THE USE OF THE CORRELATED WEIBULL AND LOGISTIC REGRESSION MODELS IN EPIDEMIOLOGY

SUBMITTED TO
THE DEPARTMENT OF STATISTICS
OF THE UNIVERSITY OF DORTMUND

IN FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF NATURAL SCIENCES

BY
REGINALD NII OTOO ODAI
FROM GHANA

DORTMUND 2003

Supervisor: Prof. Dr. Wolfgang Urfer

Co-Supervisor: PD Dr. Jürgen Groß

Date of the oral examination: July 29, 2003

THE USE OF THE CORRELATED WEIBULL AND LOGISTIC REGRESSION MODELS IN EPIDEMIOLOGY

ABSTRACT

An important factor in the analysis of family data is the dependence structure. In order to incorporate dependence within families into regression models, Bonney (1998) introduced the disposition model for the analysis of correlated binary data. In this work, the disposition model has been extended to allow for situations where quaternary-group dispositions are required. Estimation procedures for the correlated Weibull and logistic regression models have been provided for the non-nested and nested disposition models.

The correlated Weibull regression model was contrasted with the correlated logistic regression model. The results showed that both regression models were useful in explaining the familial aggregation of oesophageal cancer. The correlated logistic regression model fitted the oesophageal cancer data better than the correlated Weibull regression model. Furthermore, the correlated logistic regression model was computationally more attractive than the correlated Weibull regression model. The implications of higher level nesting of the disposition model in relation to the dimension of the parameter space have been examined and the performance of the disposition model compared to that of Cox's model using breast cancer data. It has been observed that the disposition model has a very large number of unknown parameters, and is therefore limited by the method of estimation used. In the case of the maximum likelihood method, reasonable estimates are obtained if the number of parameters in the model is at most nine. This corresponds to about four to seven covariates. Since each covariate in Cox's model provides a parameter, it is possible to include more covariates in the regression analysis. On the other hand, as opposed to Cox's model, the disposition model is fitted with parameters to capture aggregation in families, if there should be any. The choice of a particular model should therefore depend on the available data set and the purpose of the statistical analysis.

CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Prof. Wolfgang Urfer and PD Dr. Jürgen Groß, for their guidance and inspiration throughout my research. I would also like to thank my external supervisor, Prof. George E. Bonney, for encouraging me to work on the topic of my thesis and for permitting me to use the oesophageal cancer data and the breast cancer data.

My thanks go to the “Studentensekretariat” of the University of Dortmund for offering me a scholarship to pursue this research.

Special thanks go to my mother, Victoria Vanderpuije, and to my sisters, Paulina Odai and Patience Odai, for their encouragement. I am indebted to Dr. Kepher Makambi, Dr. John Kwagyan and Mr. Joachim Kötting for their helpful suggestions which improved my work. I am also indebted to Messrs. Patrice Wandja, Hayford-Seth Badu, Kayode Adekeye, Jürgen Schweiger, Joachim Gerß, Ralf Offele, to Mesdames Anja Wollsiefen, Bärbel Skopp, Brigitte Koths, Isabelle Grimmenstein, Lisete Sousa, Magdalena Thöne, to Doctors K.-H. Loesgen, J. Elsner, M. Brunnert, S. Selinsky, P. Hahnen, and to Professors H. Schulte and J. Hartung for their help and inspiration.

Finally, I would like to thank all friends and well-wishers for various forms of support they gave me during my studies.

1 Background and literature review

1.1 Introduction

The outcomes of family members are correlated because they share common risks. Thus standard methods of epidemiology, which assume independence of outcomes, are unsuitable for the analysis of family data. The disposition model is one of the possible models for the analysis of correlated binary data. It enables the characterisation of the dependence structure of a family and the response probabilities associated with it. The development of the disposition model involves the derivation and parameterisation of the joint distribution on which the likelihood function is based. Here, the experimental unit is the nuclear family and the response is the disease status. In such studies, the methods of estimating the parameters of the models are of particular importance. Here, the maximum likelihood method will be used to analyse the models. Since closed-form solutions are not possible, the Newton-Raphson iteration method is applied to obtain maximum likelihood estimates of the parameter vector. It should however be pointed out that maximum-likelihood becomes increasingly intractable as the model becomes more complex. Despite this limitation, the maximum likelihood is widely used, because it can provide accurate estimates and has some attractive optimum properties, such as asymptotically normally distributed estimators and best asymptotically normal sequence of estimators (Mood et al., 1974). Also, the maximum likelihood estimators possess the quality of functional invariance: if $\hat{\lambda}$ is the maximum likelihood estimator for λ, then $h(\hat{\lambda})$ will be the maximum likelihood estimator of $h(\lambda)$ for any function h(.) (Stuart, Ord and Arnold, 1999). In this way, the maximum likelihood estimators for a wide variety of parameterisations can be determined. With this study, potential risk factors for a disease such as smoking, age and alcohol use can be examined. Also, it can be assessed whether the disease tends to aggregate in families as a result of common shared risks. Such knowledge is decisive for counselling in the aetiology of familial disease.

The rest of the thesis is organised as follows: Section 1.2 briefly reviews the correlated regression models. Section 1.3 explains why the disposition model of the Weibull-type regression can provide more reasonable solutions than that of the logistic-type regression. In Chapter 2, the standard Weibull distribution and its parameters will be discussed. Chapter 3 briefly reviews Cox's regression model (Cox, 1972) for the analysis of failure data when explanatory variables are available. In Chapter 4, the disposition model (Bonney, 1998) and its associated likelihood function will be introduced. The first and second level extensions of the disposition model will be considered in Chapter 5. Inference for the models will be treated in the first three sections of Chapter 6. To estimate the parameters in the model, the joint function of all the clusters is required. However, there is no loss of generality if the joint function of a cluster is considered. Section 6.4 discusses the properties of the score function. The likelihood ratio test and the Wald's test will be introduced in Section 6.5 to test for the independence of familial aggregation of a disease. Section 6.6 is devoted to the comparison of the model fit of models. Chapter 7 is divided into three sections. The first, Section 7.1, contains the descriptions of two data sets: oesophageal cancer data and breast cancer data. Section 7.2 illustrates the methods with the oesophageal cancer data. The application to the breast cancer data is presented in Section 7.3. Chapter 8 gives a summary of the work and discusses experiences gained.

1.2 Review of correlated regression models

In models for clustered binary data, measures of association are of primary interest when a particular pattern of infection is suspected in a family. In the search for an appropriate model for inference on response probabilities and correlations, the equations for the estimation of the parameters become more complex. Thus, the estimation of all parameters becomes more difficult as the cluster size gets larger.

Cox (1972) reviewed several methods that had been proposed for the analysis of multivariate binary data and outlined some new proposals. He suggested the use of logistic representations, in which the joint response probability is a quadratic exponential form, as the simplest, most flexible, and in many ways the most important models. In the paper 'Partial likelihood', Cox (1975) gave a definition of partial likelihood which generalises the ideas of conditional and marginal likelihood. Here, he transformed the random variable Y into a sequence $\{X_j, S_j\}$, $j = 1,...,m$, and decomposed the full likelihood of the sequence into two products, the second product being the partial likelihood based on S in the sequence $\{X_j, S_j\}$. He pointed out that the partial likelihood is especially useful when it is appreciably simpler than the full likelihood. This is the situation when constructive procedures for finding useful partial likelihoods are provided, so that the partial likelihood involves only the parameters of interest and not nuisance parameters. To support this point, he made mention of the failure of the method of maximum likelihood as a general technique, especially in the sampling theory and pure likelihood approaches, due to excessive nuisance parameters, and hence the need to reduce dimensions. Care should however be taken to ensure that all or nearly all the relevant information is contained in the partial likelihood.

Liang and Zeger (1986) introduced the use of 'generalised estimating equations' (GEE), an extension of generalised linear models, for estimating regression parameters in situations when the vector of association parameters is a nuisance parameter. The approach is to use a working generalised linear model for the marginal distribution of the outcome variable. The method gives efficient estimates of regression coefficients, although estimates of the association among the binary outcomes can be inefficient. Liang, Zeger and Qaqish (1992) discussed the use of 'generalised estimating equations' (GEE1 and GEE2) for regression analysis of multivariate binary data, focusing on the regression and association parameters. They recommended the use of GEE1, introduced by Liang and Zeger (1986), when the association parameter is considered as a nuisance and the number of clusters is large relative to the size of each cluster. On the other hand, GEE2, introduced by Zhao and Prentice (1990), is preferable to GEE1 when there are few clusters and/or the association parameter is of primary interest. Connolly and Liang (1988) introduced the conditional logistic regression models for correlated binary data which are most useful when the dependence among observations is of main interest (such as in family data). Although the estimating functions are easily computed and have high efficiency compared to the computationally intensive maximum likelihood approach, more work is needed to determine the form of the weights used for the estimating functions $U(\beta,\theta)$. Prentice (1988) considered regression methods for the analysis of correlated binary data when each binary observation may have its own covariates. In the case of the stratified and mixture models, he generalised the binary logistic regression model for the response Y_i given the covariate x_i to blocked binary data by setting $\Pr_s\left(Y_i \mid Y_\ell, \ell \neq i, x\right) = \dfrac{\exp[(\alpha_s + x_i\beta)Y_i]}{1+\exp(\alpha_s + x_i\beta)}$, where α_s is a parameter for the sth block. In the case of the conditional models, he specified a distribution (e.g., the logistic regression model) for each binary variate given all of the other variates in the block. Here, unlike the stratified and mixture models, one may allow the logistic location parameter to depend on the other binary responses in the same block.

Zhao and Prentice (1990) reparameterised probability distribution of the model advocated by Cox (1972) in terms of marginal parameters of ready interpretation. Since this approach yields models with very complicated marginal response probabilities and pairwise correlations, they suggested the transformations of the canonical parameters (θ_k, λ_k), k = 1,...,K, to response means ($\mu_k = \mu_k(\beta)$) and covariances ($\sigma_k = \sigma_k(\beta,\alpha)$), where β and α are parameter vectors. Scoring estimating functions can then be used to evaluate mean and correlation parameters under the quadratic exponential family. Qaqish and Liang (1992) presented a model for correlated binary data, in which the marginal expectation of each binary variable and the association between pairs of outcomes are modelled separately in terms of explanatory variables. With examples, they described some drawbacks of conditional models, especially in situations where observations are missing or cluster sizes differ. On the other hand, the marginal model is reproducible, since the marginal distribution of any proper subset $(Y_1,\ldots,Y_n)$ is of the same form. Hence the situation where a subset of the cluster $(Y_1,\ldots,Y_n)$ is missing causes no problem. Carey, Zeger and Diggle (1993) proposed the use of odds ratios to measure association among responses. The approach, which alternates between two steps, estimates the association parameters by modelling the conditional distribution of one response given another. The alternating logistic regression avoids the computational burdens encountered in many problems, and its estimates are reasonably efficient relative to solutions of second-order methods.

In order to accommodate the many complicating features associated with real data, Bonney (1998) derived joint distributions for constructing likelihood functions. The central aspects of his work concern the notion of disposition to an outcome. He used a moment series representation to derive the joint distributions. Kötting, Bonney and Urfer (1998) used the ordinal-disposition-transitional model, an extension of the disposition model, to analyse dynamic changes of damage in forest-ecosystems. Odai et al. (2002) discussed the use of the correlated Weibull regression model for the analysis of multivariate binary data. The results have shown that the model provides feasible means of analysing family data.

In this dissertation, computationally attractive models with readily interpretable dependence structure for the regression analysis of correlated binary data will be presented. Estimation is based on the log-likelihood function, whose solutions can be solved by the Newton-Raphson iteration. The implications of higher level nesting in relation to the dimension of the parameter space will also be examined.

1.3 Motivation

Logistic regression is by far the most common approach to modelling the relationship between some explanatory variables and a binary response variable. This approach sometimes leads to biased estimates of covariate effects since it does not take care of dependence of outcomes. In order to incorporate dependence within families into regression models, Bonney (1998) developed the disposition models for the analysis of family data. He considered the logistic-type regression (Bonney, 1998) as the basic regression function in the non-nested disposition model. However, there are situations in which the response of interest is not a binary risk, but rather the time to failure. This is especially the case if one, for instance, wishes to know if a particular disease occurs at a certain point in time or at a certain age. The standard Weibull distribution is also inadequate for the analysis of family data, because it is not equipped with a dependence structure to take care of correlated outcomes. Furthermore, explanatory variables cannot be included in the statistical analysis. It will therefore be appropriate to consider the Weibull-type regression (Bonney, 1998) as the basic regression function in Bonney's disposition model (Bonney, 1998). Thus, in general, the correlated Weibull regresion model distinguishes itself from the correlated logistic regression model in the sense that it takes into account the special features of the underlying data (e.g., it is more suitable for the analysis of data drawn from failure distributions).

2 The standard Weibull distribution

The purpose of this section is to review some basic concepts of survival theory of the standard Weibull distribution. This is necessary since there is a link between the constructions of the likelihood functions of the standard Weibull distribution and the correlated Weibull regression model. This link will be discussed at the end of Chapter 4.

Consider the two-parameter Weibull distribution denoted by $T \sim W(\phi, \rho) \quad (\phi > 0, \rho > 0)$, where T is the lifetime of a living organism or a product, or the time until the occurrence of some specified event, ϕ is the shape parameter and ρ is the scale parameter, and let $T_1, T_2, ..., T_n$ be a random sample of size n from T.

The probability density function (PDF), which is sometimes also called the unconditional failure rate, is given by

$$f_T(t; \phi, \rho) = \begin{cases} \phi\rho t^{\phi-1} \exp(-\rho t^{\phi}), & t > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{2.1}$$

where $\phi > 0$, $\rho > 0$ are real parameters (Gross and Clark, 1975).

The cumulative distribution function (CDF)

$$F_T(t; \phi, \rho) = P(T \le t) = \begin{cases} 0, & t \le 0 \\ 1 - \exp(-\rho t^{\phi}), & t > 0 \end{cases} \tag{2.2}$$

is called the lifetime distribution or failure distribution. If T represents time at death of an individual, $F_T(t; \phi, \rho)$ is the probability that an individual dies before time t. On the other hand, if T represents age of first occurrence of a certain event (e.g., chronic disease), then $F_T(t; \phi, \rho)$ represents age of onset distribution of the event (disease) (Gross and Clark, 1975; Elandt-Johnson and Johnson, 1980).

The survival function (SF), which is defined as the probability of an individual surviving beyond time t, is given by

$$S_T(t) = \Pr(T > t) = 1 - F_T(t;\phi,\rho) = \exp(-\rho t^{\phi}) \qquad (2.3)$$

(Gross and Clark, 1975; Elandt-Johnson and Johnson, 1980). In survival analysis, $S_T(t)$ is more commonly used, instead of its complementary function, $F_T(t;\phi,\rho)$.

The hazard function (HF), which characterises the instantaneous failure rate when T = t, conditional on survival to time t, is defined mathematically as

$$h_T(t) = \lim_{\Delta t \to 0}\left[\frac{\Pr(t < T < t + \Delta t \mid T > t)}{\Delta t}\right] \qquad (2.4)$$

(Gross and Clark, 1975). The hazard function, also termed the failure rate, may also be defined as a measure of proneness to failure. This can also be expressed as

$$h_T(t) = -\frac{\frac{d}{dy}S_T(t)}{S_T(t)} = -\frac{d}{dt}\log_e S_T(t) = \phi\rho t^{\phi-1} \qquad (2.5)$$

(Gross and Clark, 1975; Nelson, 1972). For values of the shape parameter, ϕ, less than 1, the hazard function is a decreasing function, for ϕ = 1, the Weibull distribution is an exponential distribution and has a constant failure rate, and for ϕ > 1, it is an increasing function of t (Nelson, 1972). An increasing hazard rate indicates that a unit of age t is more likely to fail in a given increment of time than it would be in the same increment of time at an earlier age. For example, the probability that an individual survives to age 71, given that he has lived to age 70, is greater than the probability that an individual survives to age 72, given that he has lived to age 71. Similarly, a decreasing hazard rate means that the unit is improving with age. For example, children who have undergone an operative procedure to correct a congenital condition such as a heart defect represent a population exhibiting a decreasing hazard rate. This is because the principal risk of death is the surgery or complications immediately thereafter (Gross and Clark, 1975). A constant hazard rate results due to chance failures (e.g., accidents). Such random occurrences are often independent of age.

The failure rate function of a discrete distribution $\{p_k\}_{k=0}^{\infty}$ (e.g., geometric, binomial, poisson, etc.) is

$$h(k) = \frac{p_k}{\sum_{j=k}^{\infty} p_j}, \tag{2.6}$$

where k is the number of failures (Barlow and Proschan, 1965). We note that in this case $h(k) \leq 1$.

From (2.1), (2.3) and (2.5), it follows that

$$f_T(t) = h_T(t) S_T(t). \tag{2.7}$$

Any distribution of survival times can be characterised by the three equivalent functions $f_T(t)$, $h_T(t)$ and $S_T(t)$.

In observational studies of the time to failure of units (e.g., breakdown of a machine, death of an individual), a group of data may be incomplete in the sense that some units may not have failed by the end of the study, or may have been withdrawn before the end of the study. Such data are said to be censored (Daintith and Nelson, 1989).

Censoring is said to be on the right when the item or subject is observed prior to failure or death. Since the event time is larger than the time of observation, such an observation provides information on the survival function, $S_T(t)$, evaluated at the time of observation (Klein and Moeschberger, 1997).

On the other hand, censoring is said to be on the left when failure or death occurs prior to some designated censoring time. Since the event time has already occurred, such an observation provides information on the cumulative distribution function, $F_T(t)$, evaluated at the time of observation (Klein and Moeschberger, 1997).

An observation corresponding to an exact event time provides information on the density function of T, $f_T(t)$, at this time (Klein and Moeschberger, 1997).

The likelihood function may take the following form:

$$L \propto \prod_{j \in D} f_T(t_j) \prod_{j \in R} S_T(t_j) \prod_{j \in L} F_T(t_j), \tag{2.8}$$

where, D is the set of death times, R the set of right-censored observations and L is the set of left-censored observations (Klein and Moeschberger, 1997). If the data set comprises only right-censored and left-censored observations, the above likelihood function reduces to

$$L \propto \prod_{j \in R} S_T(t_j) \prod_{j \in L} F_T(t_j). \tag{2.9}$$

The following are some examples on censored data.

Ex. 1: In a particular clinical trial, suppose that all n patients are followed until death. Their recorded survival times are $t_1, \ldots, t_n$, and it is assumed that the death density function for the jth patient is given by the Weibull density function. The likelihood function $L(t; \phi, \rho)$ is given by

$$L(t; \phi, \rho) = \prod_{j=1}^{n} f(t_j; \phi, \rho) = \prod_{j=1}^{n} \phi \rho t_j^{\phi - 1} \exp(-\rho t_j^{\phi}) \tag{2.10}$$

(Gross and Clark, 1975).

Ex. 2: Suppose that we only know that out of n individuals starting at time zero, r died before time t', and (n – r) survived beyond t' (i.e., censored data). The statistical model for this set of data is binomial, so that the likelihood function is

$$L(t;\phi,\rho) = \binom{n}{r}[F_T(t';\theta)]^r [S_T(t';\theta)]^{n-r} \tag{2.11}$$

(Elandt-Johnson and Johnson, 1980).

3 Cox's regression model

The Cox model (also known as the proportional hazards model) is a model that can be used for the analysis of failure data when explanatory variables are available. There will be a brief review of this model and its estimation procedure in this chapter.

3.1 The model

Let $h(t;x)$ be the hazard rate at time t for an individual with risk vector $x^T = (x_1,...,x_p)$. Cox (1972) specified his model as follows:

$$h(t;x) = h_0(t)\exp(\beta^T x)\,, \qquad (3.1.1)$$

where $h_0(t)$ is an arbitrary baseline hazard rate and $\beta^T = (\beta_1,...,\beta_p)$ is a vector of unknown parameters.

The above model is often called a proportional hazards model because, the ratio of the hazard rates of two individuals with covariate values x and x' can be expressed as

$$\frac{h(t;x)}{h(t;x')} = \exp\left[\sum_{k=1}^{p}\beta_k(x_k - x_k')\right], \qquad (3.1.2)$$

which is a constant (see, for example, Klein and Moeschberger, 1997). This indicates that the hazard rates are proportional. The quantity (3.1.2), called the relative risk (hazard ratio), gives the factor by which the risk of an individual with covariate x is increased in comparison to an individual with risk factor x'.

3.2 Parameter estimation

In order to estimate the parameters in Cox's model with the maximum likelihood method, the baseline hazard, $h_0(t)$, must be specified. To deal with this situation, Cox exploited the definition of partial likelihood. Specifically, he considered the baseline hazard, $h_0(t)$, as a nuisance parameter function and concentrated mainly on the regression parameters.

Let $t_{(1)} < t_{(2)} < \ldots < t_{(n)}$ denote the ordered event times and define the risk set at time $t_{(i)}$, $R(t_{(i)})$, $i = 1,\ldots,n$, as the set of all individuals who are still under study at a time just prior to $t_{(i)}$. Further, let x_j denote the value of x for the jth individual, and $x_{(i)}$ the value for the individual failing at time $t_{(i)}$, $i = 1,\ldots,n$. Then, Cox (1972) gave the partial likelihood based on the hazard function specified by (3.1.1) as

$$L(\beta) = \prod_{i=1}^{n} \frac{\exp(\beta^T x_{(i)})}{\sum_{j \in R(t_{(i)})} \exp(\beta^T x_j)}. \tag{3.2.1}$$

It should be noted that the numerator of the likelihood in (3.2.1) depends only on information from the individual who experiences the event, whereas the denominator utilises information about all individuals who have not yet experienced the event (Klein and Moeschberger, 1997).

Direct calculation from the log-likelihood gives the score equation

$$U(\beta) = \sum_{i=1}^{n} x_{(i)} - \sum_{i=1}^{n} \frac{\sum_{j \in R(t_{(i)})} x_j \exp(\beta^T x_j)}{\sum_{j \in R(t_{(i)})} \exp(\beta^T x_j)}, \tag{3.2.2}$$

from which we obtain the Hessian matrix

$$H(\beta) = A_{(i)}(\beta) A_{(i)}^T(\beta) - \sum_{i=1}^{n} \frac{\sum_{j \in R(t_{(i)})} x_j x_j^T \exp(\beta^T x_j)}{\sum_{j \in R(t_{(i)})} \exp(\beta^T x_j)}, \tag{3.2.3}$$

where $A_{(i)} = \frac{\sum_{j \in R(t_{(i)})} x_j \exp(\beta^T x_j)}{\sum_{j \in R(t_{(i)})} \exp(\beta^T x_j)}$, $i = 1, \ldots, n$.

The Fisher information matrix is given by

$$I(\beta) = \sum_{i=1}^{n} \frac{\sum_{j \in R(t_{(i)})} x_j x_j^T \exp(\beta^T x_j)}{\sum_{j \in R(t_{(i)})} \exp(\beta^T x_j)} - \sum_{i=1}^{n} \left[\frac{\sum_{j \in R(t_{(i)})} x_j \exp(\beta^T x_j)}{\sum_{j \in R(t_{(i)})} \exp(\beta^T x_j)} \right] \left[\frac{\sum_{j \in R(t_{(i)})} x_j^T \exp(\beta^T x_j)}{\sum_{j \in R(t_{(i)})} \exp(\beta^T x_j)} \right] \qquad (3.2.4)$$

(Klein and Moeschberger, 1997). Cox (1975) has shown that the usual maximum likelihood properties hold for estimates and tests based on partial likelihoods.

4 Introduction of the non-nested disposition model

Disposition, as defined by Bonney, is the tendency of an individual or group to manifest an outcome (e.g., to be affected by a disease). The central aspect of the development of the disposition model is the derivation of joint distributions that directly capture aggregation, if there should be any. In this chapter, there will be a brief presentation of the disposition model (Bonney, 1998) and its associated joint distribution function.

Consider a binary outcome Y = 1 or 0, with q_0 group-specific covariates, $Z_0^T = (Z_{01},...,Z_{0q_0})$, and p individual-specific covariates, $X_j^T = (X_{j1},...,X_{jp})$, $j = 1,...,n$, measured on several groups of individuals. We consider two types of dispositions here: the group disposition, δ_0, which is determined by the group-specific covariates, Z_0, and the individual disposition, δ_j, which is determined by the group-specific covariates, Z_0, and the individual-specific covariates, X_j, $j = 1,...,n$.

Define the group or overall disposition, δ_0, by

$$\delta_0 = \frac{\mu_0}{\alpha_0}, \tag{4.1}$$

where μ_0 is the baseline disposition under no aggregation and α_0 is the relative disposition. Then, $\alpha_0 < 1$ corresponds to positive aggregation, $\alpha_0 = 1$ corresponds to no aggregation, and $\alpha_0 > 1$ corresponds to negative aggregation.

The logit of the group disposition can be written as

$$\log\frac{\delta_0}{1-\delta_0} = M_0(Z_0) + D_0(Z_0), \tag{4.2}$$

where

$$M_0(Z_0) = \log\frac{\mu_0}{1-\mu_0} \tag{4.3}$$

and

$$D_0(Z_0) = \log\frac{\delta_0}{1-\delta_0} - \log\frac{\mu_0}{1-\mu_0}\,. \tag{4.4}$$

We term $M_0(Z_0)$ the logit of group disposition assuming no aggregation or the cluster logit mean risk and $D_0(Z_0)$ the excess disposition due to aggregation or the excess cluster logit disposition due to dependence among members of a group.

From (4.3) and (4.4), it follows that

$$\mu_0 = \frac{1}{1+\exp\{-[M_0(Z_0)]\}}\,,\ \delta_0 = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)]\}} \tag{4.5}$$

and therefore

$$\alpha_0 = \frac{\mu_0}{\delta_0} = \frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)]\}}{1+\exp\{-[M_0(Z_0)]\}}\,. \tag{4.6}$$

Now, we decompose the logit of the individual disposition as

$$\log\frac{\delta_j}{1-\delta_j} = M_0(Z_0) + D_0(Z_0) + W_j(X_j) =: \theta_j\,, \tag{4.7}$$

j = 1,…,n, where $M_0(Z_0)$ and $D_0(Z_0)$ are as defined above, and $W_j(X_j)$ is a function of the individual-specific covariates. It follows that

$$\delta_j = \frac{1}{1+\exp(-\theta_j)} = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+W_j(X_j)]\}}, \qquad (4.8)$$

j = 1,…,n.

The joint probability for a group or cluster becomes

$$P(Y_1 = y_1,\dots,Y_n = y_n) = (1-\alpha_0)\prod_{j=1}^{n}(1-y_j) + \alpha_0\prod_{j=1}^{n}\delta_j^{y_j}(1-\delta_j)^{1-y_j}, \qquad (4.9)$$

with α_0 and δ_j as defined in (4.6) and (4.8). Explicit derivation of the joint distribution can be found in Bonney (1998). If $\alpha_0 = 1$ or $D_0(Z_0) = 0$, equation (4.9) reduces to

$$P(Y_1 = y_1,\dots,Y_n = y_n) = \prod_{j=1}^{n}\delta_j^{y_j}(1-\delta_j)^{1-y_j}, \qquad (4.10)$$

that is, the independence case. Explicit parameterisations for $M_0(Z_0)$ and $D_0(Z_0)$ are obtained by the linear models

$$M_0(Z_0) = \xi_{00} + \xi_{01}Z_{01} + \ldots + \xi_{0q_0}Z_{0q_0} \tag{4.11}$$

and

$$D_0(Z_0) = \gamma_{00} + \gamma_{01}Z_{01} + \ldots + \gamma_{0q_0}Z_{0q_0}\,. \tag{4.12}$$

The set of parameters to be determined in the model is

$$\lambda = (\xi_0, \gamma_0, \beta) = (\xi_{00}, \ldots, \xi_{0q_0}, \gamma_{00}, \ldots, \gamma_{0q_0}, \beta_1, \ldots, \beta_p)\,.$$

It is now convenient to compare and contrast the standard Weibull distribution with the correlated Weibull regression model. We denote the likelihood function of the joint distribution in Equation (4.9) by $L_k(\lambda \mid y)$, $k = 1,\ldots,K$:

$$L_k(\lambda \mid y) = (1-\alpha_0)\prod_{j=1}^{n}(1-y_j) + \alpha_0\prod_{j=1}^{n}\delta_j^{y_j}(1-\delta_j)^{1-y_j}\,,$$

$\delta_j = \dfrac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+(1-\exp(\beta_1 X_{j1}+\ldots+\beta_p X_{jp}))]\}}$, j = 1,…,n, and recall that

the likelihood function for the standard Weibull distribution based on (2.9) is

$$L \propto \prod_{j\in R} S_T(t_j) \prod_{j\in L} F_T(t_j)\,.$$

The following differences are observed. (1) In the case of the standard Weibull distribution, the response variable is a variable of time (continuous or discrete), whereas the response variable in Bonney’s disposition model presented in this dissertation is the disease status, and therefore binary. (2) As opposed to the standard Weibull distribution whose most applied characterisation revolves around its role in extreme value theory (e.g., daily maximum or minimum temperatures, precipitation, etc.), Bonney's model is fitted with parameters like δ_j and α_0 to model the effect of influential factors and to capture aggregation in families, if there should be any. Here, variables of time (e.g., age) are regarded as covariates in the model. Our concern, however, is to determine the link between the standard Weibull distribution and the correlated Weibull regression model. Suppose T is the length of time until the occurrence of a certain disease, and consider a group of size n with survival times $T_1,...,T_n$, where T_j is censored or not at time t_j with the censoring indicator $y_j = 0$ if censored, and $y_j = 1$ if uncensored. Then, in the above likelihood functions, $y_j = 0$ in the correlated Weibull regression model corresponds to the survival function in the standard Weibull distribution, and $y_j = 1$ in the correlated Weibull regression model corresponds to the cumulative distribution function in the standard Weibull distribution. In other words,

$$L_k(\lambda \mid y) = (1-\alpha_0) + \alpha_0 \prod_{j=1}^{n} (1-\delta_j)$$

$$= (1-\alpha_0) + \alpha_0 \prod_{j=1}^{n} \frac{\exp\{-[M_0(Z_0) + D_0(Z_0) + (1-\exp(\beta_1 X_{j1} + ... + \beta_p X_{jp}))]\}}{1+\exp\{-[M_0(Z_0) + D_0(Z_0) + (1-\exp(\beta_1 X_{j1} + ... + \beta_p X_{jp}))]\}}$$

corresponds to $L \propto \prod_{j \in R} S_T(t_j) = \prod_{j \in R} \exp(-\rho t_j^{\phi})$,

and

$$L_k(\lambda \mid y) = \alpha_0 \prod_{j=1}^{n} \frac{1}{1+\exp\{-[M_0(Z_0) + D_0(Z_0) + (1-\exp(\beta_1 X_{j1} + ... + \beta_p X_{jp}))]\}}$$

corresponds to $L \propto \prod_{j \in L} F_T(t_j) = \prod_{j \in L} \{1-\exp(-\rho t_j^{\phi})\}$, with the above parameters as previously defined. Thus, in this sense, the two likelihood functions are equivalent.

5 Extensions of the disposition model

In Chapter 4, we concerned ourselves with the non-nested disposition model. Consideration of the nested cases of the disposition model are to be the subjects of this chapter.

5.1 First level nesting

Consider a binary outcome Y = 1 or 0, with q_0 group-specific covariates, $Z_0^T = (Z_{01},...,Z_{0q_0})$, q subgroup-specific covariates, $Z_i^T = (Z_{i1},...,Z_{iq})$, $i = 1,...,m$, and p individual-specific covariates, $X_{ij}^T = (X_{ij1},...,X_{ijp})$, $i = 1,...,m$, $j = 1,...,n_i$, measured on several individuals. Bonney (1998) considered three types of dispositions here: the group (cluster) disposition, δ_0, which is determined by the group-specific covariates, Z_0, the subgroup disposition, δ_i, $i = 1,...,m$, which is determined by the group-specific covariates, Z_0, and the subgroup-specific covariates, Z_i, $i = 1,...,m$, and the individual disposition, δ_{ij}, $i = 1,...,m$, $j = 1,...,n_i$, which is determined by the group-specific covariates, Z_0, the subgroup-specific covariates, Z_i, $i = 1,...,m$, and the individual-specific covariates, X_{ij}, $i = 1,...,m$, $j = 1,...,n_i$.

Then, δ_0 and δ_i are given by

$$\delta_0 = \frac{\mu_0}{\alpha_0} \qquad (5.1.1)$$

and

$$\delta_i = \frac{\mu_i}{\alpha_i}, \tag{5.1.2}$$

i = 1,…,m, where μ_0 is the group baseline disposition under no aggregation, μ_i is the subgroup baseline disposition under no aggregation, α_0 is the relative disposition with respect to the group and α_i is the relative disposition with respect to subgroup i, i = 1,…,m.

The logit of the individual disposition is then

$$\log\frac{\delta_{ij}}{1-\delta_{ij}} = M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + W_{ij}(X_{ij}) =: \theta_{ij}, \tag{5.1.3}$$

i = 1,…,m, j = 1,…, n_i, where

$$M_0(Z_0) = \log\frac{\mu_0}{1-\mu_0} \tag{5.1.4}$$

is the cluster logit mean risk,

$$D_0(Z_0) = \log\frac{\delta_0}{1-\delta_0} - \log\frac{\mu_0}{1-\mu_0} \tag{5.1.5}$$

is the excess cluster logit disposition due to dependence among members of the group,

$$M_i(Z_i) = \log\frac{\mu_i}{1-\mu_i} - \log\frac{\delta_0}{1-\delta_0}, \tag{5.1.6}$$

i = 1,…,m, is the excess on the logit scale of the mean risk in subgroup i above that due to the cluster disposition,

$$D_i(Z_i) = \log\frac{\delta_i}{1-\delta_i} - \log\frac{\mu_i}{1-\mu_i}, \qquad (5.1.7)$$

i = 1,…,m, is the excess on the logit scale of the disposition in subgroup i that cannot be explained by the overall cluster disposition and the differences in μ_i, i = 1,…,m, and

$$W_{ij}(X_{ij}), \qquad (5.1.8)$$

i = 1,…,m, j = 1,…, n_i, is a function of the individual-specific covariates.

From (5.1.4) - (5.1.7), it follows that

$$\mu_0 = \frac{1}{1+\exp\{-[M_0(Z_0)]\}}, \ \delta_0 = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)]\}},$$

$$\mu_i = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}, \ i = 1,\dots,m,$$

$$\delta_i = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)]\}}, \ i = 1,\dots,m, \qquad (5.1.9)$$

and therefore

$$\alpha_0 = \frac{\mu_0}{\delta_0} = \frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)]\}}{1+\exp\{-[M_0(Z_0)]\}}, \qquad (5.1.10)$$

$$\alpha_i = \frac{\mu_i}{\delta_i} = \frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)]\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}, \qquad (5.1.11)$$

i = 1,…,m, and

$$\delta_{ij} = \frac{1}{1+\exp(-\theta_{ij})} = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+W_{ij}(X_{ij})]\}}, \tag{5.1.12}$$

$i = 1,\ldots,m,\ j = 1,\ldots,n_i$.

With these, the joint probability for the first level nesting becomes

$$P(Y_{11} = y_{11},\ldots,Y_{mn_i} = y_{mn_i}) = (1-\alpha_0)\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij})$$

$$+\ \alpha_0\prod_{i=1}^{m}\left\{(1-\alpha_i)\prod_{j=1}^{n_i}(1-y_{ij}) + \alpha_i\prod_{j=1}^{n_i}\delta_{ij}^{y_{ij}}(1-\delta_{ij})^{1-y_{ij}}\right\}. \tag{5.1.13}$$

The derivation of the joint distribution can be found in Bonney (1998). Explicit parameterisations for $M_0(Z_0)$, $D_0(Z_0)$, $M_i(Z_i)$ and $D_i(Z_i)$ are obtained by the linear models

$$M_0(Z_0) = \xi_{00} + \xi_{01}Z_{01}+\ldots+\xi_{0q_0}Z_{0q_0}, \tag{5.1.14}$$

$$D_0(Z_0) = \gamma_{00} + \gamma_{01}Z_{01}+\ldots+\gamma_{0q_0}Z_{0q_0}, \tag{5.1.15}$$

$$M_i(Z_i) = \xi_1 Z_{i1} + \ldots + \xi_q Z_{iq}, \tag{5.1.16}$$

$i = 1,\ldots,m$, and

$$D_i(Z_i) = \gamma_1 Z_{i1}+\ldots+\gamma_q Z_{iq}, \tag{5.1.17}$$

$i = 1,\ldots,m$.

The set of parameters to be determined in the model is

$$\lambda = (\xi, \gamma, \beta) = (\xi_{00}, \ldots, \xi_{0q_0}, \xi_1, \ldots, \xi_q, \gamma_{00}, \ldots, \gamma_{0q_0}, \gamma_1, \ldots, \gamma_q, \beta_1, \ldots, \beta_p) .$$

If $\alpha_i = 1$ or $D_i(Z_i) = 0$, i = 1,…,m, Equation (5.1.13) reduces to the non-nested case. Also, if $\alpha_0 = 1$ and $\alpha_i = 1$, or equivalently, if $D_0(Z_0) = 0$ and $D_i(Z_i) = 0$, Equation (5.1.13) reduces to the independence case.

5.2 Second level nesting

Consider a binary outcome Y = 1 or 0, with q_0 primary-group-specific covariates (i.e., cluster-specific covariates), $Z_0^T = (Z_{01},...,Z_{0q_0})$, q_i secondary-group-specific covariates (i.e., subgroup-specific covariates), $Z_i^T = (Z_{i1},...,Z_{iq_i})$, $i = 1,...,m$, q_{ij} tertiary-group-specific covariates, $Z_{ij}^T = (Z_{ij1},...,Z_{ijq_{ij}})$, $i = 1,...,m$, $j = 1,...,n_i$, and p unit-specific covariates, $X_{ijh}^T = (X_{ijh1},...,X_{ijhp})$, $i = 1,...,m$, $j = 1,...,n_i$, $h = 1,...,n_{ij}$, measured on several units. Four types of dispositions are considered here: the group (cluster) disposition, δ_0, which is determined by the group-specific covariates, Z_0, the subgroup disposition, δ_i, $i = 1,...,m$, which is determined by the group-specific covariates, Z_0, and the subgroup-specific covariates, Z_i, $i = 1,...,m$, the tertiary-group disposition, δ_{ij}, which is determined by the primary-group-specific covariates, Z_0, the secondary-group-specific covariates, Z_i, and the tertiary-group-specific covariates, Z_{ij}, and the unit disposition, δ_{ijh}, $i = 1,...,m$, $j = 1,...,n_i$, $h = 1,...,n_{ij}$, which is determined by the primary-group-specific covariates, Z_0, the secondary-group-specific covariates, Z_i, the tertiary-group-specific covariates, Z_{ij}, and the unit-specific covariates, X_{ijh}, $i = 1,...,m$, $j = 1,...,n_i$, $h = 1,...,n_{ij}$.

We define δ_0, δ_i and δ_{ij} as follows:

$$\delta_0 = \frac{\mu_0}{\alpha_0}, \tag{5.2.1}$$

$$\delta_i = \frac{\mu_i}{\alpha_i}, \tag{5.2.2}$$

$i = 1,...,m$, and

$$\delta_{ij} = \frac{\mu_{ij}}{\alpha_{ij}}, \tag{5.2.3}$$

$i = 1,...,m$, $j = 1,...,n_i$, where μ_0 is the primary group baseline disposition under no aggregation, μ_i is the secondary group baseline disposition under no aggregation,

μ_{ij} is the tertiary group baseline disposition under no aggregation, α_0 is the relative disposition with respect to the primary group, α_i is the relative disposition with respect to the secondary group and α_{ij} is the relative disposition with respect to the tertiary group.

The logit of the unit disposition is decomposed as

$$\log\frac{\delta_{ijh}}{1-\delta_{ijh}} = M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij}) + D_{ij}(Z_{ij}) + W_{ijh}(X_{ijh})$$

$$=: \theta_{ijh}, \tag{5.2.4}$$

$i = 1,...,m$, $j = 1,...,n_i$, $h = 1,...,n_{ij}$, where

$$M_0(Z_0) = \log\frac{\mu_0}{1-\mu_0} \tag{5.2.5}$$

is the cluster logit mean risk,

$$D_0(Z_0) = \log\frac{\delta_0}{1-\delta_0} - \log\frac{\mu_0}{1-\mu_0} \tag{5.2.6}$$

is the excess cluster logit disposition due to dependence among members of the group,

$$M_i(Z_i) = \log\frac{\mu_i}{1-\mu_i} - \log\frac{\delta_0}{1-\delta_0}, \tag{5.2.7}$$

$i = 1,\ldots,m$, is the excess on the logit scale of the mean risk in secondary group i above that due to the cluster disposition,

$$D_i(Z_i) = \log\frac{\delta_i}{1-\delta_i} - \log\frac{\mu_i}{1-\mu_i}, \tag{5.2.8}$$

$i = 1,\ldots,m$, is the excess on the logit scale of the secondary group i disposition that cannot be explained by the overall primary group disposition and the differences in μ_i,

$$M_{ij}(Z_{ij}) = \log\frac{\mu_{ij}}{1-\mu_{ij}} - \log\frac{\delta_i}{1-\delta_i}, \tag{5.2.9}$$

$i = 1,\ldots,m$, $j = 1,\ldots,n_i$, is the excess on the logit scale of the mean risk in the tertiary group j above that due to the secondary group disposition,

$$D_{ij}(Z_{ij}) = \log\frac{\delta_{ij}}{1-\delta_{ij}} - \log\frac{\mu_{ij}}{1-\mu_{ij}}, \tag{5.2.10}$$

$i = 1,...,m$, $j = 1,...,n_i$, is the excess on the logit scale of the tertiary group disposition that cannot be explained by the overall cluster disposition, the subgroup disposition and the differences in μ_{ij}, and

$$W_{ijh}(X_{ijh}), \tag{5.2.11}$$

$i = 1,...,m$, $j = 1,...,n_i$, $h = 1,...,n_{ij}$, is a function of the unit-specific covariates.

From (5.2.5) - (5.2.10), we have

$$\mu_0 = \frac{1}{1+\exp\{-[M_0(Z_0)]\}}, \quad \delta_0 = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)]\}},$$

$$\mu_i = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}, \ i = 1,...,m,$$

$$\delta_i = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)]\}}, \ i = 1,...,m,$$

$$\mu_{ij} = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})]\}},$$

$i = 1,...,m$, $j = 1,...,n_i$, and

$$\delta_{ij} = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})+D_{ij}(Z_{ij})]\}}, \tag{5.2.12}$$

$i = 1,...,m$, $j = 1,...,n_i$.

Hence,

$$\alpha_0 = \frac{\mu_0}{\delta_0} = \frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)]\}}{1+\exp\{-[M_0(Z_0)]\}}, \tag{5.2.13}$$

$$\alpha_i = \frac{\mu_i}{\delta_i} = \frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)]\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}, \tag{5.2.14}$$

$i = 1,\ldots, m$,

$$\alpha_{ij} = \frac{\mu_{ij}}{\delta_{ij}} = \frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})+D_{ij}(Z_{ij})]\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})]\}}, \tag{5.2.15}$$

$i = 1,\ldots, m$, $j = 1,\ldots, n_i$, and

$$\delta_{ijh} = \frac{1}{1+\exp\{-\theta_{ijh}\}}$$

$$= \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})+D_{ij}(Z_{ij})+W_{ijh}(X_{ijh})]\}}, \tag{5.2.16}$$

$i = 1,\ldots, m$, $j = 1,\ldots, n_i$, $h = 1,\ldots, n_{ij}$.

The joint probability for a cluster is

$$P(Y_{111}=y_{111},\ldots,Y_{mn_in_{ijh}}=y_{mn_in_{ijh}})=(1-\alpha_0)\prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})$$

$$+\ \alpha_0\prod_{i=1}^{m}\left\{(1-\alpha_i)\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})+\alpha_i\prod_{j=1}^{n_i}\left\{(1-\alpha_{ij})\prod_{h=1}^{n_{ij}}(1-y_{ijh})\right.\right.$$

$$\left.\left.+\ \alpha_{ij}\prod_{h=1}^{n_{ij}}\delta_{ijh}^{y_{ijh}}(1-\delta_{ijh})^{1-y_{ijh}}\right\}\right\}. \quad (5.2.17)$$

The following parameterisations are considered:

$$M_0(Z_0)=\xi_{00}+\xi_{01}Z_{01}+\ldots+\xi_{0q_0}Z_{0q_0}, \quad (5.2.18)$$

$$D_0(Z_0)=\gamma_{00}+\gamma_{01}Z_{01}+\ldots+\gamma_{0q_0}Z_{0q_0}, \quad (5.2.19)$$

$$M_i(Z_i)=\xi_1Z_{i1}+\ldots+\xi_qZ_{iq_i}, \quad (5.2.20)$$

$i=1,\ldots,m$,

$$D_i(Z_i)=\gamma_1Z_{i1}+\ldots+\gamma_qZ_{iq_i}, \quad (5.2.21)$$

$i=1,\ldots,m$,

$$M_{ij}(Z_{ij})=\xi_{11}Z_{ij1}+\ldots+\xi_{1q}Z_{ijq_{ij}}, \quad (5.2.22)$$

$i=1,\ldots,m$, $j=1,\ldots,n_i$, and

$$D_{ij}(Z_{ij})=\gamma_{11}Z_{ij1}+\ldots+\gamma_{1q}Z_{ijq_{ij}}, \quad (5.2.23)$$

$i = 1,\dots,m$, $j = 1,\dots,n_i$.

The set of parameters to be determined in the model is therefore

$$\lambda = (\xi, \gamma, \beta) = (\xi_{00},\dots,\xi_{0q_0}, \xi_1,\dots,\xi_{q_i}, \xi_{11},\dots,\xi_{1q_{ij}}, \gamma_{00},\dots,\gamma_{0q_0}, \gamma_1,\dots,\gamma_{q_i}, \gamma_{11},\dots,\gamma_{1q_{ij}}, \beta_1,\dots,\beta_p).$$

If $\alpha_{ij} = 1$ or $D_{ij}(Z_{ij}) = 0$, $i = 1,\dots,m$, $j = 1,\dots,n_i$, Equation (5.2.17) reduces to the first level nesting. If $\alpha_i = 1$ and $\alpha_{ij} = 1$, or equivalently, if $D_i(Z_i) = 0$ and $D_{ij}(Z_{ij}) = 0$, Equation (5.2.17) reduces to the non-nested case. Finally, if $\alpha_0 = 1$, $\alpha_i = 1$ and $\alpha_{ij} = 1$, or $D_0(Z_0) = 0$, $D_i(Z_i) = 0$ and $D_{ij}(Z_{ij}) = 0$, Equation (5.2.17) reduces to the independence case.

6 Inference

The method of maximum likelihood is used to determine estimates of the unknown model parameters, $\lambda = (\xi, \gamma, \beta)$, and make inference about them. Since closed-form solutions are not possible here, the Newton-Raphson iteration method is applied to obtain estimates of the parameter vector. The Newton-Raphson method requires the first and second derivatives of the log-likelihood functions. To estimate the parameters in the model, the joint function of all the clusters is required, but there is no loss of generality if the joint function of a cluster is considered. Unless otherwise stated, the estimation procedures developed apply to the correlated Weibull regression model.

6.1 Parameter estimation for the non-nested disposition model

Denote the likelihood function of the joint probability in Equation (4.9) by $L_k(\lambda \mid y)$, $k = 1,\dots,K$:

$$L_k(\lambda \mid y) = (1-\alpha_0)\prod_{j=1}^{n}(1-y_j) + \alpha_0\prod_{j=1}^{n}\delta_j^{y_j}(1-\delta_j)^{1-y_j}$$

$$= (1-\alpha_0)\prod_{j=1}^{n}(1-y_j) + \alpha_0 L_{\pi j}, \qquad (6.1.1)$$

where $L_{\pi j} = \prod_{j=1}^{n} L_j$, $L_j = \delta_j^{y_j}(1-\delta_j)^{1-y_j}$, $\delta_j = \dfrac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+W_j(X_j)]\}}$,

and

$W_j(X_j) = 1-\exp(\beta_1 x_{j1} + \dots + \beta_p x_{jp})$, $j = 1,\dots,n$, for the correlated Weibull regression model.

This gives the score function

$$U_k(\lambda) = A_k(\lambda)\alpha_0^* + B_k(\lambda)\left[\sum_{j=1}^{n} U_j\right], \tag{6.1.2}$$

k = 1,…,K, where $\alpha_0^*(\lambda) = -(1-\delta_0)\frac{\delta}{\delta\lambda}D_0(Z_0) + \delta_0(1-\alpha_0)\frac{\delta}{\delta\lambda}M_0(Z_0)$ (see Appendix D1 for the derivation), $A_k(\lambda) = \frac{\alpha_0\left[L_{\pi j} - \prod_{j=1}^{n}(1-y_j)\right]}{L_k}$, k = 1,…,K, $B_k(\lambda) = \frac{\alpha_0}{L_k}L_{\pi j}$, $k = 1,...,K$,

$$U_j(\lambda \mid y) = (y_j - \delta_j)\theta_j^{(1)} = (y_j - \delta_j)\frac{\delta}{\delta\lambda}[M_0(Z_0) + D_0(Z_0) + W_j(X_j)]$$

$$= (y_j - \delta_j)\begin{pmatrix} \frac{\delta}{\delta\xi_0}M_0(Z_0) \\ \frac{\delta}{\delta\gamma_0}D_0(Z_0) \\ \frac{\delta}{\delta\beta}W_j(X_j) \end{pmatrix} = (y_j - \delta_j)\begin{pmatrix} Z_0' \\ Z_0' \\ -X_j\exp(\beta^T X_j) \end{pmatrix},$$

$Z_0^{'T} = (1, Z_{01}, Z_{02}, ..., Z_{0q_0})$, $\beta^T = (\beta_1, ..., \beta_p)$ and $X_j^T = (X_{j1}, ..., X_{jp})$, $j = 1,...,n$.

The Hessian matrix is given by

$$H_k(\lambda) = \frac{\prod_{j=1}^{n}(1-y_j)}{L_k}A_k\alpha_0^*\alpha_0^{*T} + \frac{\prod_{j=1}^{n}(1-y_j)}{L_k}B_k\left[\alpha_0^*\left[\sum_{j=1}^{n}U_j\right]^T + \left[\sum_{j=1}^{n}U_j\right]\alpha_0^{*T}\right]$$

$$+ \frac{(1-\alpha_0)\prod_{j=1}^{n}(1-y_j)}{L_k}B_k\left[\sum_{j=1}^{n}U_j\right]\left[\sum_{j=1}^{n}U_j\right]^T + B_k\left[\sum_{j=1}^{n}H_j\right] + A_k\frac{\delta}{\delta\lambda}\alpha_0^*, \tag{6.1.3}$$

$k = 1,...,K$, where

$$H_j(\lambda) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -(y_j - \delta_j)X_j X_{j'}^T \exp(\beta^T X_j) \end{pmatrix}$$

$$- \delta_j(1-\delta_j)\begin{pmatrix} Z_0^{'} Z_0^{'T} & Z_0^{'} Z_0^{'T} & -Z_0^{'}[X_j \exp(\beta^T X_j)]^T \\ Z_0^{'} Z_0^{'T} & Z_0^{'} Z_0^{'T} & -Z_0^{'}[X_j \exp(\beta^T X_j)]^T \\ -[X_j \exp(\beta^T X_j)]Z_0^{'T} & -[X_j \exp(\beta^T X_j)]Z_0^{'T} & X_j X_j^T \exp(2\beta^T X_j) \end{pmatrix},$$

$j = 1,...,n$.

The Fisher Information matrix is

$$I_k(\lambda) = \alpha_0 \sum_{j=1}^{n} I_j(\lambda) - A_k^* \alpha_0^* \alpha_0^{*T} - B_k^* \left[\alpha_0^* \left[\sum_{j=1}^{n} U_j^* \right]^T + \left[\sum_{j=1}^{n} U_j^* \right] \alpha_0^{*T} \right]$$

$$- B_k^* (1-\alpha_0) \left[\sum_{j=1}^{n} U_j^* \right] \left[\sum_{j=1}^{n} U_j^* \right]^T, \qquad (6.1.4)$$

k = 1,…,K, where

$$I_j(\lambda) = \delta_j(1-\delta_j)\theta_j^{(1)}\theta_j^{(1)T}$$

$$= \delta_j(1-\delta_j)\begin{pmatrix} Z_0^{'} Z_0^{'T} & Z_0^{'} Z_0^{'T} & -Z_0^{'}[X_j \exp(\beta^T X_j)]^T \\ Z_0^{'} Z_0^{'T} & Z_0^{'} Z_0^{'T} & -Z_0^{'}[X_j \exp(\beta^T X_j)]^T \\ -[X_j \exp(\beta^T X_j)]Z_0^{'T} & -[X_j \exp(\beta^T X_j)]Z_0^{'T} & X_j X_j^T \exp(2\beta^T X_j) \end{pmatrix},$$

$j = 1,...,n$, and A_k^*, B_k^*, U_j^*, and U_j^{*T} are the resulting values of A_k, B_k, U_j, and U_j^T evaluated at y = 0 (see also Bonney, 1998).

For the correlated logistic regression model, the following are the corresponding expressions for δ_j, $U_j(\lambda)$, $H_j(\lambda)$ and $I_j(\lambda)$:

$$\delta_j = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+(\beta_1 x_{j1}+\ldots+\beta_p x_{jp})]\}}, \ j=1,\ldots,n,$$

$$U_j(\lambda \mid y) = (y_j - \delta_j)\begin{pmatrix} Z_0^{'} \\ Z_0^{'} \\ X_j \end{pmatrix}, \ j=1,\ldots,n,$$

$$H_j(\lambda) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \delta_j(1-\delta_j)\begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}X_j^T \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}X_j^T \\ X_jZ_0^{'T} & X_jZ_0^{'T} & X_jX_j^T \end{pmatrix}$$

$$= -\ \delta_j(1-\delta_j)\begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}X_j^T \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}X_j^T \\ X_jZ_0^{'T} & X_jZ_0^{'T} & X_jX_j^T \end{pmatrix}, \ j=1,\ldots,n,$$

and

$$I_j(\lambda) = \delta_j(1-\delta_j)\begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}X_j^T \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}X_j^T \\ X_jZ_0^{'T} & X_jZ_0^{'T} & X_jX_j^T \end{pmatrix}, \ j=1,\ldots,n.$$

See Appendix D4 for an equivalent form of the Fisher information matrix in Equation (6.1.4).

6.2 Parameter estimation for the first level nesting

Denote the likelihood function of the joint probability in Equation (5.1.13) by $L_k(\lambda \mid y)$, $k = 1,...,K$:

$$L_k(\lambda|y) = (1-\alpha_0)\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij}) + \alpha_0 L_{\pi i}, \tag{6.2.1}$$

where $L_{\pi i} = \prod_{i=1}^{m} L_i$, $L_i = (1-\alpha_i)\prod_{j=1}^{n_i}(1-y_{ij}) + \alpha_i L_{\pi j}$, $L_{\pi j} = \prod_{j=1}^{n_j} L_j$, $L_j = \delta_{ij}^{y_{ij}}(1-\delta_{ij})^{1-y_{ij}}$ and

$$\delta_{ij} = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+(1-\exp(\beta_1 x_{ij1}+...+\beta_p x_{ijp}))]\}},$$

$i = 1,...,m$, $j = 1,...,n_i$, for the correlated Weibull regression model.

The corresponding score function is

$$U_k(\lambda \mid y) = A_k(\lambda)\alpha_0^* + B_k(\lambda)\left[\sum_{i=1}^{m} U_i\right], \tag{6.2.2}$$

k = 1,…,K, where $\alpha_0^*(\lambda) = -(1-\delta_0)\frac{\delta}{\delta\lambda}D_0(Z_0) + \delta_0(1-\alpha_0)\frac{\delta}{\delta\lambda}M_0(Z_0)$,

$A_k(\lambda) = \frac{\alpha_0\left[L_{\pi i} - \prod_{i}^{m}\prod_{j}^{n_i}(1-y_{ij})\right]}{L_k}$, k = 1,…,K, $B_k(\lambda) = \frac{\alpha_0 L_{\pi i}}{L_k}$, k = 1,…,K,

$L_{\pi i}(\lambda \mid y) = \prod_{i=1}^{m} L_i$, $L_i = (1-\alpha_i)\prod_{j=1}^{n_i}(1-y_{ij}) + \alpha_i \prod_{j=1}^{n_i}\delta_{ij}^{y_{ij}}(1-\delta_{ij})^{1-y_{ij}}$, i = 1,…,m,

$$U_i(\lambda \mid y) = A_i(\lambda)\alpha_i^* + B_i(\lambda)\left[\sum_{j=1}^{n_i} U_j\right],$$

$\alpha_i^* = -(1-\delta_i)\frac{\delta}{\delta\lambda}D_i(Z_i) + \delta_i(1-\alpha_i)\frac{\delta}{\delta\lambda}[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]$ (see Appendix D2 for the derivation),

$$A_i(\lambda) = \frac{\alpha_i\left[L_{\pi j} - \prod_{j=1}^{n_i}(1-y_{ij})\right]}{L_i},\ B_i(\lambda) = \frac{\alpha_i}{L_i}L_{\pi j},\ i = 1,\ldots,m,\ L_{\pi j} = \prod_{j=1}^{n_j} L_j,\ L_j = \delta_{ij}^{y_{ij}}(1-\delta_{ij})^{1-y_{ij}},$$

$$U_j(\lambda \mid y) = (y_{ij}-\delta_{ij})\theta_{ij}^{(1)} = (y_{ij}-\delta_{ij})\frac{\delta}{\delta\lambda}[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+W_{ij}(X_{ij})]$$

$$= (y_{ij}-\delta_{ij})\begin{pmatrix} \frac{\delta}{\delta\xi_0}M_0(Z_0) \\ \frac{\delta}{\delta\gamma_0}D_0(Z_0) \\ \frac{\delta}{\delta\xi_i}M_i(Z_i) \\ \frac{\delta}{\delta\gamma_i}D_i(Z_i) \\ \frac{\delta}{\delta\beta}W_{ij}(X_{ij}) \end{pmatrix} = (y_{ij}-\delta_{ij})\begin{pmatrix} Z_0' \\ Z_0' \\ Z_i \\ Z_i \\ -X_{ij}\exp(\beta^T X_{ij}) \end{pmatrix},$$

$Z_0^{'T} = (1, Z_{01}, Z_{02}, \ldots, Z_{0q_0})$, $Z_i^T = (Z_{i1},\ldots,Z_{iq})$, $\beta^T = (\beta_1,\ldots,\beta_p)$ and $X_{ij}^T = (X_{ij1},\ldots,X_{ijp})$, $i = 1,\ldots,m$, $j = 1,\ldots,n_i$.

The Hessian matrix is given by

$$H_k(\lambda) = B_k\left[\sum_{i=1}^{m} H_i(\lambda)\right] + \frac{\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij})}{L_k}A_k\alpha_0^*\alpha_0^{*T} + \frac{\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij})}{L_k}B_k(1-\alpha_0)\left[\sum_{i=1}^{m}U_i\right]\left[\sum_{i=1}^{m}U_i\right]^T$$

$$+ \frac{\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij})}{L_k}B_k\left\{\left[\sum_{i=1}^{m}U_i\right]\alpha_0^{*T} + \alpha_0^*\left[\sum_{i=1}^{m}U_i\right]^T\right\} + A_k\frac{\delta}{\delta\lambda^T}\alpha_0^*, \qquad (6.2.3)$$

k = 1,…K, where

$$H_i = \frac{\prod_{j=1}^{n_i}(1-y_{ij})}{L_i} A_i \alpha_i^* \alpha_i^{*T} + \frac{\prod_{j=1}^{n_i}(1-y_{ij})}{L_i} B_i \left[\alpha_i^* \left[\sum_{j=1}^{n_i} U_j\right]^T + \left[\sum_{j=1}^{n_i} U_j\right] \alpha_i^{*T} \right]$$

$$+ \frac{(1-\alpha_i)\prod_{j=1}^{n_i}(1-y_{ij})}{L_i} B_i \left[\sum_{j=1}^{n_i} U_j\right]\left[\sum_{j=1}^{n_i} U_j\right]^T + B_i \left[\sum_{j=1}^{n_i} H_j\right] + A_i \frac{\delta}{\delta\lambda^T} \alpha_i^*$$

and

$$H_j(\lambda) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -(y_{ij}-\delta_{ij})X_{ij}X_{ij'}^T \exp(\beta^T X_{ij}) \end{pmatrix}$$

$$-\delta_{ij}(1-\delta_{ij}) \begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^T & Z_0^{'}Z_i^T & -Z_0^{'}w^T \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^T & Z_0^{'}Z_i^T & -Z_0^{'}w^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^T & Z_iZ_i^T & -Z_iw^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^T & Z_iZ_i^T & -Z_iw^T \\ -wZ_0^{'T} & -wZ_0^{'T} & -wZ_i^T & -wZ_i^T & X_{ij}X_{ij}^T \exp(2\beta^T X_{ij}) \end{pmatrix},$$

$w = [X_{ij} \exp(\beta^T X_{ij})]$, $i = 1,...,m$, $j = 1,...,n$.

The Fisher information matrix for the first level nesting is

$$I_k(\lambda) = \alpha_0 \sum_{i=1}^{m} I_i(\lambda) - A_k^* \alpha_0^* \alpha_0^{*T} - B_k^*(1-\alpha_0)\left[\sum_{i=1}^{m} U_i^*\right]\left[\sum_{i=1}^{m} U_i^*\right]^T$$

$$- B_k^* \left\{ \alpha_0^* \left[\sum_{i=1}^{m} U_i^*\right]^T + \left[\sum_{i=1}^{m} U_i^*\right] \alpha_0^{*T} \right\}, \qquad (6.2.4)$$

k = 1,…,K, where

$$I_i(\lambda) = \alpha_i \sum_{j=1}^{n_i} I_j - A_i^* \alpha_i^* \alpha_i^{*T} - B_i^*(1-\alpha_i)\left[\sum_{j=1}^{n_i} U_j^*\right]\left[\sum_{j=1}^{n_i} U_j^*\right]^T - B_i^*\left\{\alpha_i^*\left[\sum_{j=1}^{n_i} U_j^*\right]^T + \left[\sum_{j=1}^{n_i} U_j^*\right]\alpha_i^{*T}\right\},$$

i = 1,…,m,

$$I_j(\lambda) = \delta_{ij}(1-\delta_{ij})\theta_{ij}^{(1)}\theta_{ij}^{(1)T}$$

$$= \delta_{ij}(1-\delta_{ij})\begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^{T} & Z_0^{'}Z_i^{T} & -Z_0^{'}w^T \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^{T} & Z_0^{'}Z_i^{T} & -Z_0^{'}w^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^{T} & Z_iZ_i^{T} & -Z_iw^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^{T} & Z_iZ_i^{T} & -Z_iw^T \\ -wZ_0^{'T} & -wZ_0^{'T} & -wZ_i^{T} & -wZ_i^{T} & X_{ij}X_{ij}^T\exp(2\beta^T X_{ij}) \end{pmatrix},$$

$w = [X_{ij}\exp(\beta^T X_{ij})]$, $i = 1,...,m$, $j = 1,\ldots,n_i$, and A_k^*, B_k^*, A_i^*, B_i^*, U_i^*, and U_j^*, are the resulting values of A_k, B_k, A_i, B_i, U_i, and U_j evaluated at y = 0 (see also Kwagyan (2001) for the logistic-normal version).

For the correlated logistic regression model, we have the following corresponding expressions for δ_{ij}, $U_j(\lambda \mid y)$, $H_j(\lambda)$ and $I_j(\lambda)$:

$$\delta_{ij} = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+(\beta_1 x_{ij1}+...+\beta_p x_{ijp})]\}},$$

$i = 1,...,m$, $j = 1,...,n_i$,

$$U_j(\lambda \mid y) = (y_{ij}-\delta_{ij})\begin{pmatrix} Z_0^{'} \\ Z_0^{'} \\ Z_i \\ Z_i \\ X_{ij} \end{pmatrix}, \ i = 1,...,m, \ j = 1,...,n_i,$$

$$H_j(\lambda) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} - \delta_{ij}(1-\delta_{ij}) \begin{pmatrix} Z_0^{'} Z_0^{'T} & Z_0^{'} Z_0^{'T} & Z_0^{'} Z_i^{T} & Z_0^{'} Z_i^{T} & Z_0^{'} X_{ij}^{\ T} \\ Z_0^{'} Z_0^{'T} & Z_0^{'} Z_0^{'T} & Z_0^{'} Z_i^{T} & Z_0^{'} Z_i^{T} & Z_0^{'} X_{ij}^{\ T} \\ Z_i Z_0^{'T} & Z_i Z_0^{'T} & Z_i Z_i^{T} & Z_i Z_i^{T} & Z_i X_{ij}^{\ T} \\ Z_i Z_0^{'T} & Z_i Z_0^{'T} & Z_i Z_i^{T} & Z_i Z_i^{T} & Z_i X_{ij}^{\ T} \\ X_{ij} Z_0^{'T} & X_{ij} Z_0^{'T} & X_{ij} Z_i^{T} & X_{ij} Z_i^{T} & X_{ij} X_{ij}^{T} \end{pmatrix}$$

$$= -\delta_{ij}(1-\delta_{ij}) \begin{pmatrix} Z_0^{'} Z_0^{'T} & Z_0^{'} Z_0^{'T} & Z_0^{'} Z_i^{T} & Z_0^{'} Z_i^{T} & Z_0^{'} X_{ij}^{\ T} \\ Z_0^{'} Z_0^{'T} & Z_0^{'} Z_0^{'T} & Z_0^{'} Z_i^{T} & Z_0^{'} Z_i^{T} & Z_0^{'} X_{ij}^{\ T} \\ Z_i Z_0^{'T} & Z_i Z_0^{'T} & Z_i Z_i^{T} & Z_i Z_i^{T} & Z_i X_{ij}^{\ T} \\ Z_i Z_0^{'T} & Z_i Z_0^{'T} & Z_i Z_i^{T} & Z_i Z_i^{T} & Z_i X_{ij}^{\ T} \\ X_{ij} Z_0^{'T} & X_{ij} Z_0^{'T} & X_{ij} Z_i^{T} & X_{ij} Z_i^{T} & X_{ij} X_{ij}^{T} \end{pmatrix},$$

$i = 1,\ldots,m$, $j = 1,\ldots,n$, and

$$I_j(\lambda) = \delta_{ij}(1-\delta_{ij}) \begin{pmatrix} Z_0^{'} Z_0^{'T} & Z_0^{'} Z_0^{'T} & Z_0^{'} Z_i^{T} & Z_0^{'} Z_i^{T} & Z_0^{'} X_{ij}^{\ T} \\ Z_0^{'} Z_0^{'T} & Z_0^{'} Z_0^{'T} & Z_0^{'} Z_i^{T} & Z_0^{'} Z_i^{T} & Z_0^{'} X_{ij}^{\ T} \\ Z_i Z_0^{'T} & Z_i Z_0^{'T} & Z_i Z_i^{T} & Z_i Z_i^{T} & Z_i X_{ij}^{\ T} \\ Z_i Z_0^{'T} & Z_i Z_0^{'T} & Z_i Z_i^{T} & Z_i Z_i^{T} & Z_i X_{ij}^{\ T} \\ X_{ij} Z_0^{'T} & X_{ij} Z_0^{'T} & X_{ij} Z_i^{T} & X_{ij} Z_i^{T} & X_{ij} X_{ij}^{T} \end{pmatrix},$$

$i = 1,\ldots,m$, $j = 1,\ldots,n$.

See Appendix D5 for an equivalent form of the Fisher information matrix in Equation (6.2.4).

6.3 Parameter estimation for the second level nesting

Denote the likelihood function in (5.2.17) by $L_k(\lambda \mid y)$, $k = 1,...,K$:

$$L_k(\lambda|y) = (1-\alpha_0)\prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh}) + \alpha_0 L_{\pi i}$$

$$= \prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh}) + \alpha_0\left[L_{\pi i} - \prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})\right], \qquad (6.3.1)$$

where $L_{\pi i} = \prod_{i=1}^{m} L_i$, $L_i = (1-\alpha_i)\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh}) + \alpha_i L_{\pi j}$, $L_{\pi j} = \prod_{j=1}^{n_i} L_j$,

$L_j = (1-\alpha_{ij})\prod_{h=1}^{n_{ij}}(1-y_{ijh}) + \alpha_{ij} L_{\pi h}$, $L_{\pi h} = \prod_{h=1}^{n_{ij}} L_h$, $L_h = \delta_{ijh}^{y_{ijh}}(1-\delta_{ijh})^{1-y_{ijh}}$,

$$\delta_{ijh} = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})+D_{ij}(Z_{ij})+W_{ijh}(X_{ijh})]\}}$$

and

$W_{ijh}(X_{ijh}) = (1-\exp(\beta_1 X_{ijh1} + ... + \beta_p X_{ijhp}))$, $i = 1,...,m$, $j = 1,...,n_i$, $h = 1,...,n_{ij}$, for the correlated Weibull regression model.

The corresponding score function is

$$U_k(\lambda \mid y) = A_k(\lambda)\alpha_0^* + B_k(\lambda)\left[\sum_{i=1}^{m} U_i\right], \qquad (6.3.2)$$

$k = 1,...,K$, where $\alpha_0^* = -(1-\delta_0)\frac{\delta}{\delta\lambda}D_0(Z_0) + \delta_0(1-\alpha_0)\frac{\delta}{\delta\lambda}M_0(Z_0)$,

$A_k(\lambda) = \dfrac{\alpha_0\left[L_{\pi i} - \prod_{i}^{m}\prod_{j}^{n_i}\prod_{h}^{n_{ij}}(1-y_{ijh})\right]}{L_k}$, $k = 1,...,K$, $B_k(\lambda) = \dfrac{\alpha_0 L_{\pi i}}{L_k}$, $k = 1,...,K$,

$$U_i(\lambda \mid y) = A_i(\lambda)\alpha_i^* + B_i(\lambda)\left[\sum_{j=1}^{n_i} U_j\right],\ i = 1,...,m,$$

$$\alpha_i^* = -(1-\delta_i)\frac{\delta}{\delta\lambda}D_i(Z_i) + \delta_i(1-\alpha_i)\frac{\delta}{\delta\lambda}[M_0(Z_0) + D_0(Z_0) + M_i(Z_i)],\ i = 1,...,m,$$

$$A_i(\lambda) = \frac{\alpha_i\left[L_{\pi j} - \prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})\right]}{L_i},\ i = 1,...,m,\ B_i(\lambda) = \frac{\alpha_i L_{\pi j}}{L_i},\ i = 1,...,m,$$

$$U_j(\lambda \mid y) = A_j(\lambda)\alpha_{ij}^* + B_j(\lambda)\left[\sum_{h=1}^{n_{ij}} U_h\right],\ j = 1,...,n_i,$$

$$\alpha_{ij}^* = -\ (1-\delta_{ij})\frac{\delta}{\delta\lambda}D_{ij}(Z_{ij})$$

$$+\quad \delta_{ij}(1-\alpha_{ij})\frac{\delta}{\delta\lambda}\{[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}$$ (see Appendix D3 for the derivation), $A_j(\lambda) = \frac{\alpha_{ij}\left[L_{\pi h} - \prod_{h=1}^{n_{ij}}(1-y_{ijh})\right]}{L_j}$, $j = 1,...,n_i$, $B_j(\lambda) = \frac{\alpha_{ij}L_{\pi h}}{L_j}$, $j = 1,...,n_i$,

$$U_h(\lambda \mid y) = (y_{ijh} - \delta_{ijh})\theta_{ijh}^{(1)}$$

$$= (y_{ijh} - \delta_{ijh})\frac{\delta}{\delta\lambda}[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij}) + D_{ij}(Z_{ij}) + W_{ijh}(X_{ijh})]$$

$$= (y_{ijh} - \delta_{ijh})\begin{pmatrix} \frac{\delta}{\delta\xi_0}M_0(Z_0) \\ \frac{\delta}{\delta\gamma_0}D_0(Z_0) \\ \frac{\delta}{\delta\xi_i}M_i(Z_i) \\ \frac{\delta}{\delta\gamma_i}D_i(Z_i) \\ \frac{\delta}{\delta\xi_{ij}}M_{ij}(Z_{ij}) \\ \frac{\delta}{\delta\gamma_{ij}}D_{ij}(Z_{ij}) \\ \frac{\delta}{\delta\beta}W_{ij}(X_{ij}) \end{pmatrix} = (y_{ijh} - \delta_{ijh})\begin{pmatrix} Z_0' \\ Z_0' \\ Z_i' \\ Z_i' \\ Z_{ij}' \\ Z_{ij}' \\ -X_{ijh}\exp(\beta^T X_{ijh}) \end{pmatrix},$$

$Z_0^{\prime T} = (1, Z_{01}, Z_{02}, ..., Z_{0q_0})$, $Z_i^T = (Z_{i1}, ..., Z_{iq_i})$, $Z_{ij}^T = (Z_{ij1}, ..., Z_{ijq_{ij}})$, $\beta^T = (\beta_1, ..., \beta_p)$ and $X_{ijh}^T = (X_{ijh1}, ..., X_{ijhp})$, $i = 1, ..., m$, $j = 1, ..., n_i$, $h = 1, ..., n_{ij}$.

The Hessian matrix is given by

$$H_k(\lambda) = B_k\left[\sum_{i=1}^{m} H_i(\lambda)\right] + \frac{\prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})}{L_k} A_k \alpha_0^* \alpha_0^{*T} + \frac{\prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})}{L_k} B_k (1-\alpha_0)\left[\sum_{i=1}^{m} U_i\right]\left[\sum_{i=1}^{m} U_i\right]^T$$

$$+ \frac{\prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})}{L_k} B_k \left\{\left[\sum_{i=1}^{m} U_i\right]\alpha_0^{*T} + \alpha_0^*\left[\sum_{i=1}^{m} U_i\right]^T\right\} + A_k \frac{\delta}{\delta\lambda^T}\alpha_0^*, \qquad (6.3.3)$$

k = 1,…K, where

$$H_i = \frac{\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})}{L_i} A_i \alpha_i^* \alpha_i^{*T} + \frac{\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})}{L_i} B_i \left[\alpha_i^*\left[\sum_{j=1}^{n_i} U_j\right]^T + \left[\sum_{j=1}^{n_i} U_j\right]\alpha_i^{*T}\right]$$

$$+ \frac{(1-\alpha_i)\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})}{L_i} B_i\left[\sum_{j=1}^{n_i} U_j\right]\left[\sum_{j=1}^{n_i} U_j\right]^T + B_i\left[\sum_{j=1}^{n_i} H_j\right] + A_i \frac{\delta}{\delta\lambda^T}\alpha_i^*,$$

$$H_j = \frac{\prod_{h=1}^{n_{ij}}(1-y_{ijh})}{L_j} A_j \alpha_{ij}^* \alpha_{ij}^{*T} + \frac{\prod_{h=1}^{n_{ij}}(1-y_{ijh})}{L_j} B_j \left[\alpha_{ij}^*\left[\sum_{h=1}^{n_{ij}} U_h\right]^T + \left[\sum_{h=1}^{n_{ij}} U_h\right]\alpha_{ij}^{*T}\right]$$

$$+ \frac{(1-\alpha_{ij})\prod_{h=1}^{n_{ij}}(1-y_{ijh})}{L_j} B_j\left[\sum_{h=1}^{n_{ij}} U_h\right]\left[\sum_{h=1}^{n_{ij}} U_h\right]^T + B_j\left[\sum_{h=1}^{n_{ij}} H_h\right] + A_j \frac{\delta}{\delta\lambda^T}\alpha_{ij}^*$$

and

$$H_h(\lambda) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -(y_{ijh} - \delta_{ijh})X_{ijh}X_{ijh}^T \exp(2\beta^T X_{ijh}) \end{pmatrix}$$

$$-\delta_{ijh}(1-\delta_{ijh}) \begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^T & Z_0^{'}Z_i^T & Z_0^{'}Z_{ij}^T & Z_0^{'}Z_{ij}^T & -Z_0^{'}w^T \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^T & Z_0^{'}Z_i^T & Z_0^{'}Z_{ij}^T & Z_0^{'}Z_{ij}^T & -Z_0^{'}w^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^T & Z_iZ_i^T & Z_iZ_{ij}^T & Z_iZ_{ij}^T & -Z_iw^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^T & Z_iZ_i^T & Z_iZ_{ij}^T & Z_iZ_{ij}^T & -Z_iw^T \\ Z_{ij}Z_0^{'T} & Z_{ij}Z_0^{'T} & Z_{ij}Z_i^T & Z_{ij}Z_i^T & Z_{ij}Z_{ij}^T & Z_{ij}Z_{ij}^T & -Z_{ij}w^T \\ Z_{ij}Z_0^{'T} & Z_{ij}Z_0^{'T} & Z_{ij}Z_i^T & Z_{ij}Z_i^T & Z_{ij}Z_{ij}^T & Z_{ij}Z_{ij}^T & -Z_{ij}w^T \\ -wZ_0^{'T} & -wZ_0^{'T} & -wZ_i^T & -wZ_i^T & -wZ_{ij}^T & -wZ_{ij}^T & X_{ijh}X_{ijh}^T \exp(2\beta^T X_{ijh}) \end{pmatrix},$$

$w = [X_{ijh} \exp(\beta^T X_{ijh})]$, $i = 1,...,m$, $j = 1,...,n_i$, $h = 1,...,n_{ij}$.

The Fisher information matrix for the second level nesting is

$$I_k(\lambda) = \alpha_0 \sum_{i=1}^{m} I_i(\lambda) - A_k^* \alpha_0^* \alpha_0^{*T} - B_k^*(1-\alpha_0)\left[\sum_{i=1}^{m} U_i^*\right]\left[\sum_{i=1}^{m} U_i^*\right]^T$$

$$- B_k^*\left\{\alpha_0^*\left[\sum_{i=1}^{m} U_i^*\right]^T + \left[\sum_{i=1}^{m} U_i^*\right]\alpha_0^{*T}\right\}, \qquad (6.3.4)$$

$k = 1,...,K$, where

$$I_i(\lambda) = \alpha_i \sum_{j=1}^{n_i} I_j(\lambda) - A_i^* \alpha_i^* \alpha_i^{*T} - B_i^*(1-\alpha_i)\left[\sum_{j=1}^{m} U_j^*\right]\left[\sum_{j=1}^{m} U_j^*\right]^T$$

$$- B_i^*\left[\alpha_i^*\left[\sum_{j=1}^{n_i} U_j^*\right]^T + \left[\sum_{j=1}^{n_i} U_j^*\right]\alpha_i^{*T}\right],$$

$$I_j(\lambda) = \alpha_{ij}\sum_{h=1}^{n_{ij}} I_h(\lambda) - A_j^*\alpha_{ij}^*\alpha_{ij}^{*T} - B_j^*(1-\alpha_{ij})\left[\sum_{h=1}^{n_{ij}} U_h^*\right]\left[\sum_{h=1}^{n_{ij}} U_h^*\right]^T$$

$$- B_j^*\left[\alpha_{ij}^*\left[\sum_{h=1}^{n_{ij}} U_h^*\right]^T + \left[\sum_{h=1}^{n_{ij}} U_h^*\right]\alpha_{ij}^{*T}\right],$$

$$I_h(\lambda) = \delta_{ijh}(1-\delta_{ijh})\theta_{ijh}^{(1)}\theta_{ijh}^{(1)T}$$

$$= \delta_{ijh}(1-\delta_{ijh})\begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^{T} & Z_0^{'}Z_i^{T} & Z_0^{'}Z_{ij}^{T} & Z_0^{'}Z_{ij}^{T} & -Z_0^{'}w^T \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^{T} & Z_0^{'}Z_i^{T} & Z_0^{'}Z_{ij}^{T} & Z_0^{'}Z_{ij}^{T} & -Z_0^{'}w^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^{T} & Z_iZ_i^{T} & Z_iZ_{ij}^{T} & Z_iZ_{ij}^{T} & -Z_iw^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^{T} & Z_iZ_i^{T} & Z_iZ_{ij}^{T} & Z_iZ_{ij}^{T} & -Z_iw^T \\ Z_{ij}Z_0^{'T} & Z_{ij}Z_0^{'T} & Z_{ij}Z_i^{T} & Z_{ij}Z_i^{T} & Z_{ij}Z_{ij}^{T} & Z_{ij}Z_{ij}^{T} & -Z_{ij}w^T \\ Z_{ij}Z_0^{'T} & Z_{ij}Z_0^{'T} & Z_{ij}Z_i^{T} & Z_{ij}Z_i^{T} & Z_{ij}Z_{ij}^{T} & Z_{ij}Z_{ij}^{T} & -Z_{ij}w^T \\ -wZ_0^{'T} & -wZ_0^{'T} & -wZ_i^{T} & -wZ_i^{T} & -wZ_{ij}^{T} & -wZ_{ij}^{T} & X_{ijh}X_{ijh}^T\exp(2\beta^TX_{ijh}) \end{pmatrix},$$

$w = [X_{ijh}\exp(\beta^TX_{ijh})]$, $i = 1,...,m$, $j = 1,...,n_i$, $h = 1,...,n_{ij}$, and A_k^*, A_i^*, A_j^*, B_k^*, B_i^*, B_j^*, U_i^*, U_j^* and U_h^* are the resulting values of A_k, A_i, A_j, B_k, B_i, B_j, U_i, U_j and U_h evaluated at y = 0.

For the correlated logistic regression model, we have the following corresponding expressions for δ_{ijh}, $U_h(\lambda\,|\,y)$, $H_h(\lambda)$ and $I_h(\lambda)$:

$$\delta_{ijh} = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})+D_{ij}(Z_{ij})+W_{ijh}(X_{ijh})]\}},$$

$$W_{ijh}(X_{ijh}) = \beta_1X_{ijh1}+...+\beta_pX_{ijhp},\ i = 1,...,m,\ j = 1,...,n_i,\ h = 1,...,n_{ij},$$

$$U_h(\lambda \mid y) = = (y_{ijh} - \delta_{ijh}) \begin{pmatrix} Z_0^{'} \\ Z_0^{'} \\ Z_i^{'} \\ Z_i^{'} \\ Z_{ij}^{'} \\ Z_{ij}^{'} \\ X_{ijh} \end{pmatrix},$$

$Z_0^{'T} = (1, Z_{01}, Z_{02}, ..., Z_{0q_0})$, $Z_i^T = (Z_{i1}, ..., Z_{iq_i})$, $Z_{ij}^T = (Z_{ij1}, ..., Z_{ijq_{ij}})$ and $X_{ijh}^T = (X_{ijh1}, ..., X_{ijhp})$,

$i = 1, ..., m$, $j = 1, ..., n_i$, $h = 1, ..., n_{ij}$,

$$H_h(\lambda) = -\delta_{ijh}(1 - \delta_{ijh}) \begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^T & Z_0^{'}Z_i^T & Z_0^{'}Z_{ij}^T & Z_0^{'}Z_{ij}^T & -Z_0^{'}X_{ijh}^T \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^T & Z_0^{'}Z_i^T & Z_0^{'}Z_{ij}^T & Z_0^{'}Z_{ij}^T & -Z_0^{'}X_{ijh}^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^T & Z_iZ_i^T & Z_iZ_{ij}^T & Z_iZ_{ij}^T & -Z_iX_{ijh}^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^T & Z_iZ_i^T & Z_iZ_{ij}^T & Z_iZ_{ij}^T & -Z_iX_{ijh}^T \\ Z_{ij}Z_0^{'T} & Z_{ij}Z_0^{'T} & Z_{ij}Z_i^T & Z_{ij}Z_i^T & Z_{ij}Z_{ij}^T & Z_{ij}Z_{ij}^T & -Z_{ij}X_{ijh}^T \\ Z_{ij}Z_0^{'T} & Z_{ij}Z_0^{'T} & Z_{ij}Z_i^T & Z_{ij}Z_i^T & Z_{ij}Z_{ij}^T & Z_{ij}Z_{ij}^T & -Z_{ij}X_{ijh}^T \\ X_{ijh}Z_0^{'T} & X_{ijh}Z_0^{'T} & X_{ijh}Z_i^T & X_{ijh}Z_i^T & X_{ijh}Z_{ij}^T & X_{ijh}Z_{ij}^T & X_{ijh}X_{ijh}^T \end{pmatrix},$$

$i = 1, ..., m$, $j = 1, ..., n_i$, $h = 1, ..., n_{ij}$, and

$$I_h(\lambda) = \delta_{ijh}(1 - \delta_{ijh}) \begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^T & Z_0^{'}Z_i^T & Z_0^{'}Z_{ij}^T & Z_0^{'}Z_{ij}^T & -Z_0^{'}X_{ijh}^T \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^T & Z_0^{'}Z_i^T & Z_0^{'}Z_{ij}^T & Z_0^{'}Z_{ij}^T & -Z_0^{'}X_{ijh}^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^T & Z_iZ_i^T & Z_iZ_{ij}^T & Z_iZ_{ij}^T & -Z_iX_{ijh}^T \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^T & Z_iZ_i^T & Z_iZ_{ij}^T & Z_iZ_{ij}^T & -Z_iX_{ijh}^T \\ Z_{ij}Z_0^{'T} & Z_{ij}Z_0^{'T} & Z_{ij}Z_i^T & Z_{ij}Z_i^T & Z_{ij}Z_{ij}^T & Z_{ij}Z_{ij}^T & -Z_{ij}X_{ijh}^T \\ Z_{ij}Z_0^{'T} & Z_{ij}Z_0^{'T} & Z_{ij}Z_i^T & Z_{ij}Z_i^T & Z_{ij}Z_{ij}^T & Z_{ij}Z_{ij}^T & -Z_{ij}X_{ijh}^T \\ X_{ijh}Z_0^{'T} & X_{ijh}Z_0^{'T} & X_{ijh}Z_i^T & X_{ijh}Z_i^T & X_{ijh}Z_{ij}^T & X_{ijh}Z_{ij}^T & X_{ijh}X_{ijh}^T \end{pmatrix},$$

$i = 1, ..., m$, $j = 1, ..., n_i$, $h = 1, ..., n_{ij}$.

6.4 Properties of the score function

Following Henze (1995) and Van der Vaart (1998), the following properties of the score function must be satisfied, under the regularity conditions, in order for the maximum likelihood estimators to be asymptotically optimal:

(a) $E[U(\lambda)] = 0$,

(b) $Var[U(\lambda)] = I(\lambda)$.

Proofs:

6.4.1 <u>The independence case:</u>

Let n be the number of observations in a cluster and $Y_1,...,Y_n$ independent and identically distributed random variables with common distribution function:

$$P(Y_1,...,Y_n) = \prod_{j=1}^{n} \delta_j^{y_j}(1-\delta_j)^{1-y_j} \,,$$

where $\delta_j = \dfrac{1}{1+\exp\{-[M_0(Z_0)+W_j(X_j)]\}}$, $j = 1,...,n$,

$W_j(X_j) = 1-\exp(\beta_1 x_1 + ... + \beta_p x_p)$ for the correlated Weibull regression model

and

$W_j(X_j) = \beta_1 x_1 + ... + \beta_p x_p$ for the correlated logistic regression model.

It follows that $P(Y_j) = \delta_j^{y_j}(1-\delta_j)^{1-y_j} \Rightarrow P(Y_j = 0) = 1-\delta_j$, $P(Y_j = 1) = \delta_j$.

Then,

(a) $$E\left[\sum_{j=1}^{n} U_j\right] = E\left[\sum_{j=1}^{n} (Y_j - \delta_j)\theta_j^{(1)}\right] \qquad \left(\theta_j^{(1)} = \frac{\delta}{\delta\lambda}\left(M(Z_0) + D(Z_0) + W(X_j)\right)\right)$$

$$= \sum_{j=1}^{n} \{E[(Y_j - \delta_j)\theta_j^{(1)}]\}$$

$$= \sum_{j=1}^{n} \{[E(Y_j) - \delta_j]\theta_j^{(1)}\}$$

$$= \sum_{j=1}^{n} \{[\sum_{\underset{\sim}{y}} yP(Y = y) - \delta_j]\theta_j^{(1)}\}$$

$$= \sum_{j=1}^{n} \{[(0P(Y_j = 0) + 1P(Y_j = 1)) - \delta_j]\theta_j^{(1)}\}$$

$$= \sum_{j=1}^{n} \{[\delta_j - \delta_j]\theta_j^{(1)}\}$$

$$= 0.$$

(b) $$Var\left[\sum_{j=1}^{n} U_j\right] = Var\left[\sum_{j=1}^{n} (Y_j - \delta_j)\theta_j^{(1)}\right],$$

$$= \sum_{j=1}^{n} \{Var[(Y_j - \delta_j)\theta_j^{(1)}]\}$$

$$= \sum_{j=1}^{n} \{[Var(Y_j)]\theta_j^{(1)}\theta_j^{(1)T}\}$$

$$= \sum_{j=1}^{n} \{[E(Y_j^2) - [E(Y_j)]^2]\theta_j^{(1)}\theta_j^{(1)T}\}$$

$$= \sum_{j=1}^{n} \{[\sum_{\underset{\sim}{y}} y^2 P(Y = y) - \delta_j^2]\theta_j^{(1)}\theta_j^{(1)T}\}$$

$$= \sum_{j=1}^{n} \{[0^2 P(Y_j = 0) + 1^2 P(Y_j = 1) - \delta_j^2]\theta_j^{(1)}\theta_j^{(1)T}\}$$

$$= \sum_{j=1}^{n} \{\delta_j - \delta_j^2\}\theta_j^{(1)}\theta_j^{(1)T}$$

$$= \sum_{j=1}^{n} \delta_j (1-\delta_j) \theta_j^{(1)} \theta_j^{(1)T} = \sum_{j=1}^{n} I_j(\lambda) \text{ (see Equation (6.1.4)).}$$

6.4.2 The non-nested case:

(a) $$E[U_k(\lambda)] = \sum_{\underset{\sim}{y}} U_k(\lambda) L_k(\lambda)$$

$$= \sum_{\underset{\sim}{y}} L_k \frac{\delta}{\delta\lambda} \log L_k$$

$$= \sum_{\underset{\sim}{y}} \frac{\delta}{\delta\lambda} L_k$$

$$= \frac{\delta}{\delta\lambda} \sum_{\underset{\sim}{y}} L_k$$

$$= \frac{\delta}{\delta\lambda} 1$$

$$= 0.$$

(b) $$\mathrm{Var}[U_k(\lambda)] = E[U_k U_k^T] \text{ (using the results in (6.4.2(a)))}$$

$$= E[(A_k \alpha_0^* + B_k U)(A_k \alpha_0^* + B_i U)^T] \qquad \left(U = \sum_{j=1}^{n} U_j \right)$$

$$= E[A_k A_k \alpha_0^* \alpha_0^{*T} + A_k B_k \alpha_0^* U^T + A_k B_k U \alpha_0^{*T} + B_k B_k U U^T]$$

$$= \sum_{\underset{\sim}{y}} \left\{ \frac{\alpha_0}{L_k} [L_{\pi j} - \prod_{j=1}^{n} (1-y_j)] A_k \alpha_0^* \alpha_0^{*T} \right.$$

$$\left. + \frac{\alpha_0}{L_k} [L_{\pi j} - \prod_{j=1}^{n} (1-y_j)] B_k [\alpha_0^* U^T + U \alpha_0^{*T}] + \frac{\alpha_0 L_{\pi j}}{L_k} B_k U U^T \right\} L_k$$

$$= \sum_{\underset{\sim}{y}} \left\{ [\alpha_0 L_{\pi j} - \alpha_0 \prod_{j=1}^{n} (1-y_j)] A_k \alpha_0^* \alpha_0^{*T} \right.$$

$$\left. + [\alpha_0 L_{\pi j} - \alpha_0 \prod_{j=1}^{n} (1-y_j)] B_k [\alpha_0^* U^T + U \alpha_0^{*T}] + \alpha_0 L_{\pi j} B_k U U^T \right\}.$$

By reformulating and inserting (6.1.1) in the above expression, we obtain

$$\mathrm{Var}[U_k(\lambda)] = \sum_{\underset{\sim}{y}} \Bigg\{ [L_k - \prod_{j=1}^{n}(1-y_j)] A_k \alpha_0^* \alpha_0^{*T}$$

$$+ \left[L_k - \prod_{j=1}^{n}(1-y_j)] B_k [\alpha_0^* U^T + U\alpha_0^{*T} \right] + \left[L_k - (1-\alpha_0)\prod_{j=1}^{n}(1-y_j) \right] B_k UU^T \Bigg\}.$$

Grouping all terms in L_k, we obtain

$$\mathrm{Var}[U_k(\lambda)] = \sum_{\underset{\sim}{y}} \Bigg\{ -\prod_{j=1}^{n}(1-y_j) A_k \alpha_0^* \alpha_0^{*T}$$

$$- \prod_{j=1}^{n}(1-y_j) B_k [\alpha_0^* U^T + U\alpha_0^{*T}] - (1-\alpha_0)\prod_{j=1}^{n}(1-y_j)] B_k UU^T$$

$$+ L_k [A_k \alpha_0^* \alpha_0^{*T} + B_k [\alpha_0^* U^T + U\alpha_0^{*T}] + B_k UU^T] \Bigg\}.$$

By substituting for A_k and B_k in the above expression and simplifying, we obtain

$$\mathrm{Var}[U_k(\lambda)] = \sum_{\underset{\sim}{y}} \Bigg\{ -\prod_{j=1}^{n}(1-y_j) A_k \alpha_0^* \alpha_0^{*T}$$

$$- \prod_{j=1}^{n}(1-y_j) B_k [\alpha_0^* U^T + U\alpha_0^{*T}] - (1-\alpha_0)\prod_{j=1}^{n}(1-y_j)] B_k UU^T$$

$$+ \alpha_0 \left[L_{\pi j} - \prod_{j=1}^{n}(1-y_j) \right] \alpha_0^* \alpha_0^{*T} + \alpha_0 L_{\pi j} [\alpha_0^* U^T + U\alpha_0^{*T}] + \alpha_0 L_{\pi j} UU^T] \Bigg\}$$

$$\overset{\sum_{\underset{\sim}{y}} = \sum_{\underset{\sim}{y}=0} + \sum_{\underset{\sim}{y} \neq 0}}{=} - A_k^* \alpha_0^* \alpha_0^{*T} + 0 - B_k^* [\alpha_0^* U^{*T} + U^* \alpha_0^{*T}] + 0 - (1-\alpha_0) B_k^* U^* U^{*T} + 0$$

$$+ \alpha_0 \sum_{\underset{\sim}{y}} \left[L_{\pi j} - \prod_{j=1}^{n}(1-y_j) \right] \alpha_0^* \alpha_0^{*T} + \alpha_0 \alpha_0^* \sum_{\underset{\sim}{y}} \underbrace{L_{\pi j} U}_{\frac{\delta}{\delta\lambda} L_{\pi j}} + \alpha_0 \alpha_0^{*T} \sum_{\underset{\sim}{y}} \underbrace{L_{\pi j} U}_{\frac{\delta}{\delta\lambda} L_{\pi j}}$$

$$+ \alpha_0 \sum_{\underset{\sim}{y}} L_{\pi j} UU^T$$

$$= - A_k^* \alpha_0^* \alpha_0^{*T} - B_k^* [\alpha_0^* U^{*T} + U^* \alpha_0^{*T}] - B_k^* (1-\alpha_0) U^* U^{*T} + \alpha_0 E[UU^T]$$

$$= \alpha_0 \sum_{j=1}^{n} I_j(\lambda) - A_k^* \alpha_0^* \alpha_0^{*T} - B_k^* [\alpha_0^* U^{*T} + U^* \alpha_0^{*T}] - B_k^* (1-\alpha_0) U^* U^{*T},$$

where A_k^*, B_k^*, U^* and U^{*T} are the resulting values of A_k, B_k, U and U^T evaluated at $y = (y_1, \ldots, y_n) = 0$.

6.4.3 First level nesting:

(a) $E(U_k) = \sum_{\underset{\sim}{y}} U_k L_k$

$$= \sum_{\underset{\sim}{y}} \left\{ B_k \left[\sum_{i=1}^{m} U_i \right] + A_k \alpha_0^* \right\} L_k$$

$$= \sum_{\underset{\sim}{y}} \left\{ \alpha_0 L_{\pi i} \left[\sum_{i=1}^{m} U_i \right] \right\} + \sum_{\underset{\sim}{y}} \left\{ \alpha_0 \left[L_{\pi i} - \prod_{i=1}^{m} \prod_{j=1}^{n_i} (1 - y_{ij}) \right] \right\} \alpha_0^*$$

$$= \alpha_0 \sum_{\underset{\sim}{y}} \left\{ L_{\pi i} \left[\sum_{i=1}^{m} U_i \right] \right\} + \alpha_0 \alpha_0^* \sum_{\underset{\sim}{y}} \left\{ \left[L_{\pi i} - \prod_{i=1}^{m} \prod_{j=1}^{n_i} (1 - y_{ij}) \right] \right\}$$

$$= \alpha_0 \sum_{\underset{\sim}{y}} \left\{ \prod_{i=1}^{m} L_i \left[\sum_{i=1}^{m} \frac{\delta}{\delta\lambda} \log L_i \right] \right\} + 0$$

$$= \alpha_0 \sum_{\underset{\sim}{y}} \left\{ \prod_{i=1}^{m} L_i \frac{\delta}{\delta\lambda} \log \left[\prod_{i=1}^{m} L_i \right] \right\}$$

$$= \alpha_0 \sum_{\underset{\sim}{y}} \left\{ \frac{\delta}{\delta\lambda} \left[\prod_{i=1}^{m} L_i \right] \right\}$$

$$= \alpha_0 \frac{\delta}{\delta\lambda} \sum_{\underset{\sim}{y}} \left\{ \left[\prod_{i=1}^{m} L_i \right] \right\} = \alpha_0 \frac{\delta}{\delta\lambda} 1 = 0.$$

(b) $\text{Var}(U_k) = E\{[U_k - E(U_k)][U_k - E(U_k)]^T\}$

$= E\{U_k U_k^T\}$ (using the results in 6.4.3(a))

$$= E\left\{\left[B_k\left(\sum_{i=1}^{m} U_i\right) + A_k\alpha_0^*\right]\left[B_k\left(\sum_{i=1}^{m} U_i\right) + A_k\alpha_0^*\right]^T\right\}$$

$$= E\left\{B_k B_k\left(\sum_{i=1}^{m} U_i\right)\left(\sum_{i=1}^{m} U_i\right)^T + A_k B_k \alpha_0^*\left(\sum_{i=1}^{m} U_i\right)^T + A_k B_k\left(\sum_{i=1}^{m} U_i\right)\alpha_0^{*T}\right.$$

$$\left. + A_k A_k \alpha_0^* \alpha_0^{*T}\right\}$$

$$= \sum_{\underset{\sim}{y}}\left\{B_k B_k\left(\sum_{i=1}^{m} U_i\right)\left(\sum_{i=1}^{m} U_i\right)^T + A_k B_k \alpha_0^*\left(\sum_{i=1}^{m} U_i\right)^T + A_k B_k\left(\sum_{i=1}^{m} U_i\right)\alpha_0^{*T}\right.$$

$$\left. + A_k A_k \alpha_0^* \alpha_0^{*T}\right\} L_k$$

$$= \sum_{\underset{\sim}{y}}\left\{\alpha_0 L_{\pi i} B_k\left(\sum_{i=1}^{m} U_i\right)\left(\sum_{i=1}^{m} U_i\right)^T\right\}$$

$$+ \sum_{\underset{\sim}{y}}\left\{\alpha_0\left[L_{\pi i} - \prod_{i=1}^{m}\prod_{j=1}^{n_i}(1 - y_{ij})\right] B_k\left[\alpha_0^*\left(\sum_{i=1}^{m} U_i\right)^T + \left(\sum_{i=1}^{m} U_i\right)\alpha_0^{*T}\right]\right\}$$

$$+ \sum_{\underset{\sim}{y}}\left\{\alpha_0\left[L_{\pi i} - \prod_{i=1}^{m}\prod_{j=1}^{n_i}(1 - y_{ij})\right] A_k \alpha_0^* \alpha_0^{*T}\right\}.$$

Making $\alpha_0 L_{\pi i}$ and $\alpha_0\left[L_{\pi i} - \prod_{i=1}^{m}\prod_{j=1}^{n_i}(1 - y_{ij})\right]$ the subjects in (6.2.1) and substituting the resulting expressions in the above, we obtain

$$\operatorname{Var}(U_k) = \sum_{\underset{\sim}{y}} \left\{ \left[L_k - (1-\alpha_0)\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij}) \right] B_k \left(\sum_{i=1}^{m} U_i \right)\left(\sum_{i=1}^{m} U_i \right)^T \right\}$$

$$+ \sum_{\underset{\sim}{y}} \left\{ \left[L_k - \prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij}) \right] B_k \left[\alpha_0^* \left(\sum_{i=1}^{m} U_i \right)^T + \left(\sum_{i=1}^{m} U_i \right) \alpha_0^{*T} \right] \right\}$$

$$+ \sum_{\underset{\sim}{y}} \left\{ \left[L_k - \prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij}) \right] A_k \alpha_0^* \alpha_0^{*T} \right\}.$$

Grouping all terms in L_k, we obtain

$$\operatorname{Var}(U_k) = \sum_{\underset{\sim}{y}} \left\{ B_k \left(\sum_{i=1}^{m} U_i \right)\left(\sum_{i=1}^{m} U_i \right)^T + B_k \left[\alpha_0^* \left(\sum_{i=1}^{m} U_i \right)^T + \left(\sum_{i=1}^{m} U_i \right) \alpha_0^{*T} \right] + A_k \alpha_0^* \alpha_0^{*T} \right\} L_k$$

$$- \sum_{\underset{\sim}{y}} \left\{ (1-\alpha_0)\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij}) B_k \left(\sum_{i=1}^{m} U_i \right)\left(\sum_{i=1}^{m} U_i \right)^T \right\}$$

$$- \sum_{\underset{\sim}{y}} \left\{ \prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij}) B_k \left[\alpha_0^* \left(\sum_{i=1}^{m} U_i \right)^T + \left(\sum_{i=1}^{m} U_i \right) \alpha_0^{*T} \right] \right\}$$

$$- \sum_{\underset{\sim}{y}} \left\{ \prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij}) A_k \alpha_0^* \alpha_0^{*T} \right\}$$

$$= \alpha_0 \sum_{\underset{\sim}{y}} \left\{ \left(\prod_{i=1}^{m} L_i \right)\left(\sum_{i=1}^{m} U_i \right)\left(\sum_{i=1}^{m} U_i \right)^T \right\} + \alpha_0 \alpha_0^* \sum_{\underset{\sim}{y}} \left\{ \left(\prod_{i=1}^{m} L_i \right)\left(\sum_{i=1}^{m} U_i \right)^T \right\}$$

$$+ \alpha_0 \alpha_0^{*T} \sum_{\underset{\sim}{y}} \left\{ \left(\prod_{i=1}^{m} L_i \right)\left(\sum_{i=1}^{m} U_i \right) \right\} + \alpha_0 \sum_{\underset{\sim}{y}} \left[L_{\pi i} - \prod_{i=1}^{m}\prod_{j=1}^{n_i}\left(1-y_{ij}\right) \right]$$

$$- (1-\alpha_0) B_k^* \left(\sum_{i=1}^{m} U_i^* \right)\left(\sum_{i=1}^{m} U_i^* \right)^T - B_k^* \left[\alpha_0^* \left(\sum_{i=1}^{m} U_i^* \right)^T + \left(\sum_{i=1}^{m} U_i^* \right) \alpha_0^{*T} \right]$$

$$- A_k^* \alpha_0^* \alpha_0^{*T}$$

$$= \alpha_0 \sum_{\underset{\sim}{y}} \left\{ \left(\prod_{i=1}^{m} L_i \right) \left(\sum_{i=1}^{m} \frac{\delta}{\delta\lambda} \log L_i \right) \left(\sum_{i=1}^{m} \frac{\delta}{\delta\lambda^T} \log L_i \right) \right\}$$

$$+ \alpha_0 \alpha_0^* \sum_{\underset{\sim}{y}} \left\{ \left(\prod_{i=1}^{m} L_i \right) \frac{\delta}{\delta\lambda^T} \log \left(\prod_{i=1}^{m} L_i \right) \right\}$$

$$+ \alpha_0 \alpha_0^{*T} \sum_{\underset{\sim}{y}} \left\{ \left(\prod_{i=1}^{m} L_i \right) \frac{\delta}{\delta\lambda} \log \left(\prod_{i=1}^{m} L_i \right) \right\} + 0 - (1-\alpha_0) B_k^* \left(\sum_{i=1}^{m} U_i^* \right) \left(\sum_{i=1}^{m} U_i^* \right)^T$$

$$- B_k^* \left[\alpha_0^* \left(\sum_{i=1}^{m} U_i^* \right)^T + \left(\sum_{i=1}^{m} U_i^* \right) \alpha_0^{*T} \right] - A_k^* \alpha_0^* \alpha_0^{*T}$$

$$= \alpha_0 \sum_{\underset{\sim}{y}} \left\{ \left(\prod_{i=1}^{m} L_i \right) \left[\frac{\delta}{\delta\lambda} \log \left(\prod_{i=1}^{m} L_i \right) \right] \left[\frac{\delta}{\delta\lambda^T} \log \left(\prod_{i=1}^{m} L_i \right) \right] \right\} + \alpha_0 \alpha_0^* \underbrace{\sum_{\underset{\sim}{y}} \left\{ \frac{\delta}{\delta\lambda^T} \left(\prod_{i=1}^{m} L_i \right) \right\}}_{= \frac{\delta}{\delta\lambda^T} \sum_{\underset{\sim}{y}} \left(\prod_{i=1}^{m} L_i \right)}$$

$$+ \alpha_0 \alpha_0^{T*} \underbrace{\sum_{\underset{\sim}{y}} \left\{ \frac{\delta}{\delta\lambda} \left(\prod_{i=1}^{m} L_i \right) \right\}}_{= \frac{\delta}{\delta\lambda} \sum_{\underset{\sim}{y}} \left(\prod_{i=1}^{m} L_i \right)} - (1-\alpha_0) B_k^* \left(\sum_{i=1}^{m} U_i^* \right) \left(\sum_{i=1}^{m} U_i^* \right)^T$$

$$- B_k^* \left[\alpha_0^* \left(\sum_{i=1}^{m} U_i^* \right)^T + \left(\sum_{i=1}^{m} U_i^* \right) \alpha_0^{*T} \right] - A_k^* \alpha_0^* \alpha_0^{*T}$$

$$= \alpha_0 \sum_{\underset{\sim}{y}} \left\{ \left[\frac{\delta}{\delta\lambda} \left(\prod_{i=1}^{m} L_i \right) \right] \left[\frac{\delta}{\delta\lambda} \left(\prod_{i=1}^{m} L_i \right) \right]^T \left(\prod_{i=1}^{m} L_i \right)^{-1} \right\} + 0 + 0$$

$$- (1-\alpha_0) B_k^* \left(\sum_{i=1}^{m} U_i^* \right) \left(\sum_{i=1}^{m} U_i^* \right)^T - B_k^* \left[\alpha_0^* \left(\sum_{i=1}^{m} U_i^* \right)^T + \left(\sum_{i=1}^{m} U_i^* \right) \alpha_0^{*T} \right]$$

$$- A_k^* \alpha_0^* \alpha_0^{*T}$$

$$= \alpha_0 \sum_{i=1}^{m} I_i(\lambda) - A_k^* \alpha_0^* \alpha_0^{*T} - B_k^* (1-\alpha_0) \left(\sum_{i=1}^{m} U_i^* \right) \left(\sum_{i=1}^{m} U_i^* \right)^T$$

$$- B_k^*\left[\alpha_0^*\left(\sum_{i=1}^{m} U_i^*\right)^T + \left(\sum_{i=1}^{m} U_i^*\right)\alpha_0^{*T}\right] = I_k(\lambda)$$

(see Ibragimov and Has'minskii, 1981; Eberl, 1982).

6.4.4 Second level nesting

(a) $E(U_k) = \sum_{\tilde{y}} U_k L_k$

$$= \sum_{\tilde{y}}\left\{B_k\left[\sum_{i=1}^{m} U_i\right] + A_k\alpha_0^*\right\}L_k$$

$$= \sum_{\tilde{y}}\left\{\alpha_0 L_{\pi i}\left[\sum_{i=1}^{m} U_i\right]\right\} + \sum_{\tilde{y}}\left\{\alpha_0\left[L_{\pi i} - \prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1 - y_{ijh})\right]\right\}\alpha_0^*.$$

Proceeding in the same manner as in the first level nesting, we obtain

$$E(U_k) = \alpha_0 \frac{\delta}{\delta\lambda}\sum_{\tilde{y}}\left\{\left[\prod_{i=1}^{m} L_i\right]\right\}$$

$$= \alpha_0 \frac{\delta}{\delta\lambda}1$$

$$= 0.$$

(b) $\mathrm{Var}(U_k) = E\{[U_k - E(U_k)][U_k - E(U_k)]^T\}$

$= E\{U_k U_k^T\}$ (using the results in 6.4.4(a))

$$= E\left\{\left[B_k\left(\sum_{i=1}^{m} U_i\right) + A_k\alpha_0^*\right]\left[B_k\left(\sum_{i=1}^{m} U_i\right) + A_k\alpha_0^*\right]^T\right\}$$

$$= E\left\{B_k B_k\left(\sum_{i=1}^{m} U_i\right)\left(\sum_{i=1}^{m} U_i\right)^T + A_k B_k \alpha_0^*\left(\sum_{i=1}^{m} U_i\right)^T + A_k B_k\left(\sum_{i=1}^{m} U_i\right)\alpha_0^{*T}\right.$$

$$\left. + A_k A_k \alpha_0^* \alpha_0^{*T}\right\} L_k$$

$$= \sum_{\underset{\sim}{y}}\left\{\alpha_0 L_{\pi i} B_k\left(\sum_{i=1}^{m} U_i\right)\left(\sum_{i=1}^{m} U_i\right)^T\right\}$$

$$+ \sum_{\underset{\sim}{y}}\left\{\alpha_0\left[L_{\pi i} - \prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})\right] B_k\left[\alpha_0^*\left(\sum_{i=1}^{m} U_i\right)^T + \left(\sum_{i=1}^{m} U_i\right)\alpha_0^{*T}\right]\right\}$$

$$+ \sum_{\underset{\sim}{y}}\left\{\alpha_0\left[L_{\pi i} - \prod_{i=1}^{m}\prod_{j=1}^{n_i}\prod_{h=1}^{n_{ij}}(1-y_{ijh})\right] A_k \alpha_0^* \alpha_0^{*T}\right\}.$$

Following the same steps as in the first level nesting, we obtain

$$\mathrm{Var}(U_k) = \alpha_0 \sum_{i=1}^{m} I_i(\lambda) \text{ - } A_k^* \alpha_0^* \alpha_0^{*T} \text{ - } B_k^*(1-\alpha_0)\left(\sum_{i=1}^{m} U_i^*\right)\left(\sum_{i=1}^{m} U_i^*\right)^T$$

$$\text{- } B_k^*\left[\alpha_0^*\left(\sum_{i=1}^{m} U_i^*\right)^T + \left(\sum_{i=1}^{m} U_i^*\right)\alpha_0^{*T}\right] = I_k(\lambda).$$

Thus in all the four cases, the expectation of the score function is equal to zero and the variance of the score function is equal to the Fisher information matrix. With these proofs, the maximum likelihood estimates will be consistent, under the usual regularity conditions. This is because the condition for consistency (that the mean of the score statistic be 0) is unaffected by dependence (see the suggestions by Smith (1992) in the discussion of Liang et al. (1992)). Also, under fairly mild regularity conditions, the sampling distribution of the vector of parameters, $\hat{\lambda}$, approaches a normal distribution as the sample size grows larger. That is, $\mathrm{Var}(\hat{\lambda}) = \left\{E\left[\left(\frac{\delta}{\delta\lambda}\log L\right)\left(\frac{\delta}{\delta\lambda}\log L\right)^T\right]\right\}^{-1} = I^{-1}(\lambda)$ (Bickel, 1977).

In other words, the maximum likelihood estimators satisfy

$$(\hat{\lambda} - \lambda) \sim \mathrm{N}\left(0, \mathrm{I}^{-1}(\lambda)\right).$$

This implies that the maximum likelihood estimators are asymptotically optimal (Van der Vaart, 1998).

6.5 Tests of independence

The null hypothesis of 'no aggregation within a cluster' is obtained by setting the relative disposition, $\alpha(\lambda)$, to one, or equivalently by setting the excess disposition due to aggregation, D(Z), to zero. The likelihood ratio test and the Wald's test will be presented in this section to test for the presence or absence of familial aggregation of a disease.

(a) The likelihood ratio test

Consider the problem of testing the null hypothesis $H_0 : D(Z) = 0$ versus the alternative $H_1 : D(Z) \neq 0$, where D(Z) is the excess disposition due to aggregation. A likelihood ratio (LR) test statistic

$$LR = -2\ln\left(\frac{L_0}{L_1}\right) \qquad (6.5.1)$$

may be used for testing the hypothesis, where L_1 is the maximised likelihood of the model in which all the parameters are estimated and L_0 is the maximised likelihood from which the parameters are omitted (or set at some value). The likelihood ratio test statistic is asymptotically distributed as a chi-square with s degrees of freedom, where s is the difference in the number of parameters fitted between the full model and the reduced model (Wilks, 1938). We reject the above hypothesis at the significance level α if

$$LR > \chi^2_{s,1-\alpha}. \qquad (6.5.2)$$

(b) The Wald's test

Let $\lambda = (\xi, \gamma, \beta)$ be the vector of parameters, γ being the parameters of interest and ξ and β nuisance parameters. Suppose that we wish to test the null hypothesis $H_0 : \gamma = 0$ against the alternative $H_1 : \gamma \neq 0$, where $\gamma^T = (\gamma_0, ..., \gamma_q)$ is the vector of parameters that characterise dependence. Further, let $I_{\gamma\gamma}(\hat{\gamma}, \hat{\gamma})$ denote the ((q+1)x(q+1)) sub-matrix of the information matrix $I(\lambda)$ corresponding to γ. The quadratic form

$$W = (\hat{\gamma} - 0)^T I_{\gamma\gamma}(\hat{\gamma}, \hat{\gamma})(\hat{\gamma} - 0) \tag{6.5.3}$$

provides a Wald statistic, where $\hat{\gamma}$ is a set of maximum likelihood estimators of the vector parameter γ. The test statistic has an asymptotic chi-square distribution with (q+1) degrees of freedom, if the null hypothesis is true (Garthwaite, 1995). This is equivalent to stating that $(\hat{\gamma} - 0)$ has an asymptotic multivariate normal distribution with mean 0 and variance-covariance matrix $I_{\gamma\gamma}^{-1}(\hat{\gamma}, \hat{\gamma})$. The hypothesis is rejected for large values of W.

For a single parameter, γ_j, say, the standard normal test statistic

$$Z_j = \frac{\hat{\gamma}_j}{\sqrt{\mathrm{Var}[\hat{\gamma}_j]}} \tag{6.5.4}$$

is used. Z_j has an asymptotic N(0,1) distribution, if the null hypothesis that $\gamma_j = 0$ is true (Kleinbaum, 1994). The hypothesis is rejected at significance level α if

$$|Z_j| > u_{1-\alpha/2}. \tag{6.5.5}$$

6.6 Comparison of model fit

The Akaike's information criterion (Akaike, 1974) is a tool used for the comparison of competing models. It is defined as

$$\mathrm{AIC}(\hat{\lambda}) = (-2)\log(\text{maximum likelihood}) + 2k, \qquad (6.6.1)$$

where $\hat{\lambda}$ is the maximum likelihood estimate of the vector of parameters, λ, $\mathrm{AIC}(\hat{\lambda})$ is an estimate of a measure of fit of the model, and k is the number of estimated parameters in the model.

The more parameters a model contains, the less accurately they can be estimated. Thus, the term 2k adjusts for the increase of the variability of the estimates when the number of parameters in the model is increased. When there are several models, the model with the minimum AIC gives the best fit to the data.

The justification of the use of the maximum likelihood as a criterion of "fit" of a model is that its estimates are, under certain regularity conditions, asymptotically efficient. Thus the likelihood function tends to be a quantity which is most sensitive to the deviations of the model parameters from the true values (Akaike, 1974).

7 Application to real data sets

In this chapter, two examples will be used to illustrate the methods described in Chapters 4 and 5. The main objective of these analyses is to assess the presence of familial aggregation of diseases. In Section 7.1, details of the two data sets will be reported. Section 7.2 gives results of the oesophageal cancer data. The results of the breast cancer data will be given in Section 7.3.

7.1 Description of the data sets

Data were collected in the Yangcheng County, Shanxi Province, the Peoples Republic of China, designed to assess the presence of familial aggregation of oesophageal cancer. There were 2951 clusters (families), parents and siblings forming two subgroups of individuals. Cluster sizes were distributed as follows:

Cluster size	3	4	5	6	7	8	9	10	11	12	13
Number of clusters	623	819	659	412	232	129	43	23	8	2	1

The independent variables were smoking status, alcohol, age, sib size (sibsize) and mean sib age (meansibage). There were no group-specific covariates. The subgroup-specific covariates consisted of sibsize and meansibage, and the individual-specific covariates consisted of smoking status, alcohol and age. Smoking status was coded as 0 for non-smokers and 1 for smokers, alcohol was coded as 0 for non-drinkers and 1 for drinkers, and age was measured in years. The response variable Y was coded as 0 for unaffected and 1 for affected.

In the second example, data are available on 240 families with breast cancer in the national database and at the Howard University, Washington, D. C., U.S.A.. The data set comprises family data and epidemiology data. The variables to be assessed are annual household income (hinc), age at time of examination (ageat), obesity, and tumour of the breast other than breast cancer (tumour). Family-specific data consist of hinc in thousands

(<5, 5-15, 15-25, 25-35, 35-50, 50+), whereas subject-specific data consist of ageat in years and obesity (0 - not obese ; 1 - obese), and unit or breast-specific data consist of tumour (0 – absence; 1 - presence). The response variable indicates whether or not a breast is affected with breast cancer. This is coded as 0 for unaffected and 1 for affected. Two levels of nesting exist in these data: two breasts are nested within each subject and subjects are nested within families (compare with the second example of Qaqish and Liang, 1992). The objective of the analysis is to assess the presence of familial aggregation of breast cancer.

7.2 Analysis of the oesophageal cancer data

7.2.1 Descriptive analysis

The oesophageal cancer data consists of the variables age, alcohol, smoking status, sib size (sibsize) and mean sib age (meansibage). The results from the data analysis are summarised in Appendices E.1 and F.1. The sample size in the study is 14,310, consisting of 12,260 unaffected people and 2,050 affected people (see Table E.1.3). The ages of individuals in this study lie between 1 and 136. The mean age is 48.26 with standard deviation 18.18 (see Table E.1.1). Figure F.1.1 shows the distribution of age for the various age groups. The shaded area represents the proportion of affected people within the age groups. People within the ages 50 and 60 have the most affected cases of oesophageal cancer. It appears that the distribution of affected people (the shaded area) is well approximated by a normal distribution with mean age 57.61 and standard deviation 9.47 (see Table E.1.2 and Figure F.1.4). This is contrary to the case of the variable meansibage in Figure F.1.2, where the distribution of affected people (the shaded region) has a long left tail. This observation is confirmed by the normal Q-Q plot in Figure F.1.5. In Figure F.1.5, one clearly observes the deviation from a normal distribution, whereas the points on the plot in Figure F.1.4 form approximately a linear pattern. In other words, the distribution of age approximately matches the theoretical distribution.

The bar chart in Figure F.1.3 displays the distribution of sib sizes in the study. Families with two or three sibs have the highest frequency and families with eleven sibs have the lowest frequency. The mean sib size in the study is 3.37 (see Table E.1.1). Table E.1.3 gives descriptive statistics of the categorical data (i.e., the variables smoking status, alcohol and status).

7.2.2 Model for the non-nested disposition model

In this subsection, we shall determine the correlated logistic and the correlated Weibull regression models. We shall also compare the model fit of the two regression models.

We note that there are no group-specific covariates in the data set. Therefore, the cluster logit mean risk, $M_0(Z_0)$, and the excess cluster logit disposition due to dependence among members of a group, $D_0(Z_0)$, become $M_0(Z_0) = \xi_{00}$ and $D_0(Z_0) = \gamma_{00}$, respectively (see Equations (4.11) and (4.12)). We also note that in the non-nested case aggregations in sub-groups are not considered. The only variables in the model are therefore the individual-specific covariates: smoking status (X_1), alcohol (X_2) and age (X_3). Thus, the function that describes the effects of the individual-specific covariates, $X^T = (X_1, X_2, X_3)$, becomes $W_j(X_j) = 1 - \exp(\beta_1 X_{j1} + \beta_2 X_{j2} + \beta_3 X_{j3})$, j = 1,…,n, for the correlated Weibull regression model and $W_j(X_j) = \beta_1 X_{j1} + \beta_2 X_{j2} + \beta_3 X_{j3}$, j = 1,…,n, for the correlated logistic regression model, for the jth individual. The set of parameters to be determined is therefore
$\lambda = (\xi, \gamma, \beta) = (\xi_{00}, \gamma_{00}, \beta_1, \beta_2, \beta_3)$.

Table 7.2.2.1 presents the results of the correlated Weibull regression model (left panel) and of the correlated logistic regression model (right panel). The table shows regression parameter estimates, standard deviations of the parameter estimates and Wald statistics for determining whether the parameters in the model are needed.

We note that as opposed to the correlated logistic regression model, where a positive value of the coefficient of the individual-specific covariate indicates increased probability for a disease, a negative value of the coefficient of the individual-specific covariate is indicative of increased probability for a disease for the correlated Weibull regression model. For both models, a positive value of the coefficient of the group-specific covariate increases the probability for a disease. For example, the negative coefficient of age in the correlated Weibull regression model indicates that age increases the probability for oesophageal cancer. All the coefficients in Table 7.2.2.1 are statistically significant in both the correlated Weibull regression model and the correlated logistic regression model, when compared to the (1-$\alpha/2$)th quantile of the standard normal distribution at an α-level of 0.05 (i.e., $u_{0.975} = 1.96$).

Table 7.2.2.1: Parameter estimates, standard deviations and Wald statistics using the correlated Weibull and the correlated logistic regression models

Variable	Parameter	Correlated Weibull regression model			Correlated logistic regression model		
		Parameter estimate	Standard deviation	Wald statistic	Parameter estimate	Standard deviation	Wald statistic
constant	ξ_{00}	-2.4630	0.0387	63.6434	-3.7617	0.0934	40.2752
constant	γ_{00}	0.1272	0.0319	3.9875	0.0510	0.0250	2.0400
smoking	β_1	-0.6657	0.2673	2.4905	0.5006	0.0597	8.3853
alcohol	β_2	2.1720	0.2581	8.4153	-1.1208	0.1701	6.5891
age	β_3	-0.0262	0.0027	9.7037	0.0364	0.0016	22.7500

Critical value for the rejection of the null hypothesis: $u_{0.975} = 1.96$.

To test the hypothesis of 'no aggregation of oesophageal cancer in a cluster', we test the hypothesis that $D_0(Z_0) = 0$, or more specifically, $\gamma_{00} = 0$. We do this by performing the likelihood ratio test and the Wald's test.

For the correlated Weibull regression model, the log likelihood under the null hypothesis is $\log L_0 = -5673.0479$ and the log likelihood based on the full data is $\log L_1 = -5665.1874$. The likelihood ratio test statistic is therefore LR_W = -2[-5673.0479-(-5665.1874)] = 15.7210, which is significant when compared to a chi-square distribution with one degree of freedom $(\text{i.e.,} \chi_1^2 = 3.8415)$. For the correlated logistic regression model, the corresponding values are $\log L_0 = -5494.8614$ and $\log L_1 = -5492.7594$. The likelihood ratio test statistic is therefore LR_L = -2[-5494.8614 – (-5492.7594)] = 4.2040, which is also significant (see Section 6.5).

We now perform the Wald's tests. In Table 7.2.2.1, the value of γ_{00} is 0.1272 for the correlated Weibull regression model. The value of the Wald statistic is $Z_W = 3.9875$, and the critical value is $u_{0.975} = 1.96$. Because $Z_W > u_{0.975}$, the null hypothesis will be rejected (see Section 6.5). The conclusion is that there is evidence of familial aggregation of oesophageal cancer. For the correlated logistic regression model, the Wald statistic is $Z_L = 2.0400$. Since the Wald statistic is large, the null hypothesis will be rejected, indicating that there is significant aggregation of oesophageal cancer in the families.

We finally compare the model fit of the correlated Weibull regression model with that of the correlated logistic regression model using the Akaike's Information Criterion (AIC) (Akaike, 1974).

The AIC of the correlated Weibull regression model is
$AIC_W = -2\log L_1 +$ 2(number of estimated parameters) = 11330.3748 + 10 = 11340.3748,

and that of the correlated logistic regression model is
$AIC_L = -2\log L_1 +$ 2(number of estimated parameters) = 10985.5187 + 10 = 10995.5187.

The correlated logistic regression model has minimum AIC, and therefore fits the oesophageal cancer data better (see Section 6.6).

7.2.3 Model for the first level nesting

Since there are no group-specific covariates in the data set, the cluster logit mean risk, $M_0(Z_0)$, and the excess cluster logit disposition due to dependence among members of a group, $D_0(Z_0)$, become $M_0(Z_0) = \xi_{00}$ and $D_0(Z_0) = \gamma_{00}$, respectively (see Equations (5.1.14) and (5.1.15)). Two subgroups are nested within each family: parents form the first subgroup (i.e., i = 1) and siblings the second (i.e., i = 2). No variables are available for subgroup 1. The variables for subgroup 2 are sibsize and meansibage. Therefore, the excess on the logit scale of the mean risk in group 2 above that due to the cluster disposition, $M_2(Z_2)$, and the excess on the logit scale of the disposition within group 2 that cannot be explained by the overall cluster disposition and differences in baseline disposition under no aggregation in the group, $D_2(Z_2)$, become $M_2(Z_2) = \xi_1 Z_{21} + \xi_2 Z_{22}$
and $D_2(Z_2) = \gamma_1 Z_{21} + \gamma_2 Z_{22}$, respectively (see Equations (5.1.16) and (5.1.17)).

The individual-specific covariates are smoking status (X_1), alcohol (X_2) and age (X_3). Thus, the function that describes the effects of the individual-specific covariates becomes $W_{ij}(X_{ij}) = 1 - \exp(\beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3})$, i = 1,…,m, j = 1,…, n_i, for the correlated Weibull regression model and $W_{ij}(X_{ij}) = \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3}$, i = 1,…,m, j = 1,…, n_i, for the correlated logistic regression model, for the jth individual in group i. The set of parameters to be estimated is therefore $\lambda = (\xi, \gamma, \beta) = (\xi_{00}, \xi_1, \xi_2, \gamma_{00}, \gamma_1, \gamma_2, \beta_1, \beta_2, \beta_3)$.

Table 7.2.3.1 provides analysis of the oesophageal cancer data. The table gives maximum likelihood estimates, standard deviations and Wald statistics for the correlated Weibull regression model (left panel) and the correlated logistic regression model (right panel).

Table 7.2.3.1: Parameter estimates, standard deviations and Wald statistics using the correlated Weibull and the correlated logistic regression models

Variable	Parameter	Correlated Weibull regression model			Correlated logistic regression model		
		Parameter estimate	Standard deviation	Wald statistic	Parameter estimate	Standard deviation	Wald statistic
constant	ξ_{00}	-4.4426	0.1154	38.4974	-4.6036	0.1243	37.0362
sibsize	ξ_1	0.0172	0.0146	1.1781	0.0183	0.0152	1.2039
meansibage	ξ_2	0.0412	0.0019	21.6842	0.0365	0.0021	17.3810
constant	γ_{00}	-0.0965	0.0342	2.8216	-0.1042	0.0342	3.0468
sibsize	γ_1	-0.0117	0.0149	0.7852	-0.0179	0.0123	1.4553
meansibage	γ_2	0.0077	0.0015	5.1333	0.0081	0.0013	6.2308
smoking	β_1	-1.2751	0.3082	4.1372	0.5812	0.0654	8.8869
alcohol	β_2	2.2346	0.3157	7.0782	-0.9633	0.1768	5.4485
age	β_3	-0.0247	0.0046	5.3696	0.0191	0.0020	9.5500

Critical value for the rejection of the null hypothesis: $u_{0.975} = 1.96$.

The negative coefficient of age in the correlated Weibull regression model indicates that age increases the probability for oesophageal cancer. With the exception of ξ_1 and γ_1, all the coefficients of both regression models are statistically significant.

The hypotheses to be tested are $H_0 : \gamma = 0$ and $H_1 : \gamma \neq 0$. The following critical values will be used in this subsection for the rejection of the null hypothesis: $u_{0.975} = 1.96$ for the 1-parameter Wald's test and $\chi^2_{3,0.95} = 7.8147$ for the global tests.

The Wald's test rejects the null hypotheses $'\gamma_{00} = 0'$ and $'\gamma_2 = 0'$ of both the correlated Weibull regression model and the correlated logistic regression model, since the test statistics are large. The conclusion is that there is significant aggregation of oesophageal cancer in families and in siblings. It follows that the meansibage affects the familial aggregation of oesophageal cancer. On the other hand, the null hypothesis $'\gamma_1 = 0'$ of both disposition models cannot be rejected, since the test statistics are small. Hence, the sibsize does not affect the familial aggregation of oesophageal cancer.

For the correlated Weibull regression model, the maximised log likelihood from which γ is omitted is $\log L_0 = -5361.6679$, and the full log likelihood is $\log L_1 = -5323.3685$. The likelihood ratio statistic is therefore LR_W = -2[-5361.6679 – (-5323.3685)] = 76.5988.

For the correlated logistic regression model, the corresponding values are $\log L_0 = -5353.7628$ and $\log L_1 = -5309.0410$. The likelihood ratio statistic is therefore LR_L = -2[-5353.7628 – (-5309.0410)] = 89.4436. Thus, for both disposition models, significant familial aggregation is observed (see Section 6.5).

The maximum likelihood estimate of the vector of parameters that characterise dependence, $\gamma^T = (\gamma_{00}, \gamma_1, \gamma_2)$, is the same for both disposition models. The maximum likelihood estimate of γ is $\hat{\gamma} = \begin{pmatrix} 3.7424\text{x}10^{-1} \\ 5.5510\text{x}10^{-2} \\ 1.1447\text{x}10^{-3} \end{pmatrix}$ and the estimated variance-covariance matrix is

$$\hat{\text{var}}(\hat{\gamma}) = \begin{pmatrix} 0.000178 & -0.0000107 & 0.0000003 \\ -0.0000107 & 0.0000891 & -0.0000074 \\ 0.0000003 & -0.0000074 & 0.0000007 \end{pmatrix}.$$

The corresponding Wald statistic therefore has a value of $W = (\hat{\gamma} - 0)^T I_{\gamma\gamma}(\hat{\gamma}\hat{\gamma})(\hat{\gamma} - 0) = \hat{\gamma}^T[\hat{\text{var}}(\hat{\gamma})]^{-1}\hat{\gamma} = 1275.7093$, which is significant (see Section 6.5; Bickel and Doksum, 1977). Thus, the null hypothesis of no familial aggregation of oesophageal cancer can be rejected at the level $\alpha = 0.05$.

The AIC of the correlated Weibull regression model is $AIC_W = 10664.7370$ and that of the correlated logistic regression model is $AIC_L = 10636.0820$. The correlated logistic regression model minimises the AIC, and is therefore considered to be the more appropriate model.

7.3 Analysis of the breast cancer data

7.3.1 Descriptive analysis

The breast cancer data consists of the variables household income (hinc), obesity, age at time of examination (ageat), tumour and the response (bca). The nature of the data set necessitated the omission of the missing values before the descriptive analysis. Out of a total of 1198 observations, only 510 (42.57%) had no missing values. Table E.2.1 gives the frequencies of the variables hinc, obesity, tumour and bca for individuals in the study. There were altogether 510 individuals, consisting of 1020 breasts. The simple descriptive statistics of the variable ageat is given in Table E.2.2. This variable gives the ages at time of examination. If affected by breast cancer, it is evaluated as

ageat = year of diagnosis – year of birth,

otherwise it is evaluated as

ageat = date of interview – year of birth.

The average age at time of examination is 52.23, and the range of age at time of examination is 19 to 87. Figure F.2.1 presents a histogram of the variable ageat. The variable hinc has six levels having values 1, 2, 3, 4, 5 and 6. From the bar chart in Figure F.2.2, it can be seen that the fifth level of the variable hinc has the highest frequency. Families in the lowest income group have the lowest frequency. This corresponds to the first level of hinc. Table E.2.3 gives the distribution of breasts affected within subjects: 275 subjects have neither breasts affected, 110 have the right breasts affected, 121 have the left breasts affected and 4 have both breasts affected. The table also gives the variation of breast cancer side with respect to obesity. For instance, subjects whose right breasts are affected with breast cancer (coded 1) and also have obesity (coded 1) are 35 in number. Subjects whose left breasts are affected with breast cancer but do not have obesity are 85 in number. The variation of breast cancer side with respect to annual household income and age at time of examination can be seen in Tables E.2.4 and E.2.5, respectively. The pie chart in Figure F.2.3 shows the distribution of the breast cancer side.

7.3.2 Model for the second level nesting

The variables in the model are hinc (Z_{01}), ageat (Z_{ij1}), obesity (Z_{ij2}) and tumour (X_{ijh1}). That is, we have one group-specific covariate, two subject-specific covariates and one unit-specific covariate. There are no subgroup-specific covariates. The linear models in Equations (5.2.18) – (5.2.23) can therefore be specified as follows:

$M_0(Z_0) = \xi_{00} + \xi_{01}Z_{01}$, $D_0(Z_0) = \gamma_{00} + \gamma_{01}Z_{01}$, $M_i(Z_i) = 0$, $D_i(Z_i) = 0$,
$M_{ij}(Z_{ij}) = \xi_{11}Z_{ij1} + \xi_{12}Z_{ij2}$, and $D_{ij}(Z_{ij}) = \gamma_{11}Z_{ij1} + \gamma_{12}Z_{ij2}$.

For the function that describes the effects of the unit-specific covariate, we have
$W_{ijh}(X_{ijh}) = 1 - \exp(\beta X_{ijh1})$ for the correlated Weibull regression model and
$W_{ijh}(X_{ijh}) = \beta X_{ijh1}$ for the correlated logistic regression model.

The set of parameters to be determined in the model is

$$\lambda = (\xi, \gamma, \beta) = (\xi_{00}, \xi_{01}, \xi_{11}, \xi_{12}, \gamma_{00}, \gamma_{01}, \gamma_{11}, \gamma_{12}, \beta).$$

Parameter estimates and standard deviations of the estimates, along with Wald statistics are given in Table 7.3.2.1 for the correlated Weibull and the correlated logistic regression models. The function of the individual-specific covariates, $W_{ijh}(X_{ijh})$, is equal to zero, since no breast has a primary tumour other than breast cancer. Hence, the estimates for both regression models are the same. The parameter β is fixed for computational reasons. The covariates of positive (negative) coefficients increase (decrease) the probability for breast cancer.

For the 1-parameter Wald's tests, the null hypothesis that $\gamma_j = 0$ is rejected for γ_{00}, γ_{01} and γ_{11}. This is an indication of the existence of familial aggregation of breast cancer. On the other hand, the null hypothesis of $\gamma_j = 0$ for γ_{12} cannot be rejected at the level $\alpha = 0.05$. Hence, obesity does not affect the familial aggregation of breast cancer.

Table 7.3.2.1: Parameter estimates, standard deviations of the estimates and Wald statistics for the correlated Weibull and the correlated logistic regression models

Variable	Parameter	Correlated Weibull and logistic regression models		
		Parameter estimate	Standard deviation	Wald statistic
constant	ξ_{00}	-3.9134	0.2759	14.1841
hinc	ξ_{01}	0.2791	0.0189	14.7672
ageat	ξ_{11}	0.0250	0.0034	7.3529
obesity	ξ_{12}	0.1275	0.1118	1.1404
constant	γ_{00}	-0.6350	0.1530	4.1503
hinc	γ_{01}	-0.0477	0.0105	4.5429
ageat	γ_{11}	-0.0290	0.0131	2.2137
obesity	γ_{12}	0.7268	0.9904	0.7338
tumour	β	---	---	---

Critical value for the rejection of the null hypothesis: $u_{0.975} = 1.96$.

For the global tests, the hypotheses to be tested are $H_0: \gamma = 0$ and $H_1: \gamma \neq 0$, where $\gamma^T = (\gamma_{00}, \gamma_{01}, \gamma_{11}, \gamma_{12})$.

Let $\log L_0$ = the maximised log-likelihood from which γ is omitted,

$\log L_1$ = the full log-likelihood and

$\chi^2_4 = 9.4877$ (i.e., the critical value for the rejection of the null hypothesis).

Then, the likelihood ratio statistic for the correlated Weibull and the correlated logistic regression models is LR = -2[$\log L_0 - \log L_1$] = -2[-536.1829 – (-466.6963)] = 138.9732. Thus, significant familial aggregation is observed for both regression models.

The maximum likelihood estimate of γ is $\hat{\gamma} = \begin{pmatrix} 6.2049\text{x}10^{-2} \\ -2.1115\text{x}10^{-2} \\ -4.5258\text{x}10^{-3} \\ 8.1942\text{x}10^{-1} \end{pmatrix}$ and the estimated variance-covariance matrix is

$$\hat{\text{var}}(\hat{\gamma}) = \begin{pmatrix} 4.31\text{x}10^{-1} & -1.11\text{x}10^{-4} & -4.72\text{x}10^{-7} & 6.86\text{x}10^{-4} \\ -1.11\text{x}10^{-4} & 3.13\text{x}10^{-5} & 6.86\text{x}10^{-4} & -2.50\text{x}10^{-4} \\ -4.72\text{x}10^{-7} & 6.86\text{x}10^{-4} & 3.95\text{x}10^{-6} & -1.77\text{x}10^{-4} \\ 6.86\text{x}10^{-4} & -2.50\text{x}10^{-4} & -1.77\text{x}10^{-4} & 3.08\text{x}10^{-2} \end{pmatrix}.$$

The Wald statistic for $H_0 : \gamma = 0$ has a value of $W = (\hat{\gamma} - 0)^T I_{\gamma\gamma}(\hat{\gamma}, \hat{\gamma})(\hat{\gamma} - 0) = \hat{\gamma}^T [\hat{\text{var}}(\hat{\gamma})]^{-1} \hat{\gamma} = 32.2734$. Since the Wald statistic is large, the null hypothesis will be rejected (see Section 6.5; Bickel and Doksum, 1977). The conclusion is that there is aggregation of breast cancer in families.

Table 7.3.2.2 presents estimates of the parameters obtained by fitting Cox's model, with standard deviations and Wald statistics for testing effects.

Table 7.3.2.2: Parameter estimates, standard deviations of the estimates and Wald statistics resulting from Cox's model

Variable	Parameter	Parameter estimate	Standard deviation	Wald statistic
hinc	β_1	-0.0159	0.0196	0.8112
ageat	β_2	-0.0050	0.0020	2.5000
obesity	β_3	0.0441	0.0652	0.6764
tumour	β_4	---	---	---

Critical value for the rejection of the null hypothesis: $u_{0.975} = 1.96$.

The hypothesis to be tested is $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$. From Table 7.3.2.2, the covariate ageat is the only significant factor. The covariates hinc and obesity produce nonsignificant effects, since the values of their Wald statistics are less than the critical value, $u_{0.975} = 1.96$.

For the global null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$, we obtain 6.9773 for the likelihood ratio statistic and 6.9874 for the Wald statistic, both values indicating non-significance when compared to a chi-square distribution with three degrees of freedom (i.e., $\chi^2_{3,0.95} = 7.8147$).

8 Discussion

The correlated Weibull and logistic regression models for correlated binary data have been presented. The objective of the analyses has been to assess familial aggregation of diseases. In Section 7.2, the model fit of the correlated Weibull regression model was compared to that of the correlated logistic regression model using the Akaike Information Criterion (AIC). The model that minimised the AIC was considered to give a better fit to the oesophageal cancer data. The correlated logistic regression model fitted the data better than the correlated Weibull regression model for both the non-nested and nested cases. On the whole, the correlated logistic regression model was computationally more feasible than the correlated Weibull regression model. The data processing was done using the SAS programming language, and computations were made in the C programming language.

The problems associated with estimation as the level of nesting gets deeper have also been investigated and the performance of the nested disposition model compared with Cox's model (Cox, 1972). The main disadvantage of the disposition model is that, with the exception of the unit-specific covariates, each covariate in the model produces two parameters. This results in the following problems:

(1) The effect of a covariate can have different interpretations. For instance, in Table 7.3.2.1 the covariate hinc increases the cluster logit mean risk, $M_0(Z_0) = \xi_{00} + \xi_{01}Z_{01}$, whereas the same covariate decreases the excess cluster logit disposition due to dependence among members of the group, $D_0(Z_0) = \gamma_{00} + \gamma_{01}Z_{01}$. Thus, the same variable hinc gives two opposing effects with regard to the probability for breast cancer,

$$\delta_{ijh} = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_{ij}(Z_{ij}) + D_{ij}(Z_{ij})]\}}.$$

(2) The number of covariates that can be included in the model is seriously limited. In Section 7.2, we could estimate up to nine parameters from five covariates, using the maximum likelihood method. An attempt to estimate more than nine parameters from five covariates (the fifth covariate finally excluded from the analysis) in Section 7.3 resulted in over-identified parameters (i.e., parameters estimated in two or more linearly independent ways).

The disposition model however has the advantage that aggregations in families, due to common shared risks, and response probabilities can jointly be modelled.

The problem of missing values is a point worth mentioning. If the size of the data set is large with few missing values, failure to omit the missing values has very little effect on the estimates of the parameters and their standard deviations. On the other hand, if the data set is small with many missing values as in the case of the breast cancer data, failure to omit the missing values leads to erroneous estimates of the parameters. There are, of course, methods that can be used to impute values for the missing data. In this dissertation, the missing values were omitted in accordance with the conventional approach in epidemiology (see Thomas and Gauderman, 1995).

APPENDIX

Appendix A: Symbols

Symbol	Meaning
D(Z)	the excess cluster logit disposition due to dependence among members
E(U)	the expectation of U
H_0	null hypothesis
H_1	alternative hypothesis
H_h	the Hessian matrix for the h-th unit
H_i	the Hessian matrix for the ith subgroup
H_j	the Hessian matrix for the jth individual
H_k	the Hessian matrix for the kth cluster
I_h	the Fisher information matrix for the h-th unit
I_i	the Fisher information matrix for the ith subgroup
I_j	the Fisher information matrix for the jth individual
I_k	the Fisher information matrix for the kth cluster
L_0	the maximised likelihood of the reduced model
L_1	the maximised likelihood of the full model
L_i	the likelihood function for the ith subgroup
L_k	the likelihood function for the kth cluster
LR	Likelihood ratio
m	the number of subgroups in a cluster
M(Z)	the cluster logit mean risk
n_i	the number of individuals in subgroup i
$P(Y_j)$	the response probability of the jth individual
U^T	the transpose of U
U^*	the value of U evaluated at $y = (y_1,...,y_n) = 0$
U_h	the score function for the h-th unit

U_i	the score function for the ith subgroup
U_j	the score function for the jth individual
U_k	the score function for the kth cluster
Var(U)	the variance of U
W(X)	the function of the individual-specifc covariates
$X_1,...,X_p$	p individual-specific covariates
$Y^T = (Y_1,...,Y_n)$	a vector of n binary outcomes
$Z_1,...,Z_p$	q group-specific covariate vector
α_0	the relative disposition with respect to a cluster
α_i	the relative disposition with respect to subgroup i
α_{ij}	the relative disposition with respect to the tertiary group
β	the parameters from W(X)
ξ	the parameters from M(Z)
γ	the parameters from D(Z)
δ_i	the subgroup disposition (i.e., secondary group disposition)
δ_{ij}	the tertiary group disposition
δ_{ijh}	unit disposition
δ_0	the group or cluster disposition
$\lambda = (\xi,\gamma,\beta)$	the set of parameters to be determined in the model
$\theta(\lambda)$	the sum of M(Z), D(Z) and W(X)
μ_i	the subgroup baseline disposition under no aggregation
μ_{ij}	the tertiary group baseline disposition under no aggregation
μ_0	the group baseline disposition under no aggregation

Appendix B: Definitions

B1: Mood, Graybill and Boes (1974, p. 296):

A sequence of estimators $T_1^*,\ldots,T_n^*,\ldots$ of $\tau(\theta)$ is defined to be best asymptotically normal (BAN) if and only if the following conditions are satisfied:

(i) The distribution of $\sqrt{n}[T_n^* - \tau(\theta)]$ approaches the normal distribution with mean 0 and variance $\sigma^{*2}(\theta)$ as n approaches infinity.

(ii) For every $\varepsilon > 0$,

$$\lim_{n\to\infty} P_\theta[|\,T_n^* - \tau(\theta)\,| > \varepsilon] = 0 \text{ for each } \theta \text{ in } \Theta$$

(i.e., T_n^* is consistent for each θ in Θ).

(iii) Let $\{T_n\}$ be any other sequence of simple consistent estimators for which the distribution of $\sqrt{n}[T_n - \tau(\theta)]$ approaches the normal distribution with mean 0 and variance $\sigma^2(\theta)$.

(iv) $\sigma^2(\theta)$ is not less than $\sigma^{*2}(\theta)$ for all θ in any open interval (i.e., T_n^* is efficient for all θ in any open interval).

B2: Mood, Graybill and Boes (1974, p. 315-316):

Let $X_1,\ldots,X_n$ be a random sample from $f(\cdot;\theta)$, where θ belongs to Θ. Assume that Θ is a subset of the real line. Let $T = t(X_1,\ldots,X_n)$ be an unbiased estimator of $\tau(\theta)$. Then, the regularity conditions are as follows:

(i) $\frac{\delta}{\delta\theta}\log f(x;\theta)$ exists for all x and all θ

(i.e., the existence of certain partial derivatives).

(ii) $$\frac{\delta}{\delta\theta}\int\ldots\int\prod_{i=1}^{n} f(x_i;\theta)dx_1\ldots dx_n = \int\ldots\int\frac{\delta}{\delta\theta}\prod_{i=1}^{n} f(x_i;\theta)dx_1\ldots dx_n$$

(i.e., exchange of differentiation and integration).

(iii) $$\frac{\delta}{\delta\theta}\int\ldots\int t(x_1,\ldots,x_n)\prod_{i=1}^{n} f(x_i;\theta)dx_1\ldots dx = \int\ldots\int t(x_1,\ldots,x_n)\frac{\delta}{\delta\theta}\prod_{i=1}^{n} f(x_i;\theta)dx_1\ldots dx_n$$

(i.e., exchange of differentiation and integration or differentiation and summation, as in Section 6.4).

(iv) $$0 < \varepsilon_\theta\left[\left[\frac{\delta}{\delta\theta}\log f(X;\theta)\right]^2\right] < \infty \text{ for all } \theta \text{ in } \Theta$$

(i.e., the Fisher-Information of $f(x;\theta)$ with respect to θ lies between 0 and infinity).

Appendix C: Newton-Raphson method

The Newton-Raphson method is a basic numerical algorithm for finding approximations to the solutions of non-linear systems of equations. Suppose that $\ell(\lambda \mid y)$ is the log-likelihood function, which we wish to maximise. Let $U(\lambda|y)$ denote the vector of first derivatives of $\ell(\lambda \mid y)$ (i.e., the score function) and $H(\lambda)$ the matrix of second derivatives of $\ell(\lambda \mid y)$ (i.e., the Hessian matrix). Then, an estimate $\hat{\lambda}$ of λ can be obtained as a solution of the likelihood equation $U(\lambda|y) \equiv \frac{\delta}{\delta\lambda} \log L(\lambda|y) = 0$ by means of the algorithm

$$\lambda^{(t+1)} = \lambda^{(t)} + [-H^{-1}(\lambda^{(t)})]U(\lambda^{(t)} \mid y)$$

for t = 0, 1, 2, ..., where λ^{t} is the estimate at the t-th iteration (McLachlan and Krishnan, 1997).

The algorithm needs both an initial guess $\lambda^{(0)}$ (e.g., an estimate based on the completely observed observations) and a stopping criterion (e.g., the requirement of a small residual such as $|U(\lambda|y)| \leq \text{tolerance}$) (Kotz and Johnson, 1982).

By successive repetition of the above algorithm, using the result of one stage as the input for the next, convergence is achieved. Convergence may be slowed or prevented if the initial guess of $\lambda^{(0)}$ is inappropriate or two roots are close together or $H(\lambda^{t}) \to 0$ (Daintith and Nelson, 1989).

A variant of this procedure is the Method of Scoring, where the observed information, $-H(\lambda)$, is replaced by the expected information, $I(\lambda) = E[-H(\lambda)]$:

$$\lambda^{(t+1)} = \lambda^{(t)} + I^{-1}(\lambda^{(t)})U(\lambda^{(t)} \mid y)$$

(see, for example, Godambe, 1991).

Appendix D: Mathematical addendum

The purpose of this appendix is to provide detailed proofs of the partial derivatives of the relative dispositions $\alpha_0(\lambda)$, $\alpha_i(\lambda)$ and $\alpha_{ij}(\lambda)$ with respect to the unknown parameters. Equivalent forms of the Fisher information matrices in Equations (6.1.4) and (6.2.4) are also given.

D1: Define the relative disposition $\alpha_0(\lambda)$ by

$$\alpha_0(\lambda) = \frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)]\}}{1+\exp[-M_0(Z_0)]} \quad .$$

Then,

$$\alpha_0^*(\lambda) = \frac{\delta}{\delta\lambda}\log\alpha_0(\lambda)$$

$$= \frac{\delta}{\delta\lambda}\log\left\{\frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)]\}}{1+\exp[-M_0(Z_0)]}\right\}$$

$$= \frac{\delta}{\delta\lambda}\left\{\log\left[1+\exp(-(M_0(Z_0)+D_0(Z_0)))\right]-\log[1+\exp(-M_0(Z_0))]\right\}$$

$$= \frac{\frac{\delta}{\delta\lambda}\left[\exp(-(M_0(Z_0)+D(Z)))\right]}{\left[1+\exp(-(M_0(Z_0)+D_0(Z_0)))\right]} - \frac{\frac{\delta}{\delta\lambda}\left[\exp(-(M_0(Z_0)))\right]}{\left[1+\exp(-(M_0(Z_0)))\right]}$$

$$\underset{\delta_0=\frac{1}{1+\exp[-(M_0(Z_0)+D_0(Z_0))]}}{=} \delta_0\frac{\delta}{\delta\lambda}\left[\exp(-(M_0(Z_0)+D_0(Z_0)))\right] - \frac{\frac{\delta}{\delta\lambda}\left[\exp(-(M_0(Z_0)))\right]}{\left[1+\exp(-(M_0(Z_0)))\right]}$$

$$= \underbrace{\delta_0 \left[\exp(-(M_0(Z_0) + D_0(Z_0)))\right]}_{1-\delta_0} \frac{\delta}{\delta\lambda}[-(M_0(Z_0) + D_0(Z_0))]$$

$$- \frac{\exp(-(M_0(Z_0)))}{\left[1 + \exp(-(M_0(Z_0)))\right]} \frac{\delta}{\delta\lambda}\left[-(M_0(Z_0)\right]$$

$$= -(1-\delta_0)\frac{\delta}{\delta\lambda}D_0(Z_0) + \left(\frac{\exp(-(M_0(Z_0)))}{1+\exp(-(M_0(Z_0)))} - (1-\delta_0)\right)\frac{\delta}{\delta\lambda}M_0(Z_0)$$

$$= -(1-\delta_0)\frac{\delta}{\delta\lambda}D_0(Z_0) + \left(\delta_0 - \frac{1}{1+\exp(-(M_0(Z_0)))}\right)\frac{\delta}{\delta\lambda}M_0(Z_0)$$

$$= -(1-\delta_0)\frac{\delta}{\delta\lambda}D_0(Z_0) + \left(\delta_0 - \delta_0\alpha_0\right)\frac{\delta}{\delta\lambda}M_0(Z_0)$$

$$= -(1-\delta_0)\frac{\delta}{\delta\lambda}D_0(Z_0) + \delta_0\left(1 - \alpha_0\right)\frac{\delta}{\delta\lambda}M_0(Z_0)\,.$$

The above expression is a vector, whose components are obtained by the execution of the partial derivatives with respect to the parameters in the model.

D2: Let the relative disposition $\alpha_i(\lambda)$ be defined by

$$\alpha_i(\lambda) = \frac{1+\exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i)]\}}{1+\exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i)]\}}\quad .$$

Then,

$$\alpha_i^*(\lambda) = \frac{\delta}{\delta\lambda}\log\alpha_i(\lambda)$$

$$= \frac{\delta}{\delta\lambda}\log\left\{\frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)]\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}\right\}$$

$$= \frac{\delta}{\delta\lambda}\log\{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)]\}\}$$

$$- \frac{\delta}{\delta\lambda}\log\{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}\}$$

$$= \frac{\frac{\delta}{\delta\lambda}\{\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)]\}\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)]\}}$$

$$- \frac{\frac{\delta}{\delta\lambda}\{\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}$$

$$= \delta_i \frac{\delta}{\delta\lambda}\{\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)]\}\}$$

$$- \frac{\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}\frac{\delta}{\delta\lambda}\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}$$

$$= (1-\delta_i)\frac{\delta}{\delta\lambda}\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)]\}$$

$$- \frac{\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}\frac{\delta}{\delta\lambda}\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}$$

$$= -(1-\delta_i)\frac{\delta}{\delta\lambda}D_i(Z_i) + \left(\frac{\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}} - (1-\delta_i)\right)$$

$$* \frac{\delta}{\delta\lambda}[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]$$

$$= -(1-\delta_i)\frac{\delta}{\delta\lambda}D_i(Z_i)$$

$$+\left(\delta_i - \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]\}}\right)\frac{\delta}{\delta\lambda}[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]$$

$$= -(1-\delta_i)\frac{\delta}{\delta\lambda}D_i(Z_i) + \left(\delta_i - \delta_i\alpha_i\right)\frac{\delta}{\delta\lambda}[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)]$$

$$= -(1-\delta_i)\frac{\delta}{\delta\lambda}D_i(Z_i) + \delta_i\left(1-\alpha_i\right)\frac{\delta}{\delta\lambda}[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)].$$

D3: Let the relative disposition $\alpha_{ij}(\lambda)$ be defined by

$$\alpha_{ij}(\lambda) = \frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})+D_{ij}(Z_{ij})]\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})]\}} .$$

Then,

$$\alpha^*_{ij}(\lambda) = \frac{\delta}{\delta\lambda}\log\alpha_{ij}(\lambda)$$

$$= \frac{\delta}{\delta\lambda}\log\left\{\frac{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})+D_{ij}(Z_{ij})]\}}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})]\}}\right\}$$

$$= \frac{\delta}{\delta\lambda}\log\{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})+D_{ij}(Z_{ij})]\}\}$$

$$- \frac{\delta}{\delta\lambda}\log\{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})]\}\}$$

$$= \frac{\frac{\delta}{\delta\lambda}\{\exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij}) + D_{ij}(Z_{ij})]\}\}}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij}) + D_{ij}(Z_{ij})]\}}$$

$$- \frac{\frac{\delta}{\delta\lambda}\{\exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}\}}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}}$$

$$= \delta_{ij} \frac{\delta}{\delta\lambda}\{\exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij}) + D_{ij}(Z_{ij})]\}\}$$

$$- \frac{\exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}}$$

$$* \frac{\delta}{\delta\lambda}\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}$$

$$= (1 - \delta_{ij}) \frac{\delta}{\delta\lambda}\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij}) + D_{ij}(Z_{ij})]\}$$

$$- \frac{\exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}}$$

$$* \frac{\delta}{\delta\lambda}\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}$$

$$= -(1 - \delta_{ij}) \frac{\delta}{\delta\lambda} D_{ij}(Z_{ij})$$

$$+ \left(\frac{\exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}} - (1 - \delta_{ij}) \right)$$

$$* \frac{\delta}{\delta\lambda}\{[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + M_{ij}(Z_{ij})]\}$$

$$= -\ (1-\delta_{ij})\frac{\delta}{\delta\lambda}D_{ij}(Z_{ij})$$

$$+\left(\delta_{ij}-\frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})]\}}\right)$$

$$*\frac{\delta}{\delta\lambda}\{[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})]\}$$

$$= -\ (1-\delta_{ij})\frac{\delta}{\delta\lambda}D_{ij}(Z_{ij})$$

$$+\ (\delta_{ij}-\delta_{ij}\alpha_{ij})\frac{\delta}{\delta\lambda}\{[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})]\}$$

$$= -\ (1-\delta_{ij})\frac{\delta}{\delta\lambda}D_{ij}(Z_{ij})$$

$$+\ \delta_{ij}(1-\alpha_{ij})\frac{\delta}{\delta\lambda}\{[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+M_{ij}(Z_{ij})]\}\ .$$

D4: An equivalent form of the Fisher information matrix in Equation (6.1.4):

$$I_k(\lambda)=\begin{pmatrix} I_{\xi_0\xi_0} & I_{\xi_0\gamma_0} & I_{\xi_0\beta} \\ I_{\gamma_0\xi_0} & I_{\gamma_0\gamma_0} & I_{\gamma_0\beta} \\ I_{\beta\xi_0} & I_{\beta\gamma_0} & I_{\beta\beta} \end{pmatrix},$$

where the entries of $I_k(\lambda)$ are the following sub-matrices:

$$I_{\xi_0\xi_0} = \Big\{\alpha_0\sum\delta_j(1-\delta_j)-A_i^*\delta_0^2(1-\alpha_0)^2$$

$$-\ 2B_i^*(1-\alpha_0)\delta_0\sum\delta_j - B_i^*(1-\alpha_0)\sum_j\sum_{j'}\delta_j\delta_{j'}\Big\}\frac{\delta M_0}{\delta\xi_0}\left(\frac{\delta M_0}{\delta\xi_0}\right)^T,$$

$$I_{\xi_0\gamma_0} = I_{\gamma_0\xi_0} = \Big\{\alpha_0 \sum \delta_j(1-\delta_j) + A_i^*(1-\alpha_0)\delta_0(1-\delta_0)$$

$$- B_i^*[(1-\delta_0)\sum \delta_j - (1-\alpha_0)\delta_0 \sum \delta_j] - B_i^*(1-\alpha_0)\sum_j \sum_{j'} \delta_j \delta_{j'} \Big\} \frac{\delta M_0}{\delta \xi_0}\left(\frac{\delta D_0}{\delta \gamma_0}\right)^T,$$

$$I_{\xi_0\beta} = I_{\beta\xi_0} = \left\{\alpha_0 \sum \delta_j(1-\delta_j) + B_i^*(1-\alpha_0)\delta_0 \sum \delta_j - B_i^*(1-\alpha_0)\sum_j \sum_{j'} \delta_j \delta_j \right\} \frac{\delta M_0}{\delta \xi_0}\left(\frac{\delta W}{\delta \beta}\right)^T,$$

$$I_{\gamma_0\gamma_0} = \Big\{\alpha_0 \sum \delta_j(1-\delta_j) - A_i^*(1-\delta_0)^2$$

$$- 2B_i^*(1-\delta_0)\sum \delta_j - B_i^*(1-\alpha_0)\sum_j \sum_{j'} \delta_j \delta_{j'} \Big\} \frac{\delta D_0}{\delta \gamma_0}\left(\frac{\delta D_0}{\delta \gamma_0}\right)^T,$$

$$I_{\gamma_0\beta} = I_{\beta\gamma_0} = \left\{\alpha_0 \sum \delta_j(1-\delta_j) - B_i^*(1-\delta_0)\sum \delta_j - B_i^*(1-\alpha_0)\sum_j \sum_{j'} \delta_j \delta_{j'} \right\} \frac{\delta D_0}{\delta \gamma_0}\left(\frac{\delta W}{\delta \beta}\right)^T$$

and

$$I_{\beta\beta} = \left\{\Big\{\alpha_0 \sum \delta_j(1-\delta_j) - B_i^*(1-\alpha_0)\sum_j \sum_{j'} \delta_j \delta_{j'} \right\} \frac{\delta W}{\delta \beta}\left(\frac{\delta W}{\delta \beta}\right)^T.$$

D5: An equivalent form of the Fisher information matrix in Equation (6.2.4):

$$I_k(\lambda) = \begin{pmatrix} I_{\xi_0\xi_0} & I_{\xi_0\xi} & I_{\xi_0\gamma_0} & I_{\xi_0\gamma} & I_{\xi_0\beta} \\ I_{\xi\xi_0} & I_{\xi\xi} & I_{\xi\gamma_0} & I_{\xi\gamma} & I_{\xi\beta} \\ I_{\gamma_0\xi_0} & I_{\gamma_0\xi} & I_{\gamma_0\gamma_0} & I_{\gamma_0\gamma} & I_{\gamma_0\beta} \\ I_{\gamma\xi_0} & I_{\gamma\xi} & I_{\gamma\gamma_0} & I_{\gamma\gamma} & I_{\gamma\beta} \\ I_{\beta\xi_0} & I_{\beta\xi} & I_{\beta\gamma_0} & I_{\beta\gamma} & I_{\beta\beta} \end{pmatrix},$$

where the entries of $I_k(\lambda)$ are the following sub-matrices:

$$I_{\xi_0\xi_0} = \left\{\alpha_0 \sum_{i=1}^{m}\left\{\alpha_i\left[\sum_{j=1}^{n_i}\delta_{ij}(1-\delta_{ij})\right] - B^*_{\pi i}(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\} - A^*_k\delta_0^2(1-\alpha_0)^2\right.$$

$$- \left\{\alpha_0 \sum_{i=1}^{m}\delta_i(1-\alpha_i)\left[A^*_{\pi i}\delta_i(1-\alpha_i) + 2B^*_{\pi i}\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]\right]\right\}$$

$$- B^*_k\left\{2\delta_0(1-\alpha_0)\left[\left[\sum_{i=1}^{m}A^*_{\pi i}\delta_i(1-\alpha_i)\right] + \left[\sum_{i=1}^{m}B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right]\right\}$$

$$- B^*_k(1-\alpha_0)\left\{\left[\sum_{i=1}^{m}A^*_{\pi i}\delta_i(1-\alpha_i)\right]^2 + 2\left[\sum_{i=1}^{m}A^*_{\pi i}\delta_i(1-\alpha_i)\right]\left[\sum_{i=1}^{m}B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right.$$

$$\left.\left. + \left[\sum_{i=1}^{m}B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]^2\right\}\right\} \frac{\delta M_0}{\delta\xi_0}\left(\frac{\delta M_0}{\delta\xi_0}\right)^T,$$

$$I_{\xi_0\xi} = I_{\xi\xi_0} = \left\{\alpha_0 \sum_{i=1}^{m}\left\{\alpha_i\left[\sum_{j=1}^{n_i}\delta_{ij}(1-\delta_{ij})\right] - B^*_{\pi i}(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\}\right.$$

$$- \left\{\alpha_0 \sum_{i=1}^{m}\delta_i(1-\alpha_i)\left[A^*_{\pi i}\delta_i(1-\alpha_i) + 2B^*_{\pi i}\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]\right]\right\}$$

$$- B^*_k\left\{\delta_0(1-\alpha_0)\left[\left[\sum_{i=1}^{m}A^*_{\pi i}\delta_i(1-\alpha_i)\right] + \left[\sum_{i=1}^{m}B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right]\right\}$$

$$- B^*_k(1-\alpha_0)\left\{\left[\sum_{i=1}^{m}A^*_{\pi i}\delta_i(1-\alpha_i)\right]^2 + 2\left[\sum_{i=1}^{m}A^*_{\pi i}\delta_i(1-\alpha_i)\right]\left[\sum_{i=1}^{m}B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right.$$

$$+\left[\sum_{i=1}^{m} B_{\pi i}^{*}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]^2\Bigg\}\Bigg\}\ \frac{\delta M_0}{\delta\xi_0}\left(\frac{\delta M_i}{\delta\xi_i}\right)^T,$$

$$I_{\xi_0\gamma_0}=I_{\gamma_0\xi_0}=\left\{\alpha_0\sum_{i=1}^{m}\left\{\alpha_i\left[\sum_{j=1}^{n_i}\delta_{ij}(1-\delta_{ij})\right]-B_{\pi i}^{*}(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\}+A_k^{*}\delta_0(1-\delta_0)(1-\alpha_0)\right.$$

$$-\left\{\alpha_0\sum_{i=1}^{m}\delta_i(1-\alpha_i)\left[A_{\pi i}^{*}\delta_i(1-\alpha_i)+2B_{\pi i}^{*}\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]\right]\right\}$$

$$-B_k^{*}\left\{\left[\delta_0(1-\alpha_0)-(1-\delta_0)\right]\left[\left[\sum_{i=1}^{m}A_{\pi i}^{*}\delta_i(1-\alpha_i)\right]+\left[\sum_{i=1}^{m}B_{\pi i}^{*}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right]\right\}$$

$$-B_k^{*}(1-\alpha_0)\left\{\left[\sum_{i=1}^{m}A_{\pi i}^{*}\delta_i(1-\alpha_i)\right]^2+2\left[\sum_{i=1}^{m}A_{\pi i}^{*}\delta_i(1-\alpha_i)\right]\left[\sum_{i=1}^{m}B_{\pi i}^{*}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right.$$

$$+\left[\sum_{i=1}^{m}B_{\pi i}^{*}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]^2\Bigg\}\Bigg\}\ \frac{\delta M_0}{\delta\xi_0}\left(\frac{\delta D_0}{\delta\gamma_0}\right)^T,$$

$$I_{\xi_0\gamma}=I_{\gamma\xi_0}=\left\{\alpha_0\sum_{i=1}^{m}\left\{\alpha_i\left[\sum_{j=1}^{n_i}\delta_{ij}(1-\delta_{ij})\right]-B_{\pi i}^{*}(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\}+\alpha_0\sum_{i=1}^{m}A_{\pi i}^{*}\delta_i(1-\delta_i)(1-\alpha_i)\right.$$

$$-\alpha_0\sum_{i=1}^{m}B_{\pi i}^{*}\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]\left[\delta_i(1-\alpha_i)-(1-\delta_i)\right]$$

$$-B_k^{*}\left\{\delta_0(1-\alpha_0)\left[\sum_{i=1}^{m}B_{\pi i}^{*}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]-\delta_0(1-\alpha_0)\left[\left[\sum_{i=1}^{m}A_{\pi i}^{*}(1-\delta_i)\right]+\right]\right\}$$

$$- \ B_k^*(1-\alpha_0)\left\{-\left[\sum_{i=1}^{m} A_{\pi i}^*(1-\delta_i)\right]\left[\sum_{i=1}^{m} A_{\pi i}^*\delta_i(1-\alpha_i)\right]+\left[\sum_{i=1}^{m} B_{\pi i}^*\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]^2\right]\right.$$

$$\left.+\left[\sum_{i=1}^{m} B_{\pi i}^*\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\left[\left[\sum_{i=1}^{m} A_{\pi i}^*\delta_i(1-\alpha_i)\right]-\left[\sum_{i=1}^{m} A_{\pi i}^*(1-\delta_i)\right]\right]\right\}\right\} \frac{\delta M_0}{\delta \xi_0}\left(\frac{\delta D_i}{\delta \gamma_i}\right)^T,$$

$$I_{\xi_0\beta} = I_{\beta\xi_0} = \left\{\alpha_0\sum_{i=1}^{m}\left\{\alpha_i\left[\sum_{j=1}^{n_i}\delta_{ij}(1-\delta_{ij})\right]-B_{\pi i}^*(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\}\right.$$

$$- \ \alpha_0\sum_{i=1}^{m} B_{\pi i}^*\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]\delta_i(1-\alpha_i) \ - \ B_k^*\left\{\delta_0(1-\alpha_0)\left[\sum_{i=1}^{m} B_{\pi i}^*\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right\}$$

$$- \ B_k^*(1-\alpha_0)\left\{\left[\sum_{i=1}^{m} A_{\pi i}^*\delta_i(1-\alpha_i)\right]\left[\sum_{i=1}^{m} B_{\pi i}^*\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right.$$

$$\left.+\left[\sum_{i=1}^{m} B_{\pi i}^*\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]^2\right]\right\}\right\} \frac{\delta M_0}{\delta \xi_0}\left(\frac{\delta W}{\delta \beta}\right)^T,$$

$$I_{\xi\xi} = \left\{\alpha_0\sum_{i=1}^{m}\left\{\alpha_i\left[\sum_{j=1}^{n_i}\delta_{ij}(1-\delta_{ij})\right]-B_{\pi i}^*(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\} \ - \ \alpha_0\sum_{i=1}^{m} A_{\pi i}^*\delta_i^2(1-\alpha_i)^2\right.$$

$$- \ 2\alpha_0\sum_{i=1}^{m} B_{\pi i}^*\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]\delta_i(1-\alpha_i) \ - \ B_k^*(1-\alpha_0)\left\{\left[\sum_{i=1}^{m} A_{\pi i}^*\delta_i(1-\alpha_i)\right]^2\right.$$

$$\left.\left.+2\left[\sum_{i=1}^{m} A_{\pi i}^*\delta_i(1-\alpha_i)\right]\left[\sum_{i=1}^{m} B_{\pi i}^*\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]+\left[\sum_{i=1}^{m} B_{\pi i}^*\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]^2\right\}\right\} \frac{\delta M_i}{\delta \xi}\left(\frac{\delta M_i}{\delta \xi}\right)^T,$$

$$I_{\xi\gamma_0} = I_{\gamma_0\xi} = \left\{\alpha_0 \sum_{i=1}^{m} \left\{\alpha_i \left[\sum_{j=1}^{n_i} \delta_{ij}(1-\delta_{ij})\right] - B^*_{\pi i}(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\}\right.$$

$$- \alpha_0 \sum_{i=1}^{m} A^*_{\pi i}\delta_i^{\,2}(1-\alpha_i)^2 - 2\alpha_0 \sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]\delta_i(1-\alpha_i)$$

$$- B^*_k\left\{-(1-\delta_0)\left[\left[\sum_{i=1}^{m} A^*_{\pi i}\delta_i(1-\alpha_i)\right] + \left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right]\right\}$$

$$- B^*_k(1-\alpha_0)\left\{\left[\sum_{i=1}^{m} A^*_{\pi i}\delta_i(1-\alpha_i)\right]^2 + 2\left[\sum_{i=1}^{m} A^*_{\pi i}\delta_i(1-\alpha_i)\right]\left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right.$$

$$\left.\left. + \left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]^2\right]\right\}\right\} \frac{\delta M}{\delta \xi}\left(\frac{\delta D_0}{\delta \gamma_0}\right)^T,$$

$$I_{\xi\gamma} = I_{\gamma\xi} = \left\{\alpha_0 \sum_{i=1}^{m} \left\{\alpha_i \left[\sum_{j=1}^{n_i} \delta_{ij}(1-\delta_{ij})\right] - B^*_{\pi i}(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\}\right.$$

$$+ \alpha_0 \sum_{i=1}^{m} A^*_{\pi i}\delta_i(1-\delta_i)(1-\alpha_i) - \alpha_0 \sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]\left[\delta_i(1-\alpha_i) - (1-\delta_i)\right]$$

$$- B^*_k(1-\alpha_0)\left\{-\left[\sum_{i=1}^{m} A^*_{\pi i}(1-\delta_i)\right]\left[\sum_{i=1}^{m} A^*_{\pi i}\delta_i(1-\alpha_i)\right] - \left[\sum_{i=1}^{m} A^*_{\pi i}(1-\delta_i)\right]\left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right.$$

$$\left.\left. + \left[\sum_{i=1}^{m} A^*_{\pi i}\delta_i(1-\alpha_i)\right]\left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right] + \left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]^2\right\}\right\} \frac{\delta M_i}{\delta \xi}\left(\frac{\delta D_i}{\delta \gamma}\right)^T,$$

$$I_{\xi\beta} = I_{\beta\xi} = \left\{ \alpha_0 \sum_{i=1}^{m} \left\{ \alpha_i \left[\sum_{j=1}^{n_i} \delta_{ij}(1-\delta_{ij}) \right] - B^*_{\pi i}(1-\alpha_i) \left[\sum_{j=1}^{n_i} (-\delta_{ij}) \right]^2 \right\} \right.$$

$$- \alpha_0 \sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{j=1}^{n_i} (-\delta_{ij}) \right] \delta_i (1-\alpha_i) \text{ - } B^*_k (1-\alpha_0) \left\{ \left[\sum_{i=1}^{m} A^*_{\pi i} \delta_i (1-\alpha_i) \right] \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right] \right.$$

$$\left. \left. + \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right]^2 \right\} \right\} \frac{\delta M_i}{\delta \xi} \left(\frac{\delta W}{\delta \beta} \right)^T,$$

$$I_{\gamma_0\gamma_0} = \left\{ \alpha_0 \sum_{i=1}^{m} \left\{ \alpha_i \left[\sum_{j=1}^{n_i} \delta_{ij}(1-\delta_{ij}) \right] - B^*_{\pi i}(1-\alpha_i) \left[\sum_{j=1}^{n_i} (-\delta_{ij}) \right]^2 \right\} \text{ - } \alpha_0 \sum_{i=1}^{m} A^*_{\pi i} \delta_i^2 (1-\alpha_i)^2 \right.$$

$$\text{- } A^*_k (1-\delta_0)^2 \text{ - } 2\alpha_0 \sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{j=1}^{n_i} (-\delta_{ij}) \right] \delta_i (1-\alpha_i)$$

$$\text{- } B^*_k \left\{ -2(1-\delta_0) \left[\left[\sum_{i=1}^{m} A^*_{\pi i} \delta_i (1-\alpha_i) \right] + \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right] \right] \right\}$$

$$\text{- } B^*_k (1-\alpha_0) \left\{ \left[\sum_{i=1}^{m} A^*_{\pi i} \delta_i (1-\alpha_i) \right]^2 + 2 \left[\sum_{i=1}^{m} A^*_{\pi i} \delta_i (1-\alpha_i) \right] \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right] \right.$$

$$\left. \left. + \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right]^2 \right\} \right\} \frac{\delta D_0}{\delta \gamma_0} \left(\frac{\delta D_0}{\delta \gamma_0} \right)^T,$$

$$I_{\gamma_0\gamma} = I_{\gamma\gamma_0} = \left\{ \alpha_0 \sum_{i=1}^{m} \left\{ \alpha_i \left[\sum_{j=1}^{n_i} \delta_{ij}(1-\delta_{ij}) \right] - B^*_{\pi i}(1-\alpha_i) \left[\sum_{j=1}^{n_i} (-\delta_{ij}) \right]^2 \right\} \right.$$

$$+ \ \alpha_0 \sum_{i=1}^{m} A^*_{\pi i} \delta_i (1-\delta_i)(1-\alpha_i) \ - \ \alpha_0 \sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{j=1}^{n_i} (-\delta_{ij}) \right] \left[\delta_i (1-\alpha_i) - (1-\delta_i) \right]$$

$$- \ B^*_k \left\{ (1-\delta_0) \left[\left[\sum_{i=1}^{m} A^*_{\pi i} (1-\delta_i) \right] - \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right] \right] \right\}$$

$$- \ B^*_k (1-\alpha_0) \left\{ - \left[\sum_{i=1}^{m} A^*_{\pi i} (1-\delta_i) \right] \left[\sum_{i=1}^{m} A^*_{\pi i} \delta_i (1-\alpha_i) \right] \right.$$

$$+ \left[\sum_{i=1}^{m} A^*_{\pi i} \delta_i (1-\alpha_i) \right] \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right] - \left[\sum_{i=1}^{m} A^*_{\pi i} (1-\delta_i) \right] \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right]$$

$$\left. + \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right]^2 \right\} \Bigg\} \frac{\delta D_0}{\delta \gamma_0} \left(\frac{\delta D}{\delta \gamma} \right)^T ,$$

$$I_{\gamma_0 \beta} = I_{\beta \gamma_0} = \left\{ \alpha_0 \sum_{i=1}^{m} \left\{ \alpha_i \left[\sum_{j=1}^{n_i} \delta_{ij} (1-\delta_{ij}) \right] - B^*_{\pi i} (1-\alpha_i) \left[\sum_{j=1}^{n_i} (-\delta_{ij}) \right]^2 \right\} \right.$$

$$- \ \alpha_0 \sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{j=1}^{n_i} (-\delta_{ij}) \right] \delta_i (1-\alpha_i) \ + \ B^*_k \left\{ (1-\delta_0) \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right] \right\}$$

$$+ \ B^*_k (1-\alpha_0) \left\{ \left[\sum_{i=1}^{m} A^*_{\pi i} \delta_i (1-\alpha_i) \right] \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right] \right.$$

$$\left. + \left[\sum_{i=1}^{m} B^*_{\pi i} \left[\sum_{i=1}^{n_i} (-\delta_{ij}) \right] \right]^2 \right\} \Bigg\} \frac{\delta D_0}{\delta \gamma_0} \left(\frac{\delta W}{\delta \beta} \right)^T ,$$

$$I_{\gamma\gamma} = \left\{\alpha_0 \sum_{i=1}^{m}\left\{\alpha_i\left[\sum_{j=1}^{n_i}\delta_{ij}(1-\delta_{ij})\right] - B^*_{\pi i}(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\} - \alpha_0 \sum_{i=1}^{m} A^*_{\pi i}(1-\delta_i)^2\right.$$

$$+2\alpha_0 \sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right](1-\delta_i) - B^*_k(1-\alpha_0)\left\{\left[\sum_{i=1}^{m} A^*_{\pi i}(1-\delta_i)\right]^2\right.$$

$$\left.\left. - 2\left[\sum_{i=1}^{m} A^*_{\pi i}(1-\delta_i)\right]\left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right] + \left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]^2\right\}\right\} \frac{\delta D}{\delta\gamma}\left(\frac{\delta D}{\delta\gamma}\right)^T,$$

$$I_{\gamma\beta} = I_{\beta\gamma} = \left\{\alpha_0 \sum_{i=1}^{m}\left\{\alpha_i\left[\sum_{j=1}^{n_i}\delta_{ij}(1-\delta_{ij})\right] - B^*_{\pi i}(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\}\right.$$

$$+ \alpha_0 \sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right](1-\delta_i) - B^*_k(1-\alpha_0)\left\{\left[\sum_{i=1}^{m} A^*_{\pi i}(1-\delta_i)\right]\left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right.$$

$$\left.\left.+\left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]^2\right\}\right\} \frac{\delta D}{\delta\gamma}\left(\frac{\delta W}{\delta\beta}\right)^T$$

and

$$I_{\beta\beta} = \left\{\alpha_0 \sum_{i=1}^{m}\left\{\alpha_i\left[\sum_{j=1}^{n_i}\delta_{ij}(1-\delta_{ij})\right] - B^*_{\pi i}(1-\alpha_i)\left[\sum_{j=1}^{n_i}(-\delta_{ij})\right]^2\right\}\right.$$

$$\left.+ B^*_k(1-\alpha_0)\left\{\left[\sum_{i=1}^{m} B^*_{\pi i}\left[\sum_{i=1}^{n_i}(-\delta_{ij})\right]\right]\right\}\right\} \frac{\delta W}{\delta\beta}\left(\frac{\delta W}{\delta\beta}\right)^T.$$

Appendix E Tables

Appendix E.1 Tables of the oesophageal cancer data

Table E.1.1: Descriptive statistics of the variables age, meansibage and sibsize

Variable	N	Mean	Standard Deviation	Minimum	Maximum
age	14,310.00	48.26	18.18	1.00	136.00
meansibage	14,310.00	42.04	17.41	1.00	90.00
sibsize	14,310.00	3.37	1.76	1.00	11.00

Table E.1.2: Descriptive statistics of the variables age and meansibage for affected individuals

Variable	N	Mean	Standard Deviation	Minimum	Maximum
age	2,050.00	57.61	9.47	20.00	84.00
meansibage	2,050.00	52.13	11.22	5.00	79.17

Table E.1.3: Descriptive statistics of the variables smoking status, alcohol and status

Variable	Yes	No	Unknown	Total
smoking	3,024	11,286	-	14,310
alcohol	767	13543	-	14,310
status	2,050	12,260	-	14,310

Appendix E.2 Tables of the breast cancer data

Table E.2.1: Frequencies of the variables hinc, obesity, tumour and bca

Variable	0	1	2	3	4	5	6	Unknown	Total
hinc	-	48	81	84	80	116	101	-	510
obesity	319	191	-	-	-	-	-	-	510
tumour	510	0	-	-	-	-	-	-	510
bca	275	235	-	-	-	-	-	-	510

Table E.2.2: Descriptive statistics of the variable ageat

Variable	N	Mean	Standard deviation	Minimum	Maximum
ageat	510	52.23	16.07	19	87

Table E.2.3: Variation of breast cancer side with respect to obesity

Breast side affected	obesity		
	0	1	Total
Neither (0)	157	118	275
Right (1)	75	35	110
Left (2)	85	36	121
Both (3)	2	2	4
Total	319	191	510

Table E.2.4: Variation of breast cancer side with respect to household income

Breast side affected	Household income						
	1	2	3	4	5	6	Total
Neither (0)	28	42	47	50	60	48	275
Right (1)	8	18	15	12	31	26	110
Left (2)	12	19	22	17	25	26	121
Both (3)	0	2	0	1	0	1	4
Total	48	81	84	80	116	101	510

Table E.2.5: Variation of breast cancer side with respect to age at time of examination

Breast side affected	Age at time of examination							
	1-10	11-20	21-30	31-40	41-50	51-60	60+	Total
Neither (0)	0	3	38	68	52	45	69	275
Right (1)	0	0	3	16	21	25	45	110
Left (2)	0	0	3	9	29	23	57	121
Both (3)	0	0	0	2	0	0	2	4
Total	0	3	44	95	102	93	173	510

Appendix F Figures

Appendix F.1 Figures of the oesophageal cancer data

Figure F.1.1: Histogram of age

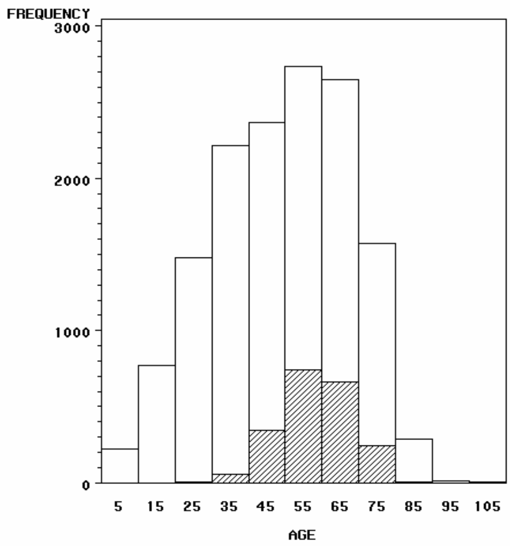


Figure F.1.2: Histogram of meansibage

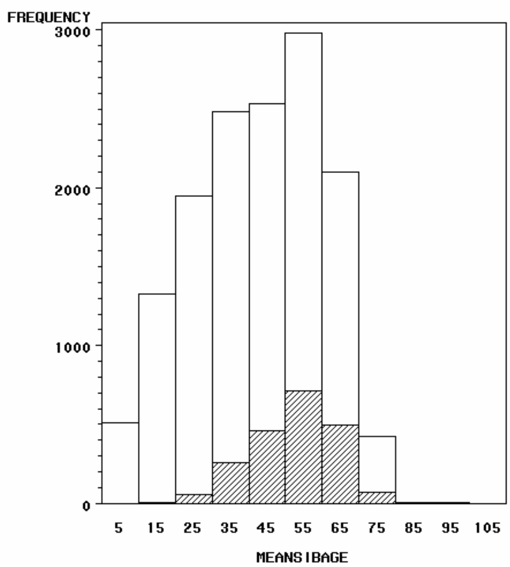


Figure F.1.3: Bar chart of sibsize

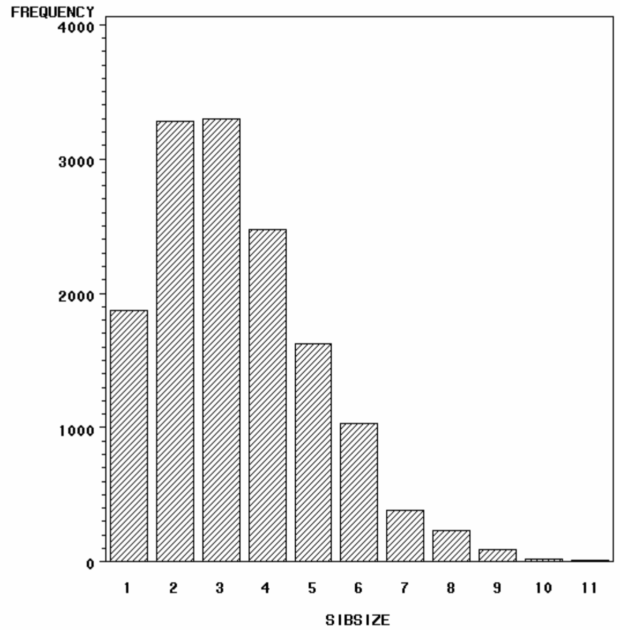


Figure F.1.4: Normal Q-Q plot of age for affected individuals

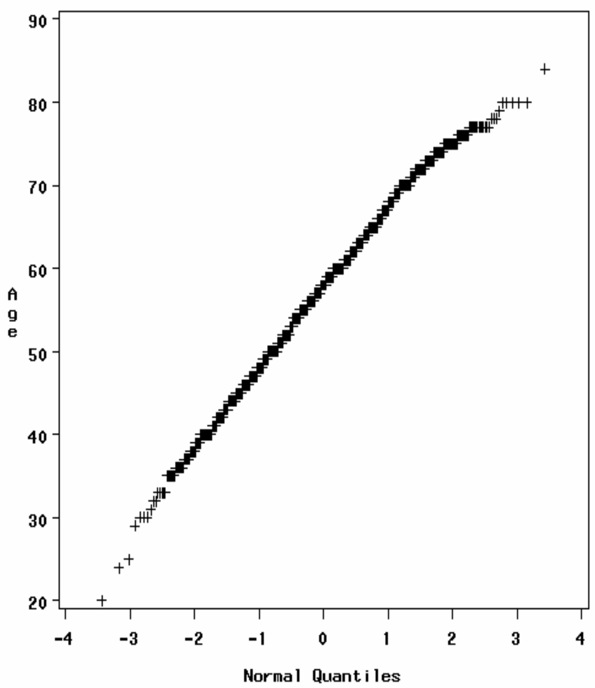
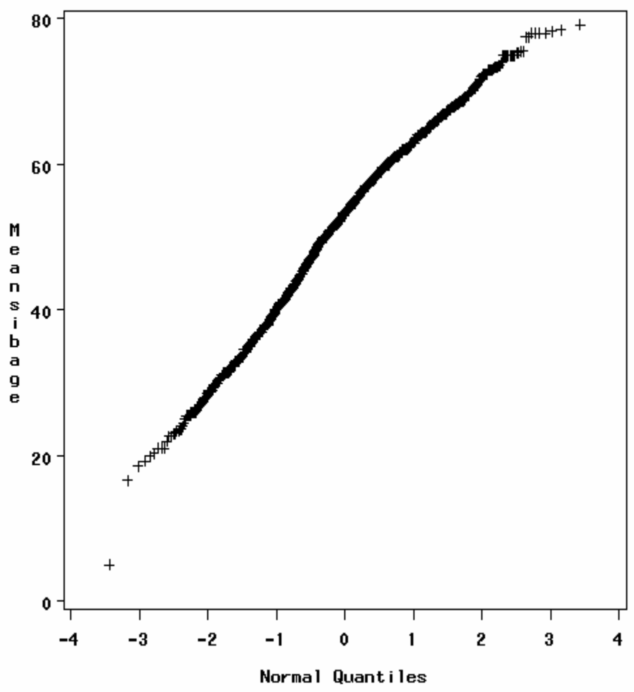


Figure F.1.5: Normal Q-Q plot of meansibage for affected individuals



Appendix F.2 Figures of the breast cancer data

Figure F.2.1. Histogram of ageat

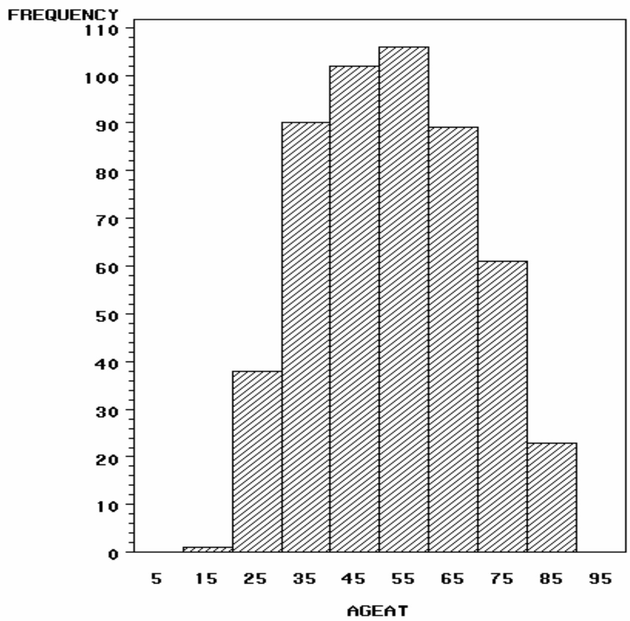


Figure F.2.2: Bar chart of hinc

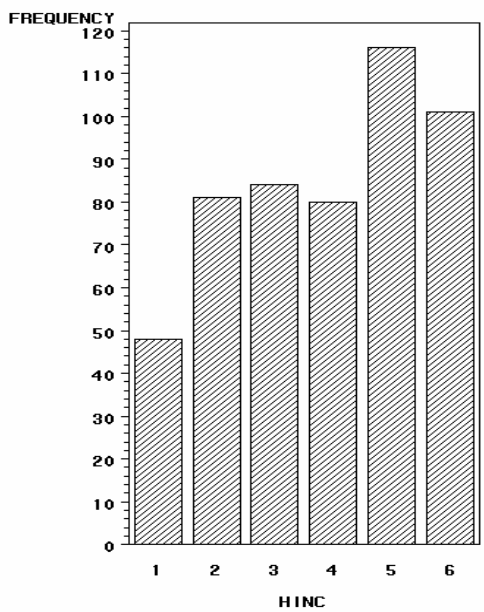
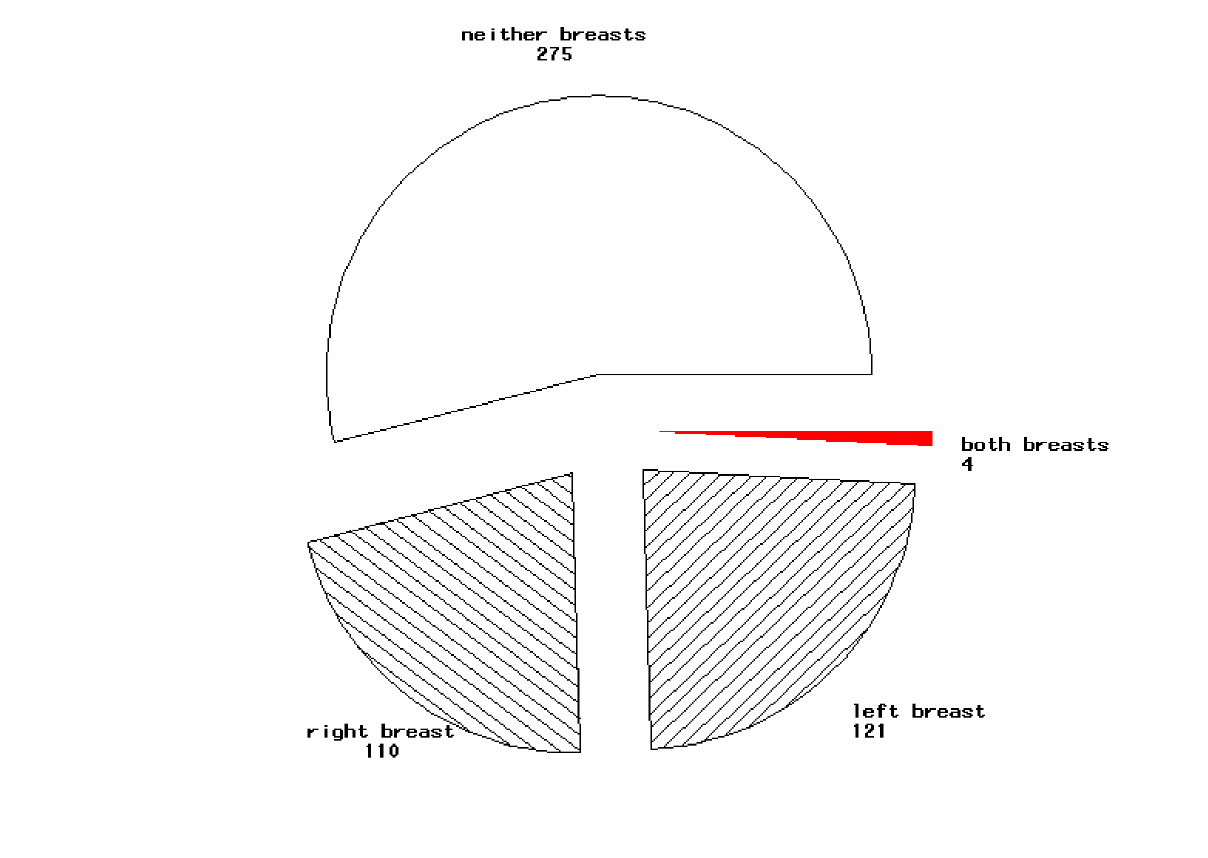


Figure F.2.3: Pie chart of breast cancer side distinguished according to neither breasts, right breast, left breast and both breasts



References

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Acoustics 19, 716-723.

[2] Barlow, R. E. and Proschan, F. (1965). Mathematical Theory of Reliability. John Wiley and Sons, New York.

[3] Bickel, P. J. and Doksum, K. A. (1977). Mathematical statistics: Basic ideas and selected topics. Holden-Day, San Francisco.

[4] Bonney, G. E. (1998). Regression analysis of disposition to correlated binary outcomes. Department of Biostatistics Technical Report No. 1998-001, Fox Chase Cancer Center, Philadelphia, PA, U.S.A.

[5] Carey, V., Zeger, S. L. and Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions. Biometrika 80, 3, 517-526.

[6] Connolly, M. A. and Liang, K. (1988). Conditional logistic regression models for correlated binary data. Biometrika 75, 501-506.

[7] Cox, D. R. (1972). The analysis of multivariate binary data. Applied Statistics 21, 113-120.

[8] Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society, Series 34, 187-220.

[9] Cox, D. R. (1975). Partial likelihood. Biometrika 62, 2, 269-276.

[10] Daintith, J. and Nelson, R. D. (1989). The Penguin Dictionary of Mathematics. Cox and Wyman Limited, Great Britain.

[11] Eberl, W. und Moeschlin, O. (1982). Mathematische Statistik. Walter de Gruyter. Berlin, New York.

[12] Elandt-Johnson, R. C. and Johnson, N. L. (1980). Survival models and data analysis. John Wiley and Sons, U.S.A.

[13] Garthwaite, P. H., Jolliffe, I. T. and Jones, B. (1995). Statistical inference. Prentice Hall, London, New York.

[14] Godambe, V. P. (1991). Estimating Functions. Clarendon Press, Oxford.

[15] Gross, A. J. and Clark, V. A. (1975). Survival distributions: Reliability Applications in the Biomedical Sciences. John Wiley and Sons, New York.

[16] Henze, N. (1995). Stochastik I: Einführung in die Wahrscheinlichkeitstheorie und Statistik. Karlsruhe.

[17] Klein, J. P. and Moeschberger, M. L. (1997). Survival Analysis: Techniques for Censored and Truncated Data. Springer-Verlag, New York.

[18] Kleinbaum, D. G. (1994). Logistic Regression. Springer-Verlag, New York.

[19] Kötting, J., Bonney, G. E. and Urfer, W. (1998). Disposition models for the analysis of dynamic changes in Forest-Ecosystems. Department of Statistics Technical Report No. 24/1998, University of Dortmund, Germany.

[20] Kotz, S. and Johnson, N. L. (1982). Encyclopedia of statistical sciences. John Wiley and Sons, New York, Volume 6, p. 210.

[21] Kwagyan, J. (2001). Further investigations of the disposition model for correlated binary outcomes. Ph.D. Thesis, The Temple University Graduate School, U.S.A..

[22] Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalised linear models. Biometrika 73, 1, 13-22.

[23] Liang, K.-Y., Zeger, S. L. and Qaqish, B. (1992). Multivariate regression analysis for categorical data. Journal of the Royal Statistical Society, Series B 54, 3-40.

[24] McLachlan, G. J. and Krishnan, T. (1997). The EM Algorithm and Extensions. John Wiley and Sons, New York.

[25] Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). Introduction to the theory of statistics. McGRAW-Hill, London.

[26] Nelson, W. B. (1972). Theory and Applications of Hazard Plotting for Censored Failure Data. Technometrics 14, 945-965.

[27] Odai, R. N. O., Kötting, J., Kwagyan, J., Bonney, G. E. and Urfer, W. (2002). Correlated Weibull Regression Model for Multivariate Binary Data. Department of Statistics Technical Report No. 66/2002, University of Dortmund, Germany.

[28] Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. Biometrics 44, 1033-1048.

[29] Qaqish, B. F. and Liang, K.-Y. (1992). Marginal models for correlated binary responses with multiple classes and multiple levels of nesting. Biometrics 48, 939-950.

[30] Stuart, A., Ord, J. K. and Arnold, S. (1999). Kendall's advanced theory of statistics, classical inference and linear model. Edward Arnold, London, Volume 2A.

[31] Thomas, D. C. and Gauderman, W. J. (1995). Gibbs sampling methods in genetics. Chapman and Hall, London.

[32] Van der Vaart, A. W. (1998). Asymptotic statistics. Cambridge University Press, United Kingdom.

[33] Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. Annals of mathematical statistics, 9, 60-62.

[34] Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. Biometrika 77, 642-648.