Available online at www.sciencedirect.com

ScienceDirect

Computational Statistics and Data Analysis 52 (2008) 3877–3888

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Some properties of regression estimators in GEE models for clustered ordinal data

María Laura Nores[a], María del Pilar Díaz[b,*]

[a] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and Facultad de Ciencias Médicas, Universidad Nacional de Córdoba, Enrique Barros y Haya de la Torre, 5000, Córdoba, Argentina*

[b] *Escuela de Nutrición, Facultad de Ciencias Médicas, Universidad Nacional de Córdoba, Enrique Barros s/n, Pabellón Chile, 5000, Córdoba, Argentina*

Received 3 April 2006; received in revised form 17 December 2007; accepted 17 December 2007
Available online 23 December 2007

Abstract

In this paper we study properties of the estimators of marginal mean parameters in the GEE1 approach of Heagerty and Zeger [Heagerty, P.J., Zeger, S.L., 1996. Marginal regression models for clustered ordinal measurements. J. Amer. Statist. Assoc. 91, 1024–1036] for clustered ordinal data. We consider two aspects: coverage probabilities and efficiency. The first point was tackled by a simulation study, calculating empirical levels of confidence intervals for regression parameters using different sample sizes. We showed that the difference between empirical and nominal levels widens when sample size decreases, especially when the probability for a given response category is low in a group of clusters with the same covariate vector. We studied asymptotic efficiency for the case of an independence working specification in relation to a correctly specified exchangeable association structure. We extended to ordinal measurements the results derived for binary outcomes, sustaining that the loss of efficiency depends both on the intensity of the association between responses and the design matrix. For equal cluster sizes, we showed that relative efficiency is high when responses are independent, when covariates are mean-balanced, or when all covariates are constant within clusters. However, relative efficiency noticeably declines with increasing association for non-mean-balanced within-cluster covariates. Simulation studies also supported these conclusions for data with an approximately exchangeable association structure.

1. Introduction

Clustered data are frequent in biological and medical experimental research. This includes longitudinal studies, where individuals are observed over time (Agresti, 2002; Diggle et al., 2002) or for other dimensions such as distance to some origin (Singer and Andrade, 1986), and also in family studies (Ziegler et al., 1998; Yan and Fine, 2004). In this context, Liang and Zeger (1986) proposed the generalized estimating equations (GEE) approach, which not only models the marginal means in terms of covariates but also incorporates the association between cluster responses.

Several GEE alternatives have been proposed. GEE1 approaches consider association as a *nuisance* characteristic, with the main interest residing in relating marginal means to covariates. When association is also focused on, GEE2

[*] Corresponding author. Tel.: +54 351 433 4020.
E-mail address: pdiaz@fcm.unc.edu.ar (M.P. Díaz).

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.12.009

methods (Zhao and Prentice, 1990; Prentice and Zhao, 1991; Liang et al., 1992; Heagerty and Zeger, 1996; Yan and Fine, 2004) offer higher efficiency in the estimation of association parameters, but are computationally costly and association structure has to be correctly specified to consistently estimate marginal mean parameters, which is not necessary in GEE1 (Liang et al., 1992; Carey et al., 1993; Heagerty and Zeger, 1996; Ziegler et al., 1998; Huang et al., 2002). A third alternative known as ALR (*Alternating Logistic Regressions*) (Carey et al., 1993; Heagerty and Zeger, 1996) is similar to GEE1, but includes some modifications in the estimating mechanism.

Clustered ordinal data are a particular case. However, since an ordinal outcome has a multivariate representation, covariates need to be rearranged and the working correlation matrix has to account for the correlation structure of the multinomial distribution (Ziegler et al., 1998). GEE approaches marginally assume a cumulative model. Global odds ratios (Williamson et al., 1995; Heagerty and Zeger, 1996) are recommended to model association (Dale, 1986; Fahrmeir and Tutz, 2001), since they are less constrained than correlations, are more easily interpreted and exploit the ordering of the categories.

In this paper we study some properties of the estimators of marginal mean parameters in the context of the GEE1 approach of Heagerty and Zeger (1996) for ordinal data. We focus on two aspects: coverage probabilities and efficiency. For the first one, we made a simulation study and calculated empirical levels of the confidence intervals for regression parameters based on the sandwich variance estimator. Lipsitz et al. (1991) studied these coverage probabilities in binary responses for a sample size of 100. We considered from 20 to 300 clusters, since our aim was to empirically verify the validity of inferences for different sample sizes. Concerning efficiency, some other authors have investigated the loss of efficiency that can occur when assuming different working association structures. However, their studies were confined to continuous or discrete data, principally binary outcomes. Clustered ordinal data have the extra complexity that observations within each cluster are also multivariate due to their multinomial nature. Fitzmaurice (1995) highlighted that the degree of efficiency depends on both the strength of association between responses and the covariate design. In the case of binary data, he showed for cluster-level covariates that assuming independence does not lead to efficiency loss in relation to exchangeable specification, but for multivariate normally distributed within-cluster covariates (with intra-cluster correlation $r < 1$), the relative efficiency declines with increasing correlation between the responses. Sutradhar and Das (1999) studied relative efficiency for binary responses with the same covariate pattern, but with the association parameter α for the working specification being replaced by the limiting value α_0 of its estimator. This was extended by Sutradhar and Das (2000) for covariates that vary within and among clusters, which are assumed to follow a multivariate normal distribution. Mancl and Leroux (1996) studied asymptotic efficiency by considering an independence working specification in relation to a correctly specified exchangeable association structure. They showed for the case of constant weights, and thus approximately for binary data, that relative efficiency is equal to one when responses within clusters are independent, when all covariates are constant within clusters, or when all covariates vary within clusters but are mean-balanced. For unequal cluster sizes, efficiency for cluster-level covariates depends also on the coefficient of variation in cluster sizes. In the present paper, we studied if the efficiency results of Mancl and Leroux (1996) could be extended for ordinal data. We also examined relative efficiency by simulations for situations with approximately exchangeable association.

This article is organized as follows. Section 2 describes briefly the GEE1 approach of Heagerty and Zeger (1996). In Section 3 we show the simulation schemes used and the results concerning confidence levels. In Section 4 we present the study of the efficiency of regression estimators. The final section discusses the results obtained.

2. The model

Let $\mathbf{y}_i = \left(y_{i1}, y_{i2}, \ldots, y_{iT_i}\right)'$ represent the vector of ordinal measurements for the ith subject or cluster, $i = 1, \ldots, K$, where $y_{it} \in \{1, \ldots, C\}$. Heagerty and Zeger (1996) proposed to use a cumulative model for the marginal means, that is:

$$g\left[E\left(y_{itc}\right)\right] = \theta_c + \mathbf{x}_{it}'\boldsymbol{\lambda}, \tag{1}$$

where $y_{itc} = I_{y_{it}>c}, c = 1, \ldots, C-1, \theta_c$ is a cutpoint, $\mathbf{x}_{it}$ denotes the covariates associated with y_{it}, $\boldsymbol{\lambda}$ is a $p \times 1$ vector of unknown parameters and g is a link function. The association between two observations t and t' from the same cluster is considered using global odds ratios (Dale, 1986), defined as

$$\Psi_{i(tt')(cc')} = \frac{P\left(y_{it} \le c, y_{it'} \le c'\right) P\left(y_{it} > c, y_{it'} > c'\right)}{P\left(y_{it} \le c, y_{it'} > c'\right) P\left(y_{it} > c, y_{it'} \le c'\right)}.$$

Then, the model for the association is $\log\left(\Psi_{i(tt')(cc')}\right) = \mathbf{e}'_{i(tt')(cc')}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a vector of unknown parameters and $\mathbf{e}_{i(tt')(cc')}$ are covariates.

The GEE1 estimating equation for $\boldsymbol{\beta} = \left(\boldsymbol{\theta}', \boldsymbol{\lambda}'\right)'$ results in:

$$\sum_{i=1}^{K} D_i' V_i^{-1}\left(\mathbf{y}_i^* - \mathbf{p}_i\right) = 0,$$

where $\mathbf{y}_i^* = \left(y_{i11}, \ldots, y_{i1,C-1}, y_{i21}, \ldots, y_{i2,C-1}, \ldots, y_{iT_i,C-1}\right)'$, $\mathbf{p}_i = E\left(\mathbf{y}_i^*\right)$, $D_i = \partial\mathbf{p}_i/\partial\boldsymbol{\beta}$ and V_i is the working covariance matrix of $\mathbf{y}_i^*$, totally specified in terms of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. The estimating equation for $\boldsymbol{\alpha}$ is:

$$\sum_{i=1}^{K}\left(\frac{\partial\boldsymbol{\sigma}_i}{\partial\boldsymbol{\alpha}}\right)'\left(\mathrm{var}\left(\mathbf{S}_i\right)\right)^{-1}\left(\mathbf{S}_i - \boldsymbol{\sigma}_i\right) = 0,$$

where $\mathbf{S}_i = \left(S_{i(12)(11)}, \ldots, S_{i(12)(1,C-1)}, \ldots, S_{i(12)(C-1,C-1)}, \ldots, S_{i(T_i-1,T_i)(C-1,C-1)}\right)'$, $S_{i(tt')(cc')} = \left(y_{itc} - p_{itc}\right) \times \left(y_{it'c'} - p_{it'c'}\right)$, $\boldsymbol{\sigma}_i = E\left(\mathbf{S}_i\right)$. A block-diagonal matrix can be used as an approximation of var $(\mathbf{S}_i)$, using only first and second moment assumptions.

A Fisher-scoring type algorithm is used to obtain $(\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\alpha}}')'$. The estimator $\widehat{\boldsymbol{\beta}}$ is consistent and asymptotically multivariate Gaussian with covariance matrix

$$\left(\sum_{i=1}^{K} D_i' V_i^{-1} D_i\right)^{-1}\left(\sum_{i=1}^{K} D_i' V_i^{-1}\mathrm{var}\left(\mathbf{y}_i^*\right) V_i^{-1} D_i\right)\left(\sum_{i=1}^{K} D_i' V_i^{-1} D_i\right)^{-1}. \tag{2}$$

A consistent estimator of (2), known as the robust or sandwich covariance estimator, is obtained by replacing var $\left(\mathbf{y}_i^*\right)$ by $\left(\mathbf{y}_i^* - \widehat{\mathbf{p}}_i\right)\left(\mathbf{y}_i^* - \widehat{\mathbf{p}}_i\right)'$ and $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by their estimates. When var $\left(\mathbf{y}_i^*\right)$ is correctly specified, then (2) reduces to:

$$\left(\sum_{i=1}^{K} D_i'\left(\mathrm{var}\left(\mathbf{y}_i^*\right)\right)^{-1} D_i\right)^{-1}. \tag{3}$$

3. Coverage probabilities

In order to study empirically the validity of inferences for the regression parameters in the GEE1 approach of Heagerty and Zeger (1996), we conducted a simulation study.

Cumulative models, as (1), assume that an observable ordinal variable y with C categories is a categorized version of an underlying unobservable continuous variable U, considering certain cutpoints θ_c. That is to say

$$y = c \Leftrightarrow \theta_c < U \le \theta_{c-1}, \quad c = 1, \ldots, C, \tag{4}$$

where $-\infty = \theta_C < \theta_{C-1} < \cdots < \theta_1 < \theta_0 = \infty$. The most common parameterization supposes that U is related to a vector of explanatory variables $\mathbf{x}$ by

$$U = -\mathbf{x}'\boldsymbol{\lambda} + \varepsilon, \tag{5}$$

where $\boldsymbol{\lambda}$ is a vector of parameters and ε is a random variable with distribution function F. Then, the model is:

$$P\left(y > c|\mathbf{x}\right) = P\left(U \le \theta_c\right) = F\left(\theta_c + \mathbf{x}'\boldsymbol{\lambda}\right).$$

This formulation of the cumulative model suggests a mechanism for generating ordinal variables from the categorization of a continuous variable considering some cutpoints, using (4). Thus, we obtained clustered ordinal simulated data by categorizing correlated normal random variables.

We considered a three-level ordinal response and a single covariate, the model for marginal means being

$$probit\left(P\left(y_{it} > c\right)\right) = \theta_c + x_{it}\lambda, \quad i = 1, \ldots, K, t = 1, \ldots, T_i, c = 1, 2. \tag{6}$$

Table 1
Description of the scenarios of the first simulation scheme (mean-balanced or cluster-level covariates)

Independence ($\rho = 0$)

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\boldsymbol{\theta}$
Ia	$(0, 1.5)'$		$(1.317, 0.183)'$
Ib	$(0, 0.6, -0.2)'$		$(0.586, -0.325)'$
Ic	$(0, 0, 0)'$	$(-2, -2, -2)'$	$(-0.309, -1.691)'$
Id	$(0, -1, -2, 1)'$	$(-1, -2, 0, 1)'$	$(0, -1)'$

Association ($\rho \neq 0$)

	ρ	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\boldsymbol{\theta}$	Ψ
Aa	0.2	$(0, 0.6, 1.2)'$		$(0.9,0.3)'$	[1.69, 1.95]
Ab	0.5	$(0, 0.4, -0.4)'$	$(-0.2, -0.2, 0.4)'$	$(0.2, -0.2)'$	[4.00, 5.33]
Ac	0.75	$(0, 0, 0)'$	$(0.2,0.2,0.2)'$	$(0.3, 0.1)'$	[11.23, 11.62]
Ad	0.85	$(0, -0.02, -0.03)'$	$(-0.01, -0.02, -0.02)'$	$(-0.01, -0.03)'$	[21.74, 21.81]
Ae	0.92	$(0, 0, 0)'$	$(-0.1, -0.1, -0.1)'$	$(0.05, -0.05)'$	[46.29, 47.68]
Af	0.97	$(0, 0.02, 0.04)'$		$(0.03, 0.01)'$	[139.09, 143.22]
Ag	0.4	$(0, 0.3, 0.6, 0.3)'$	$(0.3, 0.4, 0.5, 0)'$	$(0.45, 0.15)'$	[2.93, 3.24]
Ah	0.85	$(0, 0, 0, 0)'$	$(-0.02, -0.02, -0.02, -0.02)'$	$(-0.01, -0.03)'$	[21.74, 21.75]

For simplicity in relating (5) and (6) we used $\lambda = -1$. For each scenario (see Section 3.1), we generated 10,000 subsets of K clustered observations in an ordinal scale. For each i, a value for the corresponding vector of clustered observations was generated through a multivariate normal distribution with mean $\mathbf{x}_i = \left(x_{i1}, \ldots, x_{iT_i}\right)'$ and covariance matrix Σ_i. The choice of $\mathbf{x}_i$ and Σ_i is discussed in Section 3.1. The response was categorized at each t in three classes, considering two cutpoints θ_1 and θ_2. To model association, we used independence ($\Psi_{i(tt')(cc')} = 1$) and exchangeable ($\log\left(\Psi_{i(tt')(cc')}\right) = \alpha$) working specifications.

Asymptotic 95% confidence intervals for θ_c and λ are given, respectively, by:

$$\widehat{\theta}_c \pm 1.96\widehat{\sigma}_{\widehat{\theta}_c} \text{ and } \widehat{\lambda} \pm 1.96\widehat{\sigma}_{\widehat{\lambda}}$$

using the sandwich variance estimator for $\widehat{\sigma}_{\widehat{\theta}_c}$ and $\widehat{\sigma}_{\widehat{\lambda}}$. We determined the percentage of times these confidence intervals included after model fitting the true values used to generate the categories. With 10,000 replicates, we obtained a precision of ±0.4 (for a confidence coefficient of 0.95). Those empirical confidence levels were calculated for each scenario in subsets of $K = 20$, 30, 50, 100 and 300 clusters, in order to evaluate the validity of inferences for different sample sizes.

3.1. Simulation schemes

In a first scheme of simulations, we considered mean-balanced ($\bar{\mathbf{x}}_i = \bar{\mathbf{x}}_{i'} \forall i, i'$, with $\bar{\mathbf{x}}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} x_{it}$) or cluster-level ($x_{it} = x_{it'} \forall t, t'$) covariates and equal cluster sizes ($T_i = T \forall i, T = 2, 3, 4$). In some cases we stated the same covariate pattern for each cluster (a special case of mean-balanced covariates), generating K values from $N\left(\mathbf{x}^{(1)}, \Sigma\right)$, where Σ had diagonal elements 1 and off-diagonal elements ρ, and using those values to generate clustered ordinal observations according to the cutpoints θ_1 and θ_2. In other situations we considered two groups of clusters. Specifically, we generated $K/2$ values from $N\left(\mathbf{x}^{(1)}, \Sigma\right)$ and $K/2$ values from $N\left(\mathbf{x}^{(2)}, \Sigma\right)$ to obtain the ordinal responses, and therefore the covariates used in (6) being:

$$\mathbf{x}_i = \begin{cases} \mathbf{x}^{(1)}, & i = 1, \ldots, K/2, \\ \mathbf{x}^{(2)}, & i = K/2 + 1, \ldots, K. \end{cases}$$

Table 1 describes the scenarios considered in the first simulation scheme. Situations (Ia), (Ib), (Ic) and (Id) correspond to independent observations ($\rho = 0$), while cases from (Aa) to (Ah) represent measurements with association ($\rho \neq 0$), with an approximately exchangeable structure. The values of ρ, $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ and θ_c are indicated and also the interval in which $\Psi_{i(tt')(cc')}$ falls according to that choice, when $\rho \neq 0$. Scenarios are ordered by intensity of association within cluster size.

Table 2
Description of the scenarios of the second simulation scheme (non-mean-balanced within-cluster covariates)

Independence ($\rho = 0$)			
	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	
I'a	$(0, 0.75)'$	$(0.75, 1.5)'$	
I'b	$(-0.2, 0, 0.6)'$	$(0, 0, 0.4)'$	
I'c	(-2,0,0)$'$	$(-2, -1, -1)'$	
I'd	$(0, -1, -1, 1)'$	$(-2, 0, 0, 0)'$	
Association ($\rho \neq 0$)			
	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	Ψ
A'a	$(0, 0.6, 0.6)'$	$(0.6, 1.2, 1.2)'$	[1.69, 1.84]
A'b	$(0, 0.4, 0.4)'$	$(-0.4, -0.4, -0.4)'$	[4.04, 4.57]
A'c	$(0.2,0,0.2)'$	$(0.2, 0.4, 0.4)'$	[11.23, 12.43]
A'd	$(-0.02, 0, -0.02)'$	$(-0.02, -0.02, -0.03)'$	[21.74, 21.78]
A'e	$(-0.1, -0.05, 0)'$	$(0.1, 0.05, 0)'$	[46.29, 51.60]
A'f	$(0, 0.02, 0.04)'$	$(0.02,0,0)'$	[139.09, 143.22]
A'g	$(0, 0.3, 0.4, 0.3)'$	$(0.3, 0.4, 0.5, 0.6)'$	[2.93, 3.12]
A'h	$(0, -0.02, -0.04, 0)'$	$(0, -0.01, -0.01, -0.01)'$	[21.74, 21.84]

In a second configuration (Table 2), we considered two groups of clusters with covariates that vary within clusters but are not mean-balanced, that is, with $\overline{\mathbf{x}}^{(1)} \neq \overline{\mathbf{x}}^{(2)}$ and $\mathbf{x}^{(1)}$ or $\mathbf{x}^{(2)}$ not having all its components equal. Scenarios (I'a) to (I'd) and (A'a) to (A'h) are respectively identified with situations (Ia) to (Id) and (Aa) to (Ah) of the first scheme, with the same values for ρ, θ_1 and θ_2 being maintained and with the covariate vectors $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ being small modifications of those from the first scheme.

Finally, we analyzed the unequal cluster sizes case. We considered four situations with different design matrix and the same values of $\boldsymbol{\theta} = (0.03, 0.01)'$ and $\rho = 0.95$, with Ψ between 79 and 81:

(i) Mean-balanced covariates:
$\mathbf{x}^{(1)} = (0, 0.02, 0.04)'\,;\, \mathbf{x}^{(2)} = (0.01, 0.03, 0.03, 0.01)'\,.$
(ii) Cluster-level covariates, $T_i = 3, 4$:
$\mathbf{x}^{(1)} = (0, 0, 0)'\,;\, \mathbf{x}^{(2)} = (0, 0, 0, 0)'\,;\, \mathbf{x}^{(3)} = (0.02, 0.02, 0.02)'\,;\, \mathbf{x}^{(4)} = (0.02,0.02,0.02,0.02)'\,.$
(iii) Cluster-level covariates, $T_i = 2, 10$:
$\mathbf{x}^{(1)} = (0, 0)'\,;\, \mathbf{x}^{(2)} = (\,0,0,0,0,0,0,0,0,0,0)'\,;\, \mathbf{x}^{(3)} = (0.02,0.02)'\,;$
$\mathbf{x}^{(4)} = (0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02)'\,.$
(iv) Non-mean-balanced within-cluster covariates:
$\mathbf{x}^{(1)} = (0, 0.02, 0.04)'\,;\, \mathbf{x}^{(2)} = (0.01, -0.01, 0, 0.01)'\,.$

In (ii) and (iii), we considered four groups of $K/4$ clusters, each of them with covariate vector $\mathbf{x}^{(j)}$, $j = 1, \ldots, 4$. Values for K of 30 and 50 were changed respectively to 32 and 52 to be divisible by four.

In some scenarios of the first configuration, the cutpoints θ_c were "population terciles" based on all the observations in a data subset. For example, for the same covariate vector $\mathbf{x} = (x_1, \ldots, x_T)'$ for all clusters, these values were calculated by supposing that the TK observations were realizations of a random variable that assumes, with probability of $\frac{1}{T}$, a value from $N\left(x_t, 1\right), t = 1, \ldots, T$. Therefore, θ_c is the solution of $\frac{1}{T}\sum_{t=1}^{T}\Phi\left(\theta_c\right) = \frac{c}{3}, c = 1, 2$, where Φ is the standard normal distribution function.

To calculate the value of the global odds ratios for each scenario, it is necessary to note that

$$\Psi_{i(tt')(cc')} = \frac{P\left(U_{it} > \theta_c, U_{it'} > \theta_{c'}\right) P\left(U_{it} \leq \theta_c, U_{it'} \leq \theta_{c'}\right)}{P\left(U_{it} > \theta_c, U_{it'} \leq \theta_{c'}\right) P\left(U_{it} \leq \theta_c, U_{it'} > \theta_{c'}\right)},$$

where $\left(U_{i1}, \ldots, U_{iT_i}\right)'$ is the multivariate normal random vector used to generate the observations. Therefore, $\Psi_{i(tt')(cc')}$ involves calculating joint probabilities from a bivariate normal random vector with mean $(x_{it}, x_{it'})'$ and covariance matrix $\Sigma_{i,tt'}$ (rows and columns t and t' of Σ_i). The simplest association structure is "independence" ($\Psi_{i(tt')(cc')} = 1$ for all i, t, t', c, c'), which is equivalent to $\rho = 0$. For each i, and $T_i = 3$, say, there are 12 global odds

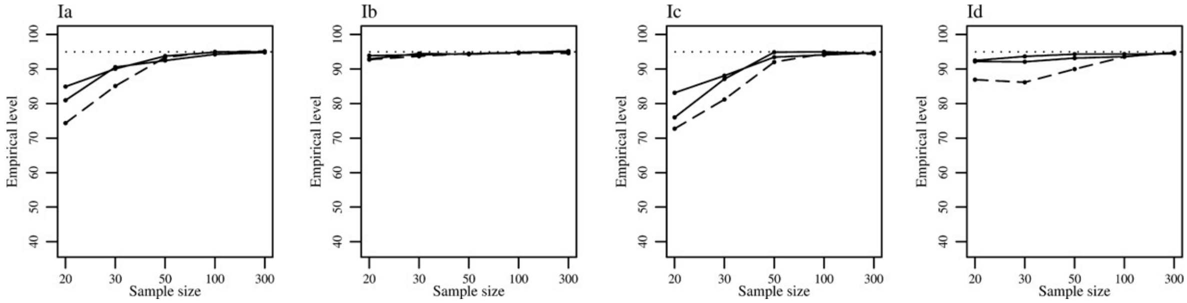

Fig. 1. Empirical confidence levels according to sample size for independent simulated data of the first simulation scheme. Solid lines correspond to θ_c, and dashed lines to λ. Dotted lines indicate the 95% nominal level.

ratios. If there exists some symmetry in the election of θ_c and x_{it}, for example $x_{i2} = x_{i1} + e, x_{i3} = x_{i1} + 2e, \theta_1 = x_{i1} + d, \theta_2 = x_{i3} - d$, there are at most seven out of the 12 global odds ratios that are different. It also can be shown that the number of different global odds ratios reduces to five for $e = 2d$. On the other hand, for fixed covariate vectors and cutpoints, Ψ increases as ρ increases, and while x_{it} values get closer, Ψ declines. Also, when cutpoints get closer, the four global odds ratios associated with each pair t and t' get more similar. All these considerations were taken into account to define scenarios with an approximately exchangeable association structure, and different strengths of association. That is, values of ρ and x_{it} (and θ_c) were varied to cover a wide range of global odds ratio values and also to get specific covariate designs. As a reviewer suggested, we set the first component of $\mathbf{x}^{(1)}$ in the first scheme at 0 in order to facilitate comparisons between scenarios.

Our simulation studies were implemented in *R*, version 2.3.1 (R Development Core Team, 2006) and the corresponding routine for the case of equal cluster sizes (first and second simulations schemes) is available at http://www.fcm.unc.edu.ar/escuelas/nutricion/catedras/estadistica/cated_Estadistica%20y%20Bioestadistica.htm. For model fitting we used the *R* function *ordgee* of package *geepack* (Yan, 2002), which handles clustered ordinal data following Heagerty and Zeger (1996). For multivariate normal vector generation, we applied the function *rmvnorm* of package *mvtnorm* (Genz et al., 2006).

3.2. *Results*

Fig. 1 shows the empirical confidence levels in situations (Ia), (Ib), (Ic) and (Id), considering an independence working association structure, for sample sizes $K = 20, 30, 50, 100$ and 300. Fig. 2 corresponds to the situations with association of the first simulation scheme, using an exchangeable working specification. Similar values of empirical levels were observed for the corresponding scenarios from the second simulation scheme (results not shown). This indicates that the behaviour of empirical levels is not related to the design matrix (in the way considered in this work). That is, for similar values in covariates, coverage patterns do not notably change whether the covariates are cluster-level, mean-balanced or not. This is also supported by the unequal cluster sizes case (Fig. 3).

The observed patterns for the different parameters are similar within each scenario. On the other hand, it can be seen in general that the difference between empirical and nominal confidence levels declines as K grows. For $K = 300$, the percentage observed is approximately 95% in every case, while for $K = 100$, it is higher than 90% in most scenarios, supporting the validity of inferences for these sample sizes in the cases considered.

For situations with association, when global odds ratios are lower than 20, the difference between empirical and nominal levels is almost insignificant for all sample sizes. However, there is an important increase in that difference for higher values, specifically in scenarios (Af) and (Ah), where the empirical levels are between approximately 60% and 70% for $K = 30$ and are around 40% and 50%, for $K = 20$. This should indicate that, as the association grows large, inferences are less reliable for small sample sizes. However, in some independence situations there also exist noticeable differences when K decreases. Furthermore, it is worth mentioning that we modified some scenarios varying the strength of association, but letting fixed the other parameters, and similar coverage patterns were obtained, showing that empirical levels are not linked to intensity of association among responses. It can be noted that smaller coverage values are observed in situations where $P\left(y_{it} = c\right)$ is low for some c and t within a group. For example, $P\left(y_{it} = 2\right) = 0.008$ for all i, t in (Ad), (Af) and (Ah), while in (Aa), where empirical levels are almost 95% for all

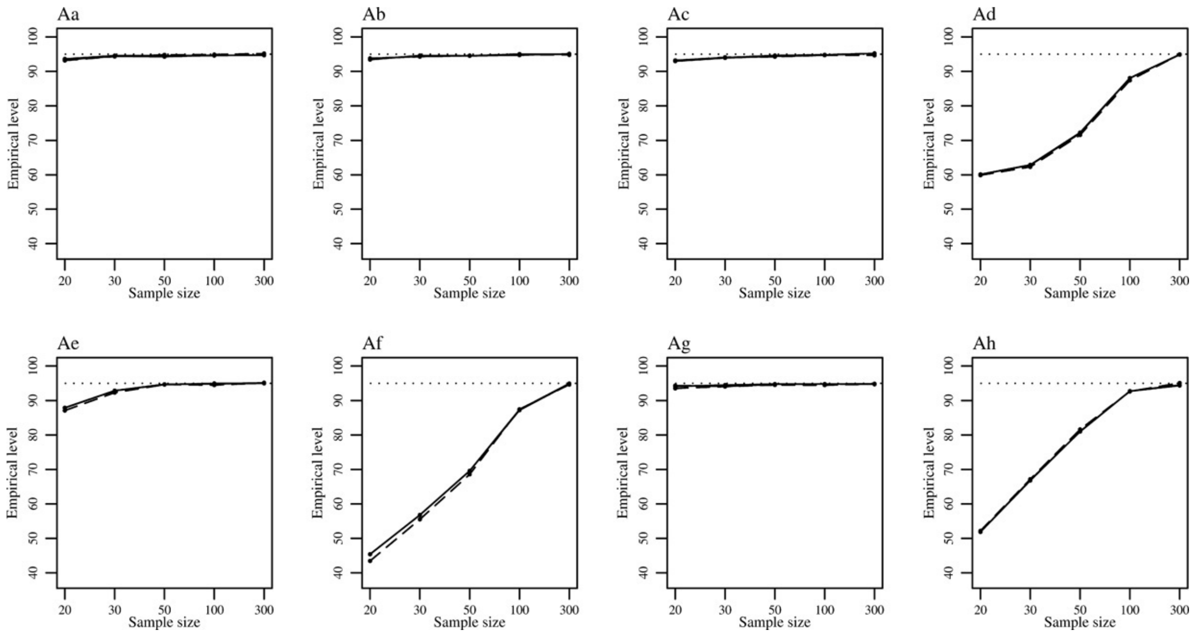

Fig. 2. Empirical confidence levels according to sample size for situations with association of the first simulation scheme. Solid lines correspond to θ_c, and dashed lines to λ. Dotted lines indicate the 95% nominal level.

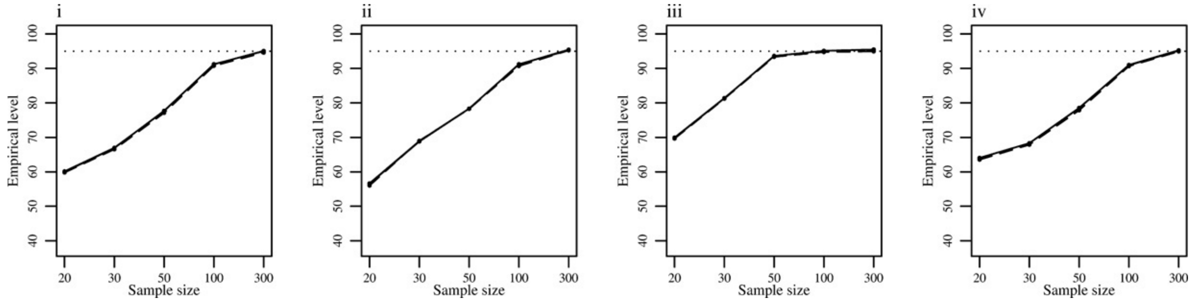

Fig. 3. Empirical confidence levels according to sample size for situations with unequal cluster sizes. Solid lines correspond to θ_c, and dashed lines to λ. Dotted lines indicate the 95% nominal level.

sample size, the lowest value of $P\,(y_{it} = c)$ is 0.184. That may occur since estimation problems arise when some response category is not well represented in a dataset, and this is more frequent for a low value of the corresponding probability and worsens when sample size decreases.

It is also remarkable that similar values of empirical levels were obtained in all the situations with association in both simulation schemes and in the unequal cluster sizes case, using an independence working association structure (considering only the replicates where the estimation algorithm converged for both independence and exchangeable working specifications).

4. Efficiency of regression estimators

4.1. Asymptotic relative efficiency under working independence versus true exchangeable association structure

In this section we present the study of the asymptotic efficiency of the estimator $\widehat{\boldsymbol{\beta}}_I$, using an independence working specification, relative to $\widehat{\boldsymbol{\beta}}_{Ex}$, assuming a correctly specified exchangeable association structure. For each element of $\boldsymbol{\beta}$, the asymptotic relative efficiency (ARE) is given by the ratio of the diagonal elements of (3) and (2) (see Appendix). We considered analogous situations to those presented in the simulation schemes in Section 3, assuming

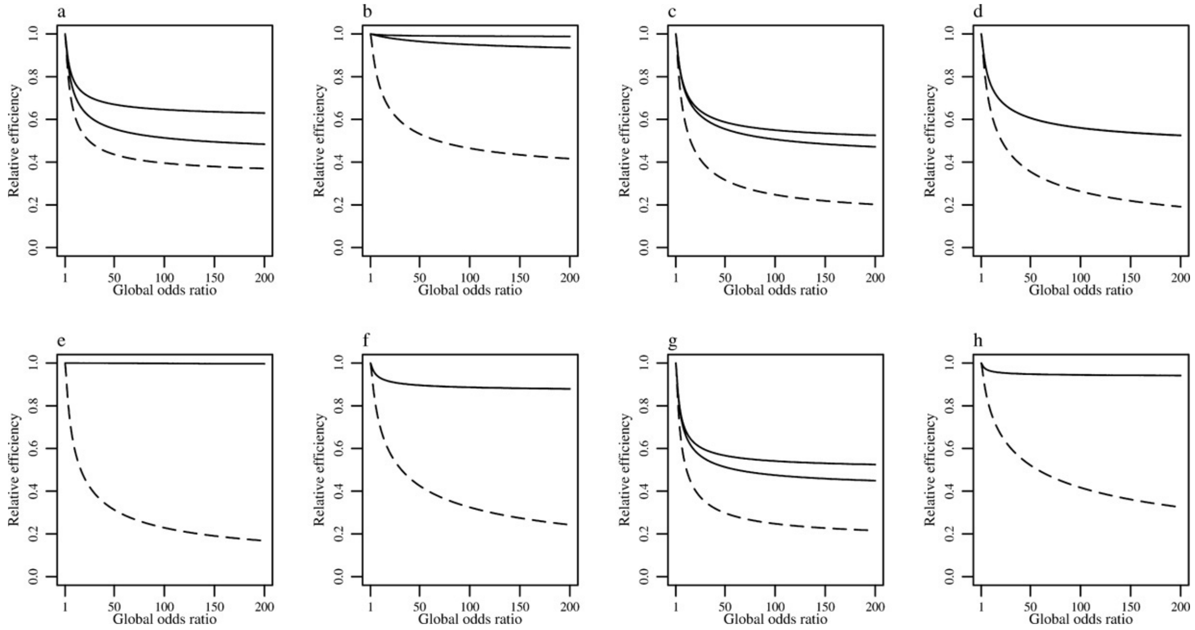

Fig. 4. Asymptotic relative efficiency of independence to exchangeable estimator for non-mean-balanced within-cluster covariates and equal cluster sizes. Solid lines correspond to θ_c, and dashed lines to λ.

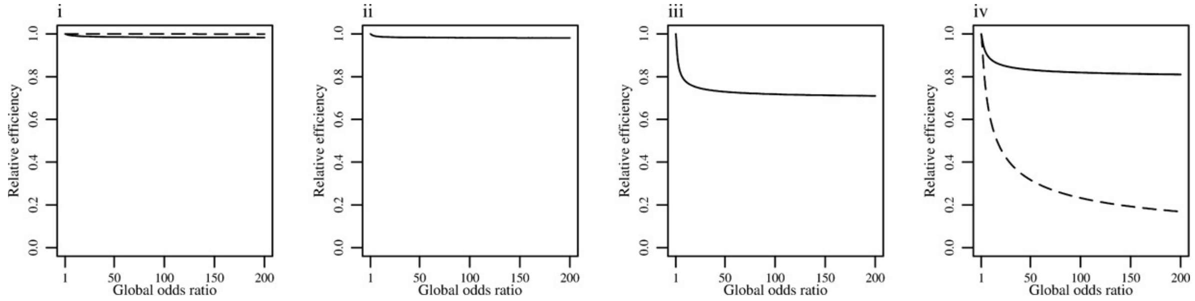

Fig. 5. Asymptotic relative efficiency of independence to exchangeable estimator for situations with unequal cluster sizes. Solid lines correspond to θ_c, and dashed lines to λ.

a cumulative probit model (6) for the marginal means, but varying the global odds ratio Ψ from 1 to 200. For all cases corresponding to the first scheme, the ARE was superior to 0.9 in general, showing that the independence estimator is almost as efficient as the exchangeable one for cluster-level and mean-balanced covariates and equal cluster sizes, without taking into account the strength of association. Fig. 4 presents the ARE of each element of $\widehat{\boldsymbol{\beta}}_I$ to $\widehat{\boldsymbol{\beta}}_{Ex}$ for situations (a) to (h) corresponding to (A'a) to (A'h) in the second configuration. It can be seen that the ARE of $\widehat{\lambda}_I$ to $\widehat{\lambda}_{Ex}$ is equal to one for independent observations ($\Psi = 1$) and it noticeably decreases as Ψ grows larger, with values lower than 0.4 for Ψ higher than 50. Therefore, there is an important efficiency loss when covariates vary within clusters but are not mean-balanced. Regarding the cutpoints, the ARE also declines as Ψ gets larger, but it is one when the covariates grand mean is zero (as in (e)), and generally high for mean value near to zero. This is in analogy with the intercept behaviour in GEE models for the case of constant weights, since it can be shown that relative efficiency is one for independent within-cluster responses, for cluster-level and mean-balanced covariates and also when the covariates grand mean is zero.

The ARE of $\widehat{\boldsymbol{\beta}}_I$ to $\widehat{\boldsymbol{\beta}}_{Ex}$ for the unequal cluster sizes situations is shown in Fig. 5. For mean-balanced covariates the ARE is approximately one and also for cluster-level covariates as in (ii). However, when the coefficient of variation in

Table 3
Relative efficiency of independence to exchangeable estimator for the second simulation scheme (non-mean-balanced within-cluster covariates)

Independence ($\rho = 0$)			
	θ_1	θ_2	λ
I'a	1.000	0.999	1.000
I'b	1.000	1.000	1.000
I'c	1.007	1.013	1.009
I'd	1.000	1.000	1.000
Association ($\rho \neq 0$)			
	θ_1	θ_2	λ
A'a	0.983	0.983	0.972
A'b	0.997	0.998	0.908
A'c	0.738	0.730	0.575
A'd	0.680	0.679	0.498
A'e	1.003	1.003	0.325
A'f	0.911	0.911	0.283
A'g	0.876	0.874	0.820
A'h	0.952	0.952	0.650

cluster sizes is greater, as in (iii) (with $T_i = 2$ in half of the clusters and $T_i = 10$ in the other half), the ARE declines as Ψ increases although covariates are constant within clusters. Finally, for non-mean-balanced within-cluster covariates, it can be seen that the ARE decreases as association gets stronger.

For all the situations presented above, similar results were obtained assuming an analogous logit cumulative model for the marginal means (results not shown).

4.2. Simulation study

We wanted to verify if the efficiency results obtained when the true association structure is exchangeable are also applicable to approximated situations, as in the case of the simulated data presented in Section 3. Then, for all the scenarios described in Section 3.1 and $K = 300$, we evaluated the relative efficiency for each parameter by calculating the ratio of the sample variances of the estimates obtained specifying exchangeable and independence working association structures. Estimators of $\boldsymbol{\beta}$ were little biased (the bias was generally in the order of 10^{-3} and 10^{-4} and only in a few situations around 10^{-2}) and therefore the efficiency values coincided with those calculated as the ratio of the sample mean squared errors. The standard error of the mean of $\boldsymbol{\beta}$ estimates was approximately 10^{-3} and 10^{-4} in general, with a maximum value of 0.078.

For the first configuration, the relative efficiency was approximately one in all situations both with cluster-level and mean-balanced within-cluster covariates.

The relative efficiencies of $\widehat{\boldsymbol{\beta}}_I$ to $\widehat{\boldsymbol{\beta}}_{Ex}$ for the second configuration are presented in Table 3. Clearly, there is an important reduction in the relative efficiency of $\widehat{\lambda}_I$ to $\widehat{\lambda}_{Ex}$ when association gets stronger. For independent data ($\Psi = 1$), efficiency is almost one. When Ψ is approximately two, efficiency is a little less than one, but for Ψ around four, efficiency is slightly higher than 0.9, descending to 0.6 for Ψ between 10 and 20. Finally, values lower than 0.33 can be seen for Ψ higher than 40. As above, it can be observed that relative efficiency of $\widehat{\boldsymbol{\theta}}_I$ to $\widehat{\boldsymbol{\theta}}_{Ex}$ depends on the association strength and on the covariates grand mean. For example, in (A'e) Ψ is around 50 but covariates mean is equal to zero, and thus efficiency is almost one. Note also that the values observed in simulations are similar to those presented in Fig. 4 for the corresponding situation and odds ratio value.

In the unequal cluster sizes case, $\widehat{\boldsymbol{\beta}}_I$ was almost as efficient as $\widehat{\boldsymbol{\beta}}_{Ex}$ in cases (i) and (ii), while for cluster-level covariates with higher coefficient of variation in cluster sizes (iii), the efficiency was around 0.7 and for non-mean-balanced within-cluster covariates (iv), it was 0.81 for θ_1 and θ_2 and 0.261 for λ.

Relative efficiency was also studied for small sample sizes ($K = 20, 30, 50, 100$). In general it coincided that in the scenarios where empirical confidence levels differed strongly from nominal ones (see Section 3.2), there were problems in calculating efficiency for $K \neq 300$ because some $\boldsymbol{\beta}$ estimates were extremely large, mainly when using an exchangeable specification. However, when coverage probabilities were always around 0.95, lower sample sizes were needed to obtain efficiency values similar to those for $K = 300$ (and to the ARE values in Section 4.1). For example, in scenario (A'g) approximately the same efficiency values were observed for all sample sizes.

5. Discussion

In this article we studied some properties of the estimators of regression parameters in the GEE1 approach of Heagerty and Zeger (1996) for clustered ordinal data. We focused on this alternative since it includes specific models for ordinal data, both for marginal means and for association between responses. Furthermore, it uses centered variables $\mathbf{y}_i^* - \mathbf{p}_i$ in the estimating equation for $\boldsymbol{\alpha}$ instead of $\mathbf{y}_i^*$, as proposed by Williamson et al. (1995), which results in $\boldsymbol{\alpha}$ estimators that are more efficient and invariant to the codification of the response in the case of binary data (Heagerty and Zeger, 1996).

In the simulation study, the generation of ordinal data was based on the cumulative probit model formulation, which assumes an underlying normal variable. For $K = 300$, the estimation algorithm was successful for both working specifications. Some convergence problems appeared when sample size got smaller and they worsened when the probability of getting certain outcome was low within a group of clusters having the same covariate vector. From our experience, GEE with working independence structure converges much more frequently. Other authors also reported convergence difficulties for binary data (Lipsitz et al., 1991) and ordinal data with the association modelled through correlation between category indicator variables (Miller et al., 1993). In those simulation studies, GEE with an independence structure converged in all replicates, while GEE including some association structures encountered estimation problems for small sample sizes and/or high association.

The empirical confidence levels for regression parameters were approximately 95% for sample sizes greater than or equal to 100. The difference between empirical and nominal levels widened when sample size decreased and this rise was noticeable when the probability for a given response category was low for some t in a group of clusters with the same covariate vector. Therefore, some caution should be exercised when making inferences when a response category is not well represented in a dataset, especially with small size samples. In similar studies, but only for binary data, empirical confidence levels higher than 84% were obtained, for 50 (Carey et al., 1993) and 100 (Lipsitz et al., 1991) clusters. Furthermore, empirical confidence levels might not be related to either the working association structure used or the covariates design in the sense considered here (mean-balanced, cluster-level or not).

We also studied asymptotic efficiency of the independence estimator $\widehat{\boldsymbol{\beta}}_I$ relative to the exchangeable estimator $\widehat{\boldsymbol{\beta}}_{Ex}$, when the true association structure is exchangeable. We showed that, when responses are independent, when covariates are mean-balanced or when all covariates are constant within clusters, $\widehat{\lambda}_I$ maintains a high efficiency in relation to $\widehat{\lambda}_{Ex}$, whatever the strength of the association is. On the other hand, when covariates vary within clusters but are not mean-balanced, the ARE of $\widehat{\lambda}_I$ to $\widehat{\lambda}_{Ex}$ markedly declines as association becomes stronger. These results are also applicable to the unequal cluster size cases, except for cluster-level covariates, where efficiency declines as association increases when the coefficient of variation in cluster sizes gets higher. This extends results of Mancl and Leroux (1996) for the case of ordinal data. Regarding the cutpoints, another situation of high efficiency is when the covariates grand mean is zero; however, this is less relevant since in applications the interest usually focuses in λ.

Sutradhar and Das (1999, 2000) used the limiting value α_0 of the α estimator for the working specification. In our study, exchangeable was assumed as the true association structure, and independence specification does not require estimation of an association parameter α. Then, in this case, their approach coincides with that of Fitzmaurice (1995). In all these articles, the authors also conclude that independent observations within clusters and cluster-level covariates produce ARE equal to one for binary data, and they show that ARE declines with increasing association between the responses for normally distributed covariates with zero mean and intra-cluster correlation $r < 1$.

Our study then supports the affirmation, now also for the case of ordinal measurements, that the degree of efficiency depends not only on the strength of association but also on the covariate design. Moreover, the simulation study showed that the results described above are also applicable to data with an approximately exchangeable association structure.

Circumstances in which assuming independence leads to estimators nearly as efficient as exchangeable ones are very specific, thus in general it is expected that in practical situations there could be efficiency loss. Therefore, we agree with Fitzmaurice (1995) in that some effort should be made in modelling association even when it is considered a *nuisance* characteristic of the data.

Acknowledgements

We thank Dr. Oscar Bustos and the referees for their helpful comments. Nores' research was supported by fellowships provided by Agencia Nacional de Promoción Científica y Tecnológica and CONICET.

Appendix

Assuming a correctly specified exchangeable association structure (all global odds ratios equal to Ψ), a cumulative probit model (6) for the marginal means and the same covariate vector $\mathbf{x} = (x_1, x_2, x_3)'$ for every cluster, then var $\left(\widehat{\boldsymbol{\beta}}_I\right)$ and var $\left(\widehat{\boldsymbol{\beta}}_{Ex}\right)$, following (2) and (3), are given by:

$$\operatorname{var}\left(\widehat{\boldsymbol{\beta}}_I\right) = \frac{1}{K}\left(D'V_I^{-1}D\right)^{-1}\left(D'V_I^{-1}V_{Ex}V_I^{-1}D\right)\left(D'V_I^{-1}D\right)^{-1} \quad \text{and}$$

$$\operatorname{var}\left(\widehat{\boldsymbol{\beta}}_{Ex}\right) = \frac{1}{K}\left(D'V_{Ex}^{-1}D\right)^{-1}, \quad \text{where}$$

$$D = \begin{pmatrix} \varphi(\eta_{11}) & 0 & \varphi(\eta_{21}) & 0 & \varphi(\eta_{31}) & 0 \\ 0 & \varphi(\eta_{12}) & 0 & \varphi(\eta_{22}) & 0 & \varphi(\eta_{32}) \\ \varphi(\eta_{11})\,x_1 & \varphi(\eta_{12})\,x_1 & \varphi(\eta_{21})\,x_2 & \varphi(\eta_{22})\,x_2 & \varphi(\eta_{31})\,x_3 & \varphi(\eta_{32})\,x_3 \end{pmatrix}',$$

$$V_I = \begin{pmatrix} V_1 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & V_3 \end{pmatrix}, \qquad V_t = \begin{pmatrix} p_{t1}(1-p_{t1}) & p_{t2}(1-p_{t1}) \\ p_{t2}(1-p_{t1}) & p_{t2}(1-p_{t2}) \end{pmatrix},$$

$$\eta_{tc} = \theta_c + x_t\lambda,\ p_{tc} = \Phi(\eta_{tc})$$

and Φ and φ are, respectively, the standard normal distribution and density functions. The elements V_t of V_I are obtained considering that $\{y_{itc}\}$ are cumulative indicators of categories. V_{Ex} has the same block-diagonal elements of V_I, but the off-diagonal elements are replaced by $cov\left(y_{itc}, y_{it'c'}\right)$, for $t \neq t'$. After to Williamson et al. (1995), it can be shown that these values are given by

$$\frac{1}{2(\Psi-1)}\left[1 + \left(p_{tc} + p_{t'c'}\right)(\Psi-1) - Z_{(tt')(cc')}\right] - p_{tc}p_{t'c'},$$

where $Z_{(tt')(cc')} = \left\{\left[1 + \left(p_{tc} + p_{t'c'}\right)(\Psi-1)\right]^2 + 4\Psi(1-\Psi)\,p_{tc}p_{t'c'}\right\}^{1/2}$.

In the case of two groups of clusters with covariate vectors $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, var $\left(\widehat{\boldsymbol{\beta}}_{Ex}\right) = \frac{2}{K}(D_1'V_{Ex1}^{-1}D_1 + D_2'V_{Ex2}^{-1}D_2)^{-1}$, where D_j and V_{Exj} are as D and V_{Ex} above, but with $\mathbf{x}$ replaced by $\mathbf{x}^{(j)}$, $j = 1, 2$. In an analogous way, var $\left(\widehat{\boldsymbol{\beta}}_I\right)$ can be derived. This extends clearly to the case of cluster sizes different from three and unequal cluster sizes.

References

Agresti, A., 2002. Categorical Data Analysis, second ed. Wiley, New York.

Carey, V.J., Zeger, S.L., Diggle, P., 1993. Modelling multivariate binary data with alternating logistic regressions. Biometrika 80, 517–526.

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

Diggle, P.J., Heagerty, P.J., Liang, K.Y., Zeger, S.L., 2002. Analysis of Longitudinal Data, second ed. Oxford University Press, Oxford.

Fahrmeir, L., Tutz, G., 2001. Multivariate Statistical Modelling Based on Generalized Linear Models, second ed. Springer-Verlag, New York.

Fitzmaurice, G.M., 1995. A caveat concerning independence estimating equations with multivariate binary data. Biometrics 51, 309–317.

Genz, A., Bretz, F., Hothorn, T., 2006. Mvtnorm: Multivariate normal and *t* distribution. *R* package version 0.7–5.

Heagerty, P.J., Zeger, S.L., 1996. Marginal regression models for clustered ordinal measurements. J. Amer. Statist. Assoc. 91, 1024–1036.

Huang, G.H., Bandeen-Roche, K., Rubin, G.S., 2002. Building marginal models for multiple ordinal measurements. Appl. Statist. 51, 37–57.

Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.

Liang, K.Y., Zeger, S.L., Qaqish, B., 1992. Multivariate regression analyses for categorical data (with discussion). J. Roy. Statist. Soc. Ser. B 54, 3–40.

Lipsitz, S.R., Laird, N.M., Harrington, D.P., 1991. Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. Biometrika 78, 153–160.

Mancl, L.A., Leroux, B.G., 1996. Efficiency of regression estimates for clustered data. Biometrics 52, 500–511.

Miller, M.E., Davis, C.C., Landis, J.R., 1993. The analysis of longitudinal polytomous responses: Generalized estimating equations and connections with weighted least squares. Biometrics 49, 1033–1044.

Prentice, R.L., Zhao, L.P., 1991. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics 47, 825–839.

R Development Core Team, 2006. *R*: A language and environment for statistical computing. *R* Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org.

Singer, J.M., Andrade, D.F., 1986. Analise de dados longitudinais. Asociação Brasileira de Estatística, São Paulo.

Sutradhar, B.C., Das, K., 1999. On the efficiency of regression estimators in generalised linear models for longitudinal data. Biometrika 86 (2), 459–465.

Sutradhar, B.C., Das, K., 2000. On the accuracy of efficiency of estimating equation approach. Biometrics 56, 622–625.

Williamson, J.M., Kim, K.M., Lipsitz, S.R., 1995. Analyzing bivariate ordinal data using a global odds ratio. J. Amer. Statist. Assoc. 90, 1432–1437.

Yan, J., 2002. Geepack: Yet another package for generalized estimating equations. *R* News 2 (3), 12–14.

Yan, J., Fine, J., 2004. Estimating equations for association structures. Statist. Med. 23, 859–874.

Zhao, L.P., Prentice, R.L., 1990. Correlated binary regression using a quadratic exponential model. Biometrika 77, 642–648.

Ziegler, A., Kastner, C., Blettner, M., 1998. The generalised estimating equations: An annotated bibliography. Biometrical J. 40, 115–139.