Journal of the Korean Data & Information Science Society
2012, **23**(1), 209–218

한국데이터정보과학회지

Property of regression estimators in GEE models for ordinal responses

Hyun Yung Lee[1]

[1]Department of Mathematics Education, Silla University

Received 12 November 2011, revised 2 December 2011, accepted 14 December 2011

Abstract

The method of generalized estimating equations (GEEs) provides consistent estimates of the regression parameters in a marginal regression model for longitudinal data, even when the working correlation model is misspecified (Liang and Zeger, 1986). In this paper we compare the estimators of parameters in GEE approach. We consider two aspects: coverage probabilites and efficiency. We adopted to ordinal responses the results derived from binary outcomes.

Keywords: Generalized estimating equations, ordinal responses, parameter estimation, repeated measures.

1. Introduction

In medical research, disease can be measured with categorical data including a dichotomous, polychotomous or ordinal score for conditions such as the presence or absence of dysplasia or the severity of wheeze. Furthermore, the same observed unit can be assessed by different medical investigators.

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). Cho (2010) also recommended about the longitudinal data. 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. For the case of clustered ordinal data Nores and Diez (2008) investigated some properties of GEE according to working correlation structures. They compared the coverage probability of confidence interval and the efficiency in the sense of variance estimates. The asymptotic efficiency of a correctly specified exchangeable association structure relative to the independence was discussed.

In this paper we study some properties of the estimators of marginal mean parameters in the context of the GEE 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

[1] Full time lecturer, Department of Mathematics Education, Silla University, Pusan 617-736, Korea. E-mail: hylee@silla.ac.kr

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 100, 120 clusters. 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. Mancl and Leroux (1996) studied asymptotic efficiency by considering an independence working specification in relation to a correctly specified exchangeable association structure. Choi (2010) suggested a mixed-effects model for analyzing split-plot data when there is a repeated measures factor that affects on the response variable. And Choi (2008a, 2008b, 2008c) suggested a marginal probability model for analyzing repeated polytomous response data and binary response data.

This article is organized as follows. Section 2 describes the notations in GEE model for the ordinal responses. In Section 3 we show the simulation schemes used and the results concerning confidence levels. The final section discusses the results obtained.

2. Notations in GEE model for ordinal responses

Suppose that a longitudinal study consists of ordinal responses with $(J+1)$ categories and p-dimensional covariate vectors (Y_{it}, x_{it}), for $i = 1, \cdots, n$ and $t = 1, \cdots, n_i$, where Y_{it} denotes the observation for subject i at occasion t and the covariate vector $\boldsymbol{x}_{it}$ can be discrete or continuous. For simplicity we assume equal occasions, $n_i = T$. Denote Y_{it} as a vector of J indicator variables, $\boldsymbol{y}_{it} = (y_{it}^1, \cdots, y_{it}^J)'$ with $y_{it}^j = 1$ if response $Y_{it} = j$ and 0 otherwise. Let $\boldsymbol{\pi}_{it}$ and $\eta_{it}^{(j)}$ represent the vector of marginal probabilities and the marginal cumulative probabilities, respectively, where $\boldsymbol{\pi}_{it} = (\pi_{it}^{(1)}, \cdots, \pi_{it}^{(J)})'$ with $\pi_{it}^{(j)} = P(Y_{it} = j|\boldsymbol{x}_{it}) = P(Y_{it}^{(j)} = 1|\boldsymbol{x}_{it})$ and $\eta_{it}^{(j)} = P(Y_{it} \leq j|\boldsymbol{x}_{it}) = \sum_{k=1}^{j} \pi_{it}^{(k)}$. It can be straightforwardly shown that $E(\boldsymbol{y}_{it}) = \pi_{it}$ and $Var(\boldsymbol{y}_{it}) = \boldsymbol{V}_{it} = diag(\boldsymbol{\pi}_{it}) - \boldsymbol{\pi}_{it}\boldsymbol{\pi}_{it}^{'}$. The cumulative logit model with proportional odds assumption for describing the dependence of Y_{it} on $\boldsymbol{x}_{it}$ is given by

$$\text{logit}\left(\eta_{it}^{(j)}\right) = \log\left(\frac{\eta_{it}^{(j)}}{1-\eta_{it}^{(j)}}\right) = \lambda_j + \boldsymbol{x}_{it}'\beta, \text{ for } j = 2, \cdots, J \tag{2.1}$$

where the intercepts $\lambda_1, \cdots, \lambda_J$ satisfy $\lambda_1 \leq \cdots \leq \lambda_J$, is the vector of regression coefficients with $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$, and $\boldsymbol{\zeta}_{it}^{(j)}$ is the jth element of a J- dimensional linear predictor $\boldsymbol{\zeta}_{it} = (\zeta_{it}^{(1)}, \cdots, \zeta_{it}^{(J)})'$. Since $\pi_{it}^{(1)} = \eta_{it}^{(1)}$ and

$$\begin{aligned}
\pi_{it}^{(j)} &= \eta_{it}^{(j)} - \eta_{it}^{(j-1)} \\
&= P(Y_{it} = j|\boldsymbol{x}_{it}) - P(Y_{it} = j-1|\boldsymbol{x}_{it}) \\
&= \frac{\exp\left(\eta_{it}^{(j)}\right)}{1+\exp\left(\eta_{it}^{(j)}\right)} - \frac{\exp\left(\eta_{it}^{(j-1)}\right)}{1+\exp\left(\eta_{it}^{(j-1)}\right)}, \text{ for } j = 2, \cdots, J
\end{aligned}$$

the linear predictor ζ_{it} can be rewritten as $\boldsymbol{\zeta}_{it}' = \boldsymbol{Z}_{it}'\boldsymbol{\theta}$ with the parameter vector $\boldsymbol{\theta} =$

$(\lambda_1, \cdots, \lambda_J, \beta')'$ and the $J \times (J + p)$ design matrix

$$\boldsymbol{Z}'_{it} = \begin{bmatrix} 1 & & & \boldsymbol{X}'_{it} \\ & \ddots & & \boldsymbol{X}'_{it} \\ & & 1 & \boldsymbol{X}'_{it} \end{bmatrix}$$

A J-dimensional link function g connects π_{it} and the linear predictor $\boldsymbol{Z}'_{it}\boldsymbol{\theta}$ as $\boldsymbol{\pi}_{it} = g^{-1}(\boldsymbol{Z}'_{it}\boldsymbol{\theta})$. Denote the responses, the marginal probabilities and the design matrix for subject i as $\boldsymbol{Y}_i = (\boldsymbol{y}'_{i1}, \cdots, \boldsymbol{y}'_{iT})$, $\boldsymbol{\pi}_i = (\pi'_{i1}, \cdots, \pi'_{iT})$, and $\boldsymbol{Z}_i = (\boldsymbol{Z_{i1}}, \cdots, \boldsymbol{Z}_{iT})'_{TJ\times(J+P)}$, respectively. The multivariate generalized estimating equations proposed by Lipsitz *et al.* (1994) and Liang and Zeger (1986) for estimating θ is the solution to

$$\sum_{i=1}^{n} \boldsymbol{D}'_i \boldsymbol{V}_i'^{-1} (\boldsymbol{Y}_i - \pi_i) \tag{2.2}$$

where $\boldsymbol{D}_i = \partial\boldsymbol{\pi}_i/\partial\boldsymbol{\theta} = (\boldsymbol{D}'_{i1}, \cdots, D'_{iT})'$ with the jth row vector of D_{it} expressed by $(\partial\boldsymbol{\pi}_{it}^{(j)}/\partial\lambda_1, \cdots, \partial\boldsymbol{\pi}_{it}^{(j)}/\partial\beta_p)'$ for $j = 1, \cdots, J, t = 1, \cdots, T$, and $\boldsymbol{V}_i = \boldsymbol{A}_i^{1/2}\boldsymbol{R}_i(\alpha)\boldsymbol{A}_i^{1/2}$. Here $\boldsymbol{A}_i = diag(\boldsymbol{A}_{i1}, \cdots, \boldsymbol{A}_{iT})$ with diagonal block $\boldsymbol{A}_{iT} = diag(\pi_{it}^{(1)}(1-\pi_{it}^{(1)}), \cdots, \pi_{it}^{(J)}(1-\pi_{it}^{(J)}))$, the 'working correlation matrix $\mathbf{R}_i(\alpha)$ is the correlation of $\boldsymbol{Y}_\mathrm{i}$, and α is a vector of parameters involved in the working correlation structure. In general, $\boldsymbol{V}_i \neq Var(\boldsymbol{V}_i)$. Liang and Zeger (1986) proposed a 'working' correlation matrix to gain efficiency in estimating $\boldsymbol{\theta}$. The solution to (2.2) is a consistent estimate of $\boldsymbol{\theta}$ for a variety of settings of the $TJ\times TJ$ 'working' correlation matrices. Let $\mathbf{R}_i(\alpha)$ be expressed as follows:

$$\boldsymbol{R}_i(\alpha) = \begin{bmatrix} \boldsymbol{M}_{i1} & \boldsymbol{B}_{i12} & \cdots & \boldsymbol{B}_{iT1} \\ \boldsymbol{B}_{i21} & \boldsymbol{M}_{i2} & \cdots & \boldsymbol{B}_{iT2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_{iT1} & \boldsymbol{B}_{iT2} & \cdots & \boldsymbol{M}_{iT} \end{bmatrix}_{TJ\times TJ}$$

where the t^{th} diagonal block $\boldsymbol{M}_{it} = \boldsymbol{A}_{it}^{-1/2}\boldsymbol{V}_{it}\boldsymbol{A}_{it}^{-1/2}$ denoted throughout by π_i for $t = 1, \cdots, T$, and the off-diagonal matrix $\boldsymbol{B}_{ist} = \boldsymbol{A}_{is}^{-1/2}\mathrm{E}[(\boldsymbol{y}_{is} - \boldsymbol{\pi}_{is})(\boldsymbol{y}_{it} - \boldsymbol{\pi}_{it})']\ \mathbf{A}_{it}^{-1/2}$ parameterized by α for $s \neq t$. The correlations of longitudinal ordinal data include two parts. One is the correlation between the repeated responses for subject i specified by $\mathbf{R}_\mathrm{i}(\alpha)$ and the other is the correlation between ordinal categories specified by $\boldsymbol{B}_{ist}$. The estimation of θ based on the Fisher scoring algorithm and iterative proportional fitting can be obtained by the geepack package in R software. Once the estimate of the parameter vector $\hat{\theta}$ is derived, the standardized residual vector, $\widehat{\boldsymbol{e}}_i = (\widehat{\boldsymbol{e}}'_{i1}, \cdots, \widehat{\boldsymbol{e}}_{iT})'$, can be computed with the element vector $\widehat{\boldsymbol{e}}_{it} = \widehat{\boldsymbol{A}}^{-1/2}(\boldsymbol{y}_{it} - \widehat{\boldsymbol{\pi}}_{it})$, where $\widehat{\boldsymbol{\pi}}_{it} = g^{-1}(\widehat{\boldsymbol{Z}}'_{it}\widehat{\boldsymbol{\theta}})$ for $i = 1, \cdots, n; t = 1, \cdots, T$.

In this article, four 'working' correlation matrices, independence ($\boldsymbol{R}_\mathrm{i}(\alpha) = \boldsymbol{I}$, the identity matrix), AR(1) structure, exchangeable structure ($\boldsymbol{B}_{ist} = \boldsymbol{B}$) and unspecified structure

($\boldsymbol{B}_{ist} = \boldsymbol{B}$), are considered; namely

$$\boldsymbol{R}_{ind,i}(\alpha) = \boldsymbol{I}_{TJ\times TJ},\ \boldsymbol{R}_{ar(1),i}(\alpha) = \begin{bmatrix} \boldsymbol{M}_{i1} & \boldsymbol{L} & \boldsymbol{L}^2 & \cdots & \boldsymbol{L}^{T-1} \\ L & \boldsymbol{M}_{i2} & \boldsymbol{L} & \cdots & \boldsymbol{L}^{T-2} \\ \boldsymbol{L}^2 & \boldsymbol{L} & \boldsymbol{M}_{i3} & \cdots & L^{T-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{L}^{T-1} & \boldsymbol{L}^{T-2} & \boldsymbol{L}^{T-3} & \cdots & \boldsymbol{M}_{iT} \end{bmatrix},$$

$$\boldsymbol{R}_{ex,i}(\alpha) = \begin{bmatrix} \boldsymbol{M}_{i1} & \boldsymbol{B} & \cdots & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{M}_{i2} & \cdots & \boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B} & \boldsymbol{B} & \cdots & \boldsymbol{M}_{iT} \end{bmatrix} \text{ and } \boldsymbol{R}_{un,i}(\alpha) = \begin{bmatrix} \boldsymbol{M}_{i1} & \boldsymbol{B}_{12} & \cdots & \boldsymbol{B}_{1T} \\ B_{21} & \boldsymbol{M}_{i2} & \cdots & \boldsymbol{B}_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_{T1} & \boldsymbol{B}_{T2} & \cdots & \boldsymbol{M}_{iT} \end{bmatrix}$$

for $i = 1, \cdots, n$ as well as the corresponding off-diagonal matrices are estimated by

$$\widehat{\boldsymbol{L}} = \frac{\sum_{t=1}^{T-1}\sum_{i=1}^{n}\widehat{e}_{it}\widehat{e}'_{i,t+1}}{(T-1)(n-p)},\ \widehat{\boldsymbol{B}} = \frac{\sum_{i=1}^{n}\sum_{s<t}\widehat{e}_{is}\widehat{e}'_{it}}{[\frac{1}{2}\sum_{i=1}^{n}T(T-1)] - p},$$

$$\widehat{\boldsymbol{B}}_{st} = \frac{\sum_{i=1}^{n}\widehat{e}_{is}\widehat{e}'_{it}}{n-p} \text{ for } s \neq t = 1, \cdots, T.$$

3. Coverage probabilities of the regression estimators

3.1. Example

As an illustration, we present an analysis for the dataset from Costa *et al.* (2006), which consists of 48 patients suffering from rupture of the tendo achilles.

We assume the following GLM relationship

$$\text{logit}(P[Y \leq j]) = \alpha_j + \beta_1 t + \beta_2 x.$$

where t is time ($t = 1, 2, 3$) which is composed of 3 month, 6 month and 1 year and x is treat ($x = 1, 2$; operativelyand nonoperatively). The response variable is a status of activity ($j = 1, 2, 3$) those are normal sporting activity, walking, stair climbing and work activity.

Table 3.1 Estimated parameter estimates and their standard errors

Assumed working correlation	α_1	α_2	β_1	β_2
Indep	0.9980[1)]	-1.5969	2.5521	3.1640
	0.7464[2)]	1.0865	1.3157	0.7364
Exchangeable	-4.7937	-1.6097	2.5864	3.1386
	1.1675	1.0776	1.3003	0.7154
AR1	-4.7827	-1.5969	2.5521	3.1309
	1.1781	1.0865	1.3157	0.7212

1) GEE estimate 2) Srandard Error (S. E.) of GEE estimate

3.2. Simulation schemes

In order to study empirically the validity of inferences for the regression parameters in the GEE approach of Heagerty and Zeger (1996), we conducted a simulation study. In this section we present the study of the asymptotic efficiency of the estimator $\widehat{\beta}_I$, using an independence working specification, relative to $\widehat{\beta}_{EX}$, assuming a correctly specified exchangeable association structure. For each element of β, the asymptotic relative efficiency (ARE) is given by the ratio of the variance of the regression estimators.

To evaluate the performance of the proposed tests for the proportional odds model fit in terms of type I error rate and the power, the simulated longitudinal ordinal data are generated from the following models:

$$Model1;\ \text{logit}(P[Y \leq j]) = \alpha_j + \beta_1 D_{it} + \beta_2 x_{it},$$

where the monotone difference intercepts are assigned by $\alpha = (-1.0, 0, 0.5)'$, $\beta_1 = 0, \beta_2 = -0.1$; D_{it} equals to 0 for $i \leq n/2$ and 1 elsewhere, and $x_{it} \sim U(-1, 1)$ for $i = 1, \cdots, n$, $t = 1, 2, 3$, $j = 1, 2, 3$ and $n = (100, 120)$, $\rho = (0.5, 0.7)$. Here D_{1i} is time-stationary covariate and X_{it} is a time-dependent covariate. The pairwise correlations between the observations at the three occasions within a subject are assumed to be 0.5. The number of repitition is set to be 1,000.

The simulation study has been implemented through R software and its library functions. We may refer to a R Development Team (2006) for the R language and its environment.

3.3. Simulation results

The empirical coverage of confidence intervals for α_1, α_2, α_3, β_1 and β_2 are listed in Table 3.2 through Table 3.5 according to the assumed working correlation structure, the sample sizes, the nominal confidence level, and the correlation parameter ρ. The empirical coverages attain the nominal confidence levels when $n = (100, 120)$. When the true correlation structure is exchangeable with $\rho = 0.5$, the length of confidence intervals under the independence working correlation are wider than the others.

The results of exchangeable correlation and AR1 are very similar when the true correlation structure of repeated responses is exchangeable with ρ =0.5 or 0.7. As we see in Table 3.2 through Table 3.3 the lengths of confidence intervals for the exchangeable correlation structure are little shorter than those under AR1 when the true correlation structure is correctly assumed as exchangeable. On the other hand the empirical coverages of AR1 are sometimes better than those of exchangeable working correlation specification.

Table 3.2 Empirical coverages of confidence limits for α_1, α_2, α_3, β_1, and β_2 among 1000 repetitions when the true correlation structure is exchangeable with ρ =0. 5, 0. 7. (a) $\rho = 0.5$

n	Assumed working correlation	α	Confidence levels				
			α_1	α_2	α_3	β_1	β_2
100	Indep	90%	0.9980 1)	0.9960	0.9970	0.8590	0.8570
			0.7464 2)	0.7102	0.7319	1.0060	0.5771
		95%	1.0000	0.9980	1.0000	0.9990	0.8570
			0.8893	0.8462	0.8721	1.1990	0.6876
		99%	1.0000	0.9990	1.0000	1.0000	0.9990
			1.1710	1.1140	1.1480	1.5780	0.9051
	Exchangeable	90%	0.9980	0.9960	0.8570	0.9980	0.7130
			0.7459	0.7100	0.7328	1.0050	0.4643
		95%	1.0000	0.9980	1.0000	0.9990	0.9930
			0.8887	0.8459	0.8731	1.1970	0.5532
		99%	1.0000	0.9990	1.0000	1.0000	0.9980
			1.1700	1.1140	1.1490	1.5760	0.7281
	AR1	90%	0.9980	0.8560	0.8570	0.9980	0.7150
			0.7511	0.7144	0.7352	1.0080	0.4804
		95%	1.0000	0.9980	1.0000	0.9990	0.9960
			0.8949	0.8512	0.8760	1.2010	0.5723
		99%	1.0000	0.9990	1.0000	1.0000	0.9980
			1.1780	1.1200	1.1530	1.5810	0.7534
	Var Ratio	Indep vs Uniform	1.0012	1.0007	0.9977	1.0026	1.5452
		AR1vs Uniform	1.0140	1.0120	1.0070	1.0070	1.0710
120	Indep	90%	0.9260	0.8750	0.9300	0.9040	0.8360
			0.6832	0.6459	0.6592	0.9182	0.5566
		95%	0.9720	0.9560	0.9740	0.9350	0.9050
			0.8140	0.7696	0.7855	1.0940	0.6632
		99%	1.0000	0.9810	0.9990	0.9600	0.9950
			1.0720	1.0130	1.0340	1.4400	0.8730
	Exchangeable	90%	0.9070	0.8940	0.9320	0.9030	0.7540
			0.6841	0.6466	0.6598	0.9190	0.4507
		95%	0.9720	0.9560	0.9740	0.9350	0.8450
			0.8151	0.7704	0.7861	1.0950	0.5370
		99%	1.0000	0.9990	0.9990	0.9600	0.9920
			1.0730	1.0140	1.0350	1.4410	0.7069
	AR1	90%	0.9450	0.8540	0.9290	0.9040	0.8120
			0.6873	0.6491	0.6626	0.9219	0.4779
		95%	0.9720	0.9550	0.9730	0.9540	0.8410
			0.8190	0.7734	0.7895	1.0980	0.5694
		99%	0.9990	1.0000	0.9980	0.9790	0.9930
			1.0780	1.0180	1.0390	1.4460	0.7495
	Var Ratio	Indep vs Uniform	0.9974	0.9980	0.9983	0.9985	1.5249
		AR1vs Uniform	1.0090	1.0080	1.0090	1.0060	1.1240

1) Empirical coverage of confidence interval 2) Length of confidence interval

Table 3.3 Empirical coverages of confidence limits for α_1, α_2, α_3, β_1 and β_2 among 1000 repetitions when the true correlation structure is exchangeable with $\rho = 0.5, 0.7$. (b) $\rho = 0.7$

n	Assumed working correlation	α	Confidence levels				
			α_1	α_2	α_3	β_1	β_2
100	Indep	90%	0.9620 1)	0.9680	0.8750	0.9630	0.7490
			0.8153 2)	0.7738	0.7941	1.0980	0.5846
		95%	0.9840	0.9780	0.8930	0.9790	0.8670
			0.9714	0.9220	0.9492	1.3080	0.6966
		99%	0.9960	0.9940	0.9960	0.9900	0.8950
			1.2790	1.2140	1.2460	1.7220	0.9169
	Exchangeable	90%	0.9630	0.8770	0.8760	0.9620	0.5420
			0.8167	0.7752	0.7968	1.0990	0.3796
		95%	0.9830	0.9790	0.8920	0.9780	0.8500
			0.9731	0.9237	0.9493	1.3090	0.4522
		99%	0.9970	0.9940	0.9960	0.9890	0.8920
			1.2810	1.2160	1.2500	1.7230	0.5953
	AR1	90%	0.9650	0.8760	0.8780	0.9610	0.7310
			0.8229	0.7797	0.7995	1.1030	0.4112
		95%	0.9850	0.9790	0.8910	0.9770	0.8530
			0.9805	0.9290	0.9526	1.3140	0.4900
		99%	0.9970	0.9950	0.9960	0.9900	0.9770
			1.2910	1.2230	1.2540	1.7290	0.6450
	Var Ratio	Indep vs Uniform	0.9964	0.9963	0.9934	0.9988	2.3724
		AR1vs Uniform	1.0150	1.0120	1.0070	1.0070	1.1740
120	Exchangeable	90%	0.9550	0.9550	0.8870	0.9080	0.8190
			0.7487	0.7034	0.7244	1.0070	0.5569
		95%	0.9780	0.9770	0.9330	0.9310	0.8880
			0.8920	0.8381	0.8630	1.1990	0.6635
		99%	1.0000	1.0000	0.9770	0.9770	0.9780
			1.1740	1.1030	1.1360	1.5790	0.8734
	Uniform	90%	0.9550	0.9100	0.8870	0.9080	0.6840
			0.7496	0.7047	0.7258	1.0080	0.3656
		95%	0.9780	0.9770	0.9330	0.9310	0.7530
			0.8931	0.8396	0.8648	1.2010	0.4356
		99%	1.0000	1.0000	0.9770	0.9770	0.9780
			1.1760	1.1050	1.1380	1.5810	0.5734
	AR1	90%	0.9550	0.9100	0.8870	0.8850	0.7750
			0.7560	0.7095	0.7300	1.0120	0.3981
		95%	0.9780	1.0000	0.9330	0.9540	0.8210
			0.9008	0.8453	0.8698	1.2060	0.4743
		99%	1.0000	1.0000	1.0000	0.9770	0.9560
			1.1860	1.1130	1.1450	1.5870	0.6243
	Var Ratio	Indep vs Uniform	0.9976	0.9963	0.9959	0.9973	2.3197
		AR1vs Uniform	1.0170	1.0140	1.0120	1.0080	1.1850

1) Empirical coverage of confidence interval 2) Length of confidence interval

Table 3.4 Empirical coverages of confidence limits for α_1, α_2, α_3, β_1 and β_2 among 1000 repetitions when the true correlation structure is AR1 with $\rho = 0.5, 0.7$. (a) $\rho = 0.5$

n	Assumed working correlation	α	Confidence levels				
			α_1	α_2	α_3	β_1	β_2
100	Indep	90%	1.0000 1)	1.0000	0.8570	0.8570	0.5710
			0.7150 2)	0.6852	0.6949	0.9583	0.5722
		95%	1.0000	1.0000	1.0000	1.0000	0.8570
			0.8519	0.8165	0.8280	1.1420	0.6818
		99%	1.0000	1.0000	1.0000	1.0000	1.0000
			1.1210	1.0750	1.0900	1.5030	0.8974
	Exchangeable	90%	1.0000	0.8570	0.8570	0.8570	0.5710
			0.7144	0.6846	0.6955	0.9566	0.4932
		95%	1.0000	1.0000	1.0000	1.0000	1.0000
			0.8512	0.8157	0.8287	1.1400	0.5877
		99%	1.0000	1.0000	1.0000	1.0000	1.0000
			1.1200	1.0740	1.0910	1.5000	0.7736
	AR1	90%	1.0000	0.8570	0.8570	0.8570	0.5710
			0.7127	0.6816	0.6917	0.9480	0.4674
		95%	1.0000	1.0000	0.8570	1.0000	1.0000
			0.8491	0.8121	0.8241	1.1300	0.5569
		99%	1.0000	1.0000	1.0000	1.0000	1.0000
			1.1180	1.0690	1.0850	1.4870	0.7331
	Var Ratio	Indep vs Uniform	1.0070	1.0110	1.0090	1.0220	1.4990
		AR1vs Uniform	1.0050	1.0090	1.0110	1.0180	1.1140
120	Indep	90%	0.8690	0.8920	0.8480	0.9040	0.8580
			0.6570	0.6229	0.6353	0.8800	0.5516
		95%	0.9020	0.9250	0.9250	0.9370	0.9010
			0.7828	0.7422	0.7569	1.0490	0.6572
		99%	0.9790	0.9680	0.9780	0.9900	0.9890
			1.0300	0.9769	0.9963	1.3800	0.8651
	Exchangeable	90%	0.8690	0.8920	0.8370	0.9040	0.8580
			0.6570	0.6231	0.6358	0.8806	0.4790
		95%	0.9020	0.9250	0.9250	0.9260	0.9020
			0.7828	0.7424	0.7575	1.0490	0.5708
		99%	0.9790	0.9680	0.9890	0.9900	0.9680
			1.0300	0.9773	0.9972	1.3810	0.7513
	AR1	90%	0.8690	0.8920	0.8700	0.9040	0.8490
			0.6523	0.6182	0.6308	0.8727	0.4650
		95%	0.9130	0.9250	0.9140	0.9260	0.9020
			0.7772	0.7366	0.7516	1.0400	0.5540
		99%	0.9790	0.9780	0.9890	0.9900	0.9780
			1.0230	0.9696	0.9893	1.3690	0.7293
	Var Ratio	Indep vs Uniform	1.0140	1.0150	1.0140	1.0170	1.4070
		AR1vs Uniform	1.0140	1.0160	1.0160	1.0180	1.0610

1) Empirical coverage of confidence interval 2) Length of confidence interval

Table 3.5 Empirical coverages of confidence limits for α_1, α_2, α_3, β_1 and β_2 among 1000 repetitions when the true correlation structure is AR1 with $\rho = 0.5, 0.7$. (b) $\rho = 0.7$

n	Assumed working correlation	α	Confidence levels				
			α_1	α_2	α_3	β_1	β_2
100	Indep	90%	1.0000 1)	0.8570	0.8570	0.8570	0.7140
			0.7787 2)	0.7531	0.7703	1.0600	0.5761
		95%	1.0000	1.0000	0.8570	1.0000	0.8570
			0.9278	0.8974	0.9178	1.2630	0.6864
		99%	1.0000	1.0000	1.0000	1.0000	1.0000
			1.2210	1.1810	1.2080	1.6620	0.9035
	Exchangeable	90%	1.0000	0.8570	0.8570	0.8570	0.7140
			0.7791	0.7533	0.7720	1.0590	0.4124
		95%	1.0000	1.0000	0.8570	1.0000	0.8570
			0.9282	0.8976	0.9198	1.2620	0.4914
		99%	1.0000	1.0000	1.0000	1.0000	1.0000
			1.2220	1.1820	1.2110	1.6610	0.6469
	AR1	90%	1.0000	0.8570	0.8570	0.8570	0.7140
			0.7781	0.7497	0.7647	1.0480	0.3770
		95%	1.0000	0.8570	0.8570	1.0000	0.7140
			0.9271	0.8933	0.9112	1.2490	0.4492
		99%	1.0000	1.0000	1.0000	1.0000	1.0000
			1.2200	1.1760	1.1990	1.6440	0.5913
	Var Ratio	Indep vs Uniform	1.0020	1.0090	1.0150	1.0210	2.3350
		AR1vs Uniform	1.0020	1.0100	1.0190	1.0200	1.1970
120	Indep	90%	0.8630	0.8880	0.8610	0.8710	0.8460
			0.7285	0.6903	0.7051	0.9764	0.5528
		95%	0.9160	0.9240	0.9230	0.9240	0.9220
			0.8680	0.8225	0.8401	1.1630	0.6587
		99%	0.9880	0.9770	0.9880	1.0000	0.9870
			1.1430	1.0830	1.1060	1.5310	0.8670
	Exchangeable	90%	0.8640	0.8890	0.8510	0.8710	0.8680
			0.7290	0.6911	0.7060	0.9775	0.3980
		95%	0.9160	0.9250	0.9230	0.9230	0.9220
			0.8686	0.8234	0.8412	1.1650	0.4742
		99%	0.9880	0.9770	0.9880	1.0000	0.9670
			1.1430	1.0840	1.1070	1.5330	0.6242
	AR1	90%	0.8720	0.8880	0.8620	0.8720	0.7930
			0.7238	0.6857	0.7010	0.9687	0.3776
		95%	0.9060	0.9340	0.9330	0.9240	0.9220
			0.8624	0.8170	0.8352	1.1540	0.4499
		99%	0.9880	0.9670	0.9880	0.9990	0.9780
			1.1350	1.0750	1.0990	1.5190	0.5922
	Var Ratio	Indep vs Uniform	1.0130	1.0140	1.0120	1.0160	2.1430
		AR1vs Uniform	1.0140	1.0160	1.0140	1.0180	1.1110

1) Empirical coverage of confidence interval 2) Length of confidence interval

4. Concluding remarks

In this article we studied some properties of the estimators of regression parameters in the GEE 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 $y_i^* - p_i$ in the estimating equation for α instead of y_i^*, as proposed by Williamson *et al.* (1995), which results in estimators that are more efficient and invariant to the codification of the

response in the case of binary data (Heagerty and Zeger, 1996). We studied asymptotic efficiency of the independence, (Exchangeable, AR1) estimator $\widehat{\beta}_I$, $(\widehat{\beta}_{Ex}, \widehat{\beta}_{AR1})$ relative to the exchangeable estimator $\widehat{\beta}_{Ex}$, when the true association structure is exchangeable. And we studied asymptotic efficiency of the independence, (Exchangeable, AR1) estimator $\widehat{\beta}_I$, $(\widehat{\beta}_{Ex}, \widehat{\beta}_{AR1})$ relative to the AR1 estimator $\widehat{\beta}_{AR1}$, when the true association structure is AR1. When the variance ratio is larger than 1, the true association structure is better than the fitted association structure. Because in the majority of cases the variance ratio of the regression estimates of the GEE model is larger than 1, the exchangeable association structure is recommended. But in the simulation design the case of the repeated time $T = 4$ was not considered owing to the running time of simulation. And the case of unspecified working correlation matrix was not considered because it requires so many parameters.

References

Agresti, A. (2002). *Categorical data analysis*, 2nd ed., Wiley, New York.

Cho, D. H. (2010). Mixed-effects LS-SVR for longitudinal data. *Journal of the Korean Data & Information Science Society*, **21**, 363-369.

Choi, J. (2008a). A logit model for repeated binary response data. *The Korean Journal of Applied Statistics*, **21**, 291-299.

Choi, J. (2008b). A marginal logit mixed-effects model for repeated binary response data. *Journal of the Korean Data & Information Science Society*, **19**, 413-420.

Choi, J (2008c). A marginal probability model for repeated polytomous response data. *Journal of the Korean Data & Information Science Society*, **19**, 577-585.

Choi, J. (2010). A mixed model for repeated split-plot data. *Journal of the Korean Data & Information Science Society*, **21**, 1-9.

Diggle, P. J., Heagerty, P. J., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of longitudinal data*, 2nd ed., Oxford University Press, Oxford.

Heagerty, P. J. and Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association*, **91**, 1024-1036.

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

Lipsitz, S. R., Laird, N. M. and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika*, **78**, 153-60.

Mancl, L. A. and Leroux, B. G. (1996). Efficiency of regression estimates for clusterd data. *Biometrics*, **52**, 500-511.

Nores, M. L. and Diez, M. P. (2008). Some properties of regression estimators in GEE models for clustered ordinal data. *Computational Statistics & Data Analysis*, **52**, 3877-3888.

R Development 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. and Andrade, D. F. (1986). Analise de dados longitudinais. *Associacao Brasileira de Estatistica*, Sao Paulo, 106.

Yan, J. and Fine, J. (2004). Estimating equations for association structures. *Statistic in Medicine*, **23**, 859-874.

Ziegler, A., Kastner, C. and Blettner, M. (1998). The generalised estimating equations: An annotated bibliography. *Biometrical Journal*, **40**, 115-139.