DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 5 | Issue 1 Article 18

5-1-2006

Variance Estimation and Construction of Confidence Intervals for GEE Estimator

Shenghai Zhang
Public Health Agency of Canada, s.zhang@rogers.com

Mary E. Thompson
University of Waterloo

Journal of Modern Applied Statistical Methods
May, 2006, Vol. 5, No. 1, 217-229

1538 – 9472/06/$95.00

Variance Estimation and Construction of Confidence Intervals for GEE Estimator

Shenghai Zhang
Centre for Infectious Disease Prevention and Control
Public Health Agency of Canada

Mary E. Thompson
Department of Statistics & Actuarial Sciences
University of Waterloo

The sandwich estimator, also known as the robust covariance matrix estimator, has achieved increasing use in the statistical literature as well as with the growing popularity of generalized estimating equations (GEE). A modified sandwich variance estimator is proposed, and its consistency and efficiency are studied. It is compared with other variance estimators, such as a model based estimator, the sandwich estimator and a corrected sandwich estimator. Confidence intervals for regression parameters based on these estimators are discussed. Simulation studies using clustered data to compare the performance of variance estimators are reported.

Key words: Generalized estimating equation, sandwich estimator, bias corrected estimator, variance-covariance matrix

Introduction

Once the estimators of regression parameters are obtained from a generalized estimating equation (GEE) (see Diggle, Liang & Zeger,1994; Liang & Zeger,1986), one needs the variance estimator to conduct inferences about the parameters. The sandwich estimator, also known as the robust covariance matrix estimator, has been used to achieve this goal. Its virtue is that it provides consistent estimates of the covariance matrix for parameter estimates even if the correlation structure in the parametric model is mis-specified. However, the properties of the sandwich method, other than consistency, had been little discussed until Kauermann and Carroll (2001). Further discussion about the properties will be provided, as well as a new variance estimator. This will be compared with other variance estimators: (a) a model based estimator, (b) the sandwich estimator, and (c) a corrected sandwich estimator.

Dr. Shenghai Zhang is Senior Biostatistician. Email: s.zhang@rogers.com. Mary E. Thompson is Professor of Statistics. The authors thank Professors J. F. Lawless, V. P. Godambe, P. Song and C. Wu at University of Waterloo for their valuable suggestions.

Estimation of $\text{cov}(Y_i)$ will be discussed first, where $Y_i = (y_{i1}, \cdots, y_{im})^T$ is a vector of repeated measurements taken on the ith subject; associated with each measurement y_{ij} is a vector of covariates $x_{ij} = (x_{ij1}, \cdots, x_{ijp})^T$ $(1 \le j \le m, \quad 1 \le i \le n)$. The mean of the marginal distribution of y_{ij} is denoted by μ_{ij}. It is assumed that Y_i and Y_k are independent vectors for all $i \ne k$. A bias reduced variance estimator will be provided next, and its consistency and efficiency will be discussed. Also, methods of constructing confidence intervals based on the variance estimators will be discussed. The simulation studies using clustered data to compare the performance of variance estimators will be reported.

Estimating Covariance

The main parameter of interest is $\beta = (\beta_1, \cdots, \beta_p)^T$, where β, covariates x_{ij} and the mean μ_{ij} of the marginal distribution are connected by a link function $h(.)$. The variance $\text{var}(y_{ij}) = \phi^{-1} v(\mu_{ij})$, where $v(.)$ is a known function, and where ϕ is a dispersion scalar that is either unknown or a known constant. Let $R(\alpha)$ be a $m\times m$ symmetric matrix which is a 'working' correlation matrix. The estimation of the nuisance parameter α will not be discussed and will be assumed to be known. The results could be generalized to the estimated $\hat{\alpha}$ of the α. Let

$$\eta_{ij} = x_{ij}^T \beta .$$

Then

$$\mu_{ij} = h(\eta_{ij}) ,$$

and

$$X_i = (x_{i1}, \cdots, x_{im})^T ;$$

$$A_i = diag(h'(\eta_{ij})) ;$$

$$\Gamma_i = diag(v(h(\eta_{ij}))) \tag{1}$$

are matrices with order $m\times p$, $m\times m$ and $m\times m$ respectively. It is well known that the general estimating function is defined as the following (Liang & Zeger, 1986):

$$g_n(\beta,\alpha) = \frac{1}{n}\sum_{i=1}^{n}[D_i]^T[V_i]^{-1}S_i , \tag{2}$$

where

$$V_i = V_i(\beta,\alpha) = \Gamma_i^{\frac{1}{2}} R_i(\alpha)\Gamma_i^{\frac{1}{2}} ;$$

$$D_i = D_i(\beta) = -\frac{\partial S_i}{\partial \beta^T} = A_i X_i ;$$

and

$$S_i = Y_i - (h(x_{ij}{}^T\beta), \cdots, h(x_{ij}{}^T\beta))^T$$
$$= (S_{i1}, \cdots, S_{im})^T .$$

$\phi^{-1}V_i$ was used to replace the true covariance $\text{cov}(Y_i)$ in the optimal estimating function linear in S_i. Because $\text{cov}(Y_i)$ is usually unknown, the estimation of $\text{cov}(Y_i)$ is first discussed. Typically, the residual estimator $(Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^T$ is used to estimate $\text{cov}(Y_i)$, where $\hat{\mu}_i = (\hat{\mu}_{i1}, \cdots, \hat{\mu}_{im})$ is the vector of fitted values based on the estimated parameters $\hat{\beta}_{GEE}$ obtained by solving equation $g_n(\beta,\alpha) = 0$. Because the fitted values tend to be closer to the observed values than the true values are, the residuals tend to be too small. Therefore, $\text{cov}(Y_i)$ tends to be underestimated by this method. To reduce the bias in general, another estimator of $\text{cov}(Y_i)$ will be proposed.

Considering a first-order Taylor series expansion of $\hat{\mu}_i = \mu_i(\hat{\beta}_{GEE})$ at the true parameter β_0, one has the following expressions:

$$\begin{aligned} S_i(\hat{\beta}_{GEE}) &\equiv Y_i - \hat{\mu}_i \\ &= Y_i - \mu_i(\beta_0) - \frac{\partial \mu_i}{\partial \beta^T}(\hat{\beta}_{GEE} - \beta_0) - O_p(n^{-1}) \\ &= S_i(\beta_0) - D_i(\beta_0)(\hat{\beta}_{GEE} - \beta_0) - O_p(n^{-1}). \end{aligned} \tag{3}$$

Based on an expansion for $\hat{\beta}_{GEE} - \beta_0$ (see Zhang , 2003),

$$\begin{aligned} &(Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^T = S_i(\beta_0)[S_i(\beta_0)]^T \\ &-S_i(\beta_0)H_i(\beta_0,\alpha) - [H_i(\beta_0,\alpha)]^T[S_i(\beta_0)]^T \\ &+[H_i(\beta_0,\alpha)]^T H_i(\beta_0,\alpha) \\ &+O_p(n^{-\frac{3}{2}}) \end{aligned} \tag{4}$$

where

$$\begin{aligned} &H_i(\beta_0,\alpha) \\ &= \frac{1}{n}\sum_{k=1}^{n} S_k{}^T V_k{}^{-1} D_k(\dot{g}_{n,0})^{-1} D_i{}^T \Big|_{(\beta_0,\alpha)} , \end{aligned} \tag{5}$$

and

$$\dot{g}_{n,0}(\beta,\alpha) = \frac{1}{n}\sum_{i=1}^{n} D_i^{\ T}[V_i]^{-1} D_i \Big|_{(\beta_0,\alpha)}. \tag{6}$$

$f\big|_{(\beta,\alpha)}$ is used to denote the value of a function f at (β,α). For example,

$$\frac{1}{n}\sum_{i=1}^{n} D_i^{T}[V_i]^{-1} D_i \Big|_{(\beta_0,\alpha)} \equiv \frac{1}{n}\sum_{i=1}^{n}[D_i(\beta_0)]^T[V_i(\beta_0,\alpha)]^{-1}D_i(\beta_0).$$

Taking expectation on both sides of (4), under certain integral conditions,

$$\begin{aligned} &\mathrm{E}[(Y_i-\hat{\mu}_i)(Y_i-\hat{\mu}_i)^T] \\ &= \{\mathrm{cov}(Y_i) - n^{-1}\mathrm{cov}(Y_i)h_{ii} - n^{-1}h_{ii}^{\ T}\mathrm{cov}(Y_i) \\ &+ n^{-2}h_{ii}^{\ T}\mathrm{cov}(Y_i)h_{ii} \\ &+ n^{-2}\sum_{k=1,k\neq i}^{n} h_{ki}^{\ T}\mathrm{cov}(Y_k)h_{ki}\}\Big|_{(\beta_0,\alpha)} + O(n^{-\frac{3}{2}}) \\ &= \{(I_i - n^{-1}h_{ii}^{\ T})\mathrm{cov}(Y_i)(I_i - n^{-1}h_{ii}) \\ &+ n^{-2}\sum_{k=1,k\neq i}^{n} h_{ki}^{\ T}\mathrm{cov}(Y_k)h_{ki}\}\Big|_{(\beta_0,\alpha)} + O(n^{-\frac{3}{2}}) \end{aligned}$$

where

$$h_{ki} = [V_k]^{-1}D_k[\dot{g}_{n,0}]^{-1}D_i^{\ T}, \tag{7}$$

for $i,k = 1,\cdots,n$, and I_i is an identity matrix of the same dimension as that of h_{ii}. An alternative estimator for $\mathrm{cov}(Y_i)$ was proposed by Mancl and DeRouen (2001) that is intended to compensate for the bias of the residual estimator in hypothesis testing: $\mathrm{cov}(Y_i)$ could be estimated by

$$(I_i - n^{-1}\hat{h}_{ii}^{\ T})^{-1}\hat{S}_i[\hat{S}_i]^T(I_i - n^{-1}\hat{h}_{ii})^{-1}, \tag{8}$$

under the assumption that

$$n^{-2}\sum_{k=1,k\neq i}^{n} h_{ki}^{\ T}\mathrm{cov}(Y_k)h_{ki}$$

is negligible. Let $\hat{h}_{ii} = h_{ii}(\hat{\beta}_{GEE},\alpha)$ and $\hat{S}_i = S_i(\hat{\beta}_{GEE})$. It is hard to tell whether (8) is a good estimator, because the assumption is not always reasonable. If $R(\alpha)$ correctly specifies the correlation structure, the expectation of the estimator defined by (8) has the following expression:

$$\begin{aligned} &\mathrm{cov}(Y_i)\big|_{(\beta_0,\alpha)} + \{(I_i - n^{-1}h_{ii}^{\ T})^{-1}n^{-1} \\ &D_i[\dot{g}_{n,0}]^{-1}D_i^{\ T}(I_i - n^{-1}h_{ii})^{-1}\}\big|_{(\beta_0,\alpha)} \\ &+ O(n^{-\frac{3}{2}}) \end{aligned}$$

and the estimator is biased upwards with order $O(n^{-1})$. This makes it more conservative than the residual estimation. For the residual estimator of $\mathrm{cov}(Y_i)$,

$$\begin{aligned} &\mathrm{E}[(Y_i-\hat{\mu}_i)(Y_i-\hat{\mu}_i)^T] \\ &= \mathrm{cov}(Y_i)\big|_{(\beta_0,\alpha)} - n^{-1}D_i[\dot{g}_{n,0}]^{-1}D_i^{T}\big|_{(\beta_0,\alpha)}. \\ &+ O_p(n^{-\frac{3}{2}}) \end{aligned} \tag{9}$$

Because

$$n^{-1}D_i(\beta_0)[\dot{g}_{n,0}(\beta_0,\alpha)]^{-1}[D_i(\beta_0)]^T$$

is positively definite, the residual estimator appears to be biased downward with order $O(n^{-1})$.

If the parameter values were known, one could use the following covariance estimator of the $\mathrm{cov}(Y_i)$:

$$\hat{\mathrm{cov}}(\mathrm{Y}_i)_c = \{(I_i - n^{-1}h_{ii}{}^T)^{-1} S_i{}^c (I_i - n^{-1}h_{ii})^{-1}\}\Big|_{(\beta_0,\alpha)} \quad (10)$$

where

$$S_i{}^c = (Y_i - \mu_i)(Y_i - \mu_i)^T - n^{-2}\sum_{k=1,k\neq i}^{n} h_{ki}{}^T (Y_k - \mu_k)(Y_k - \mu_k)^T h_{ki}.$$

The notation $\hat{\mathrm{cov}}(Y_i)$ in (10) means an estimation of the $\mathrm{cov}(Y_i)$. In this case, the first order asymptotic bias disappears, because

$$\mathrm{E}[\hat{\mathrm{cov}}(Y_i)_c] = \mathrm{cov}(Y_i) + O(n^{-\frac{3}{2}}).$$

Therefore, if the covariance estimator (10) was able to be used, the first order bias reduction would hold even if the correlation structure were not correctly specified. In practice, plug-in estimates are proposed

$$\hat{h}_{ik} = h_{ik}(\hat{\beta}_{GEE}, \alpha)$$

and

$$\hat{\mu}_i = \mu_i(\hat{\beta}_{GEE})$$

to get $\hat{\mathrm{cov}}(Y_i)_c$.

If there is a common correlation structure $R(\alpha) = R_i(\alpha) = corr(Y_i)$, observations are pooled across different clusters to estimate $R(\alpha)$ by

$$\hat{R} = \frac{\phi}{n}\sum_{i=1}^{n} \Gamma_i{}^{-\frac{1}{2}} \hat{\mathrm{cov}}(Y_i)_c \Gamma_i{}^{-\frac{1}{2}}, \quad (11)$$

where $\hat{\mathrm{cov}}(Y_i)_c$ and Γ_i are the same as before. The estimator $\hat{R}$ is similar to Liang and Zeger's suggestion for estimation of correlation structure (see Zeger & Liang, 1992; Zhao & Prentice, 1990; Fahrmeir & Tutz, 2001). Once estimation of the correlation matrix R is obtained, then, the $\mathrm{cov}(Y_i)$ may be estimated by another way (also see Pan, 2001):

$$\hat{\mathrm{cov}}(Y_i)_{new} = \phi^{-1}\Gamma_i^{\frac{1}{2}}\hat{R}\Gamma_i^{\frac{1}{2}} = \Gamma_i^{\frac{1}{2}}[\frac{1}{n}\sum_{k=1}^{n}\Gamma_k{}^{-\frac{1}{2}}\hat{\mathrm{cov}}(Y_k)_c\Gamma_k{}^{-\frac{1}{2}}]\Gamma_i^{\frac{1}{2}}. \quad (12)$$

The $\hat{\mathrm{cov}}(Y_i)_{new}$ is a consistent estimator of $\mathrm{cov}(Y_i)$.

If there is not a common correlation structure $R(\alpha)$ across all clusters, one may classify clusters into several groups such that all subjects in the same group have the same correlation structure, and then apply (12) to obtain a correlation matrix for that group.

Estimating Covariance Matrix Of GEE Estimator

It is known that the covariance matrix of the estimator $\hat{\beta}_{GEE}$ has the following approximation:

$$\mathrm{cov}(\hat{\beta}_{GEE}) \approx \frac{1}{n^2}\{[\dot{g}_{n,0}]^{-1}\sum_{i=1}^{n} D_i{}^T V_i{}^{-1}\mathrm{cov}(Y_i) V_i{}^{-1} D_i [\dot{g}_{n,0}]^{-1}\}\Big|_{(\beta_0,\alpha)}. \quad (13)$$

If the $R(\alpha)$ is correctly specified, that is, if

$$\mathrm{cov}(Y_i) = \phi^{-1}\Gamma_i^{\frac{1}{2}} R_i(\alpha)\Gamma_i^{\frac{1}{2}}$$

then the first order approximation to $\mathrm{cov}(\hat{\beta}_{GEE})$ is $n^{-1}\phi^{-1}[\dot{g}_{n,0}]^{-1}\big|_{(\beta_0,\alpha)}$. So, one can estimate $\mathrm{cov}(\hat{\beta}_{GEE})$ by

$$\hat{\mathrm{cov}}(\hat{\beta}_{GEE})_{model} = n^{-1}\hat{\phi}^{-1}\{[\dot{g}_{n,0}]^{-1}\Big|_{(\hat{\beta}_{GEE},\alpha)}\}. \quad (14)$$

The estimate $\hat{\phi}$ may be obtained by

$$\hat{\phi} = \frac{1}{nm}\sum \hat{Z}_i^T \hat{Z}_i$$

where $\hat{Z}_i = \Gamma_i^{-\frac{1}{2}}(Y_i - \hat{\mu}_i)$. It was suggested (see Chaganty, 1997) that the $\hat{\phi}$ can be replaced by $\hat{\phi}_{bc} = nm\hat{\phi}/(nm-p)$ if a bias-corrected estimate for ϕ is preferable. However, the correlation structure could be mis-specified, that is

$$\operatorname{cov}(Y_i) \neq \phi^{-1}\Gamma_i^{\frac{1}{2}} R_i(\alpha)\Gamma_i^{\frac{1}{2}},$$

because the correlation matrix may not be known in practice. In this case, it is well known that the variance $\operatorname{cov}(\hat{\beta}_{GEE})$ can be estimated consistently by the sandwich formula

$$\hat{\operatorname{cov}}(\hat{\beta}_{GEE})_{sand} = \{[\dot{g}_{n,0}]^{-1} \frac{1}{n^2}\sum_{i=1}^{n} D_i^T V_i^{-1}\varepsilon_i\varepsilon_i^T V_i^{-1} D_i [\dot{g}_{n,0}]^{-1}\}\Big|_{(\hat{\beta}_{GEE},\alpha)} \quad (15)$$

where $\varepsilon_i = (y_{i1} - \mu_{i1}, \cdots, y_{im} - \mu_{im})^T$ are the residuals. As previously discussed, estimating $\operatorname{cov}(Y_i)$ by fitted $\hat{\varepsilon}_i\hat{\varepsilon}_i^T$ ($\hat{\varepsilon}_i = \varepsilon_i\big|_{(\hat{\beta}_{GEE},\alpha)}$) could be biased downward. Thus, the sandwich estimate $\hat{\operatorname{cov}}(\hat{\beta}_{GEE})_{sand}$ will be biased downward for estimating $\operatorname{cov}(\hat{\beta}_{GEE})$. Recently, the bias corrected sandwich estimators have been provided by Mancl and DeRouen (2001) and Kauermann and Carroll (2001), where the estimation of $\operatorname{cov}(\hat{\beta}_{GEE})$ is obtained by replacing $\hat{\varepsilon}_i\hat{\varepsilon}_i^T$ by $\hat{\operatorname{cov}}(Y_i)_c$ defined by (10), that is

$$\hat{\operatorname{cov}}(\hat{\beta}_{GEE})_{sand_u} = [\dot{g}_{n,0}]^{-1} \frac{1}{n^2}\sum_{i=1}^{n} D_i^T V_i^{-1}\hat{\operatorname{cov}}(Y_i)_c V_i^{-1} D_i [\dot{g}_{n,0}]^{-1}\Big|_{(\hat{\beta}_{GEE},\alpha)} \quad (16)$$

Finally, if $\hat{\operatorname{cov}}(Y_i)_{new}$ is used, a more efficient sandwich estimator could be obtained:

$$\hat{\operatorname{cov}}(\hat{\beta}_{GEE})_{new} = \{[\dot{g}_{n,0}]^{-1} \frac{1}{n^2}\sum_{i=1}^{n} D_i^T V_i^{-1}\hat{\operatorname{cov}}(Y_i)_{new} V_i^{-1} D_i [\dot{g}_{n,0}]^{-1}\}\Big|_{(\hat{\beta}_{GEE},\alpha)} \quad (17)$$

Consider the following:

Theorem:

$$\operatorname{cov}(vec(\hat{\operatorname{cov}}(\hat{\beta}_{GEE})_{sand})) - \operatorname{cov}(vec(\hat{\operatorname{cov}}(\hat{\beta}_{GEE})_{new})) = \Omega_n - \delta_n$$

where Ω_n is nonnegative definite, δ_n has higher order than Ω_n and the operator "vec" is used to stack the columns of a matrix together to obtain a vector.

Proof: Because $\hat{\beta}_{GEE}$ is $\sqrt{n}$-consistent, expand $\hat{\operatorname{cov}}(\hat{\beta}_{GEE})_{new}$ and $\hat{\operatorname{cov}}(\hat{\beta}_{GEE})_{sand}$ at (β_0, α). Then, the following expansions are obtained:

$$\hat{\operatorname{cov}}(\hat{\beta}_{GEE})_{new} = \{[\dot{g}_{n,0}]^{-1}\frac{1}{n^2}\sum_{i=1}^{n} D_i^T V_i^{-1}\Gamma_i^{\frac{1}{2}} \frac{1}{n}\sum_{k=1}^{n}\Gamma_k^{-\frac{1}{2}} H_{ii}^T \varepsilon_k\varepsilon_k^T H_{ii}\Gamma_k^{-\frac{1}{2}} \Gamma_i^{\frac{1}{2}} V_i^{-1} D_i [\dot{g}_{n,0}]^{-1}\}\Big|_{(\beta_0,\alpha)} + O_p(n^{-\frac{3}{2}})$$

where $H_{ii} = (I - n^{-1}h_{ii})^{-1}$. Similarly,

$$\hat{\text{cov}}(\hat{\beta}_{GEE})_{sand}$$
$$=\{[\dot{g}_{n,0}]^{-1}\frac{1}{n^2}\sum_{i=1}^{n}D_i^{\ T}V_i^{-1}\varepsilon_i$$
$$\varepsilon_i^{\ T}V_i^{-1}D_i[\dot{g}_{n,0}]^{-1}\}\Big|_{(\beta_0,\alpha)}+O_p(n^{-\frac{3}{2}})$$

By Theorem 7.16 in Schott (1997),

$$vec(\hat{\text{cov}}(\hat{\beta}_{GEE})_{sand})$$
$$\approx\frac{1}{n^2}\sum_{i=1}^{n}\{A_{i,n}vec(\varepsilon_i\varepsilon_i^{\ T})\}\Big|_{(\beta_0,\alpha)} \tag{18}$$

and

$$vec(\hat{\text{cov}}(\hat{\beta}_{GEE})_{new})$$
$$\approx\frac{1}{n^2}\sum_{i=1}^{n}\{A_{i,n}vec\{\Gamma_i^{\frac{1}{2}}\frac{1}{n}\sum_{k=1}^{n}\Gamma_k^{-\frac{1}{2}}H_{ii}^{\ T}\varepsilon_k$$
$$\varepsilon_k^{\ T}H_{ii}\Gamma_k^{-\frac{1}{2}}\Gamma_i^{\frac{1}{2}}\}\}\Big|_{(\beta_0,\alpha)}$$
$$=\frac{1}{n^2}\sum_{i=1}^{n}\{A_{i,n}\{\frac{1}{n}\sum_{k=1}^{n}B_{k,i}vec(\varepsilon_k\varepsilon_k^{\ T})\}\}\Big|_{(\beta_0,\alpha)}, \tag{19}$$

where

$$A_{i,n}=([\dot{g}_{n,0}]^{-1}D_iV_i^{-1})\otimes([\dot{g}_{n,0}]^{-1}D_iV_i^{-1})$$

and

$$B_{k,i}=(\Gamma_i^{\frac{1}{2}}\Gamma_k^{-\frac{1}{2}}H_{ii})\otimes(\Gamma_i^{\frac{1}{2}}\Gamma_k^{-\frac{1}{2}}H_{ii})\,.$$

The covariance matrices of

$$vec(\hat{\text{cov}}(\hat{\beta}_{GEE})_{new})$$

and

$$vec(\hat{\text{cov}}(\hat{\beta}_{GEE})_{sand})$$

can be obtained from (19) and (18):

$$\text{cov}(vec(\hat{\text{cov}}(\hat{\beta}_{GEE})_{sand}))$$
$$\approx\frac{1}{n^4}\sum_{i=1}^{n}\{A_{i,n}\,\text{cov}(vec(\varepsilon_i\varepsilon_i^T)A_{i,n}^{\ T}\}\Big|_{(\beta_0,\alpha)}$$

and

$$\text{cov}(vec(\hat{\text{cov}}(\hat{\beta}_{GEE})_{new}))$$
$$\approx\frac{1}{n^6}\sum_{i=1}^{n}\{\sum_{k=1}^{n}A_{i,n}B_{k,i}\,\text{cov}(vec(\varepsilon_k\varepsilon_k^{\ T})$$
$$B_{k,i}^{\ T}A_{i,n}^{\ T}\}\Big|_{(\beta_0,\alpha)}\,.$$

Notice that $vec(\varepsilon_i\varepsilon_i^{\ T})$ $(i=1,\cdots,n)$ are independent and free of n. It is clear that $\|B_{k,i}\|$ is bounded when $n\to\infty$. Hence, under some regularity conditions (see details in Zhang, 2003), there is the following result:

$$\frac{1}{n^2}\sum_{k=1}^{n}\{B_{k,i}\,\text{cov}(vec(\varepsilon_k\varepsilon_k^T))B_{k,i}^{\ T}\}\Big|_{(\beta_0,\alpha)}$$
$$=O(n^{-1}),$$

as $n\to\infty$. Finally,

$$\text{cov}(vec(\hat{\text{cov}}(\hat{\beta}_{GEE})_{sand}))$$
$$-\text{cov}(vec(\hat{\text{cov}}(\hat{\beta}_{GEE})_{new}))$$
$$=\frac{1}{n^4}\sum_{i=1}^{n}A_{i,n}(\text{cov}(vec(\varepsilon_i\varepsilon_i^T)$$
$$-O_p(n^{-1}))A_{i,n}^{\ T}\Big|_{(\beta_0,\alpha)}$$
$$=(\Omega_n-\delta_n)\Big|_{(\beta_0,\alpha)}\,.$$

where

$$\Omega_n=\frac{1}{n^4}\sum_{i=1}^{n}A_{i,n}\,\text{cov}(vec(\varepsilon_i\varepsilon_i^{\ T}))A_{i,n}^{\ T}$$

is a non-negative definite matrix and the δ_n has higher order of convergence to zero than Ω_n. Thus, it has been proven that

$$\text{cov}(vec(\hat{\text{cov}}(\hat{\beta}_{GEE})_{sand})) - \text{cov}(vec(\hat{\text{cov}}(\hat{\beta}_{GEE})_{new})) \geq 0$$

asymptotically. The proof of the Theorem is completed.

In summary, the estimator of the covariance matrix of regression parameters could gain some efficiency. Also it is expected that the method is more plausible for small sample sizes n than other estimators of the covariance.

For construction of confidence intervals, inference about $L^T\beta$ is of interest, where L^T is a $1\times p$ dimensional contrast vector of unit length, that is, $L^T L = 1$. If the $R(\alpha)$ is correctly specified, then the first-order of approximation of $\text{var}(L^T\hat{\beta}_{GEE})$ is $\phi^{-1}L^T[\dot{g}_{n,0}]^{-1}L\Big|_{(\beta_0,\alpha)}$. Thus, $\text{var}(L^T\hat{\beta}_{GEE})$ can be estimated by

$$\hat{\sigma}^2_{model} = \hat{\text{var}}(L^T\hat{\beta}_{GEE})_{model} = \frac{1}{n}\hat{\phi}^{-1}L^T[\dot{g}_{n,0}]^{-1}L\Big|_{(\hat{\beta}_{GEE},\alpha)} \tag{20}$$

Based on (20), a symmetric confidence interval is given by

$$(\hat{\theta} \pm z_q\hat{\sigma}_{model}), \tag{21}$$

where z_q is the q quantile of the standard normal distribution and $\hat{\theta} = L^T\hat{\beta}_{GEE}$.

Based on the estimation of the covariance matrix (17), if the $R_i(\alpha)$ is mis-specified, the variance $\text{var}(L^T\hat{\beta}_{GEE})$ can be estimated consistently by the sandwich formula

$$\hat{\sigma}^2_{sand} = \hat{\text{var}}(L^T\hat{\beta}_{GEE})_{sand} = L^T[\dot{g}_{n,0}]^{-1}\frac{1}{n^2}\{\sum_{i=1}^{n} D_i^T V_i^{-1}\varepsilon_i \varepsilon_i^T V_i^{-1} D_i\}[\dot{g}_{n,0}]^{-1}L\Big|_{(\hat{\beta}_{GEE},\alpha)} \tag{22}$$

where the ε_i's are the same as before. Then, based on (22), the symmetric confidence interval is given by

$$(\hat{\theta} \pm z_q\hat{\sigma}_{sand}). \tag{23}$$

It follows from the discussion that the sandwich estimate appears to be biased downward. Therefore, the bias corrected sandwich estimation of $\text{var}(L^T\hat{\beta}_{GEE})$ can be obtained by replacing $\varepsilon_i\varepsilon_i^T$ by $\text{cov}(Y_i)_c$ defined by (10). Thus, the bias reduced sandwich estimate of the variance $\text{var}(L^T\hat{\beta}_{GEE})$ is obtained by

$$\hat{\sigma}^2_{sand_u} = \hat{\text{var}}(L^T\hat{\beta}_{GEE})_{sand_u} = L^T[\dot{g}_{n,0}]^{-1}\frac{1}{n^2}\{\sum_{i=1}^{n} D_i^T V_i^{-1}\text{cov}(Y_i)_c V_i^{-1} D_i\}[\dot{g}_{n,0}]^{-1}L\Big|_{(\hat{\beta}_{GEE},\alpha)} \tag{24}$$

Corresponding to this estimate, another symmetric confidence interval is obtained

$$(\hat{\theta} \pm z_q\hat{\sigma}_{sand_u}). \tag{25}$$

$$\hat{\sigma}^2_{new} = \hat{\text{var}}(L^T\hat{\beta}_{GEE})_{new} = L^T\,\hat{\text{cov}}(\hat{\beta}_{GEE})_{new}L. \tag{26}$$

Then, a confidence interval is obtained:

$$(\hat{\theta} \pm z_q\hat{\sigma}_{new}). \tag{27}$$

Simulation Study and Discussions

Suppose that y_{ij} has marginally a negative binomial distribution, that is, $y_{ij} \sim NB(1, \mu_{ij})$, i =1,...,n and j =1,...,m. The link function is *log*, i.e. $\log(\mu_{ij}) = x_{ij}^T \beta$, where $\beta = (\beta_0, \beta_1, \beta_2)^T$ and $x_{ij} = (1, x_{ij1}, x_{ij2})^T$ are the covariates: $x_{ij2} \sim N(0,1)$ and x_{ij1} are constants. The correlation structure among $y_{i1}, \cdots, y_{im}$ is assumed to be given as an $AR(1)$ with ρ =0.8. Now, the procedures developed in the last two sections are applied to the model $\mathrm{E}(y_{ij}) = e^{x_{ij}^T \beta}$. The simulation study is completed for the number n of clusters as 10, 20, 30, $\cdots$, 90, 100 respectively.

A comparison of the performance of the estimators of the asymptotic variances is required. The estimators, $\hat{\sigma}^2_{model}(\hat{\beta}_{GEE})$, $\hat{\sigma}^2_{sand}(\hat{\beta}_{GEE})$, $\hat{\sigma}^2_{sand_u}(\hat{\beta}_{GEE})$, and $\hat{\sigma}^2_{new}(\hat{\beta}_{GEE})$, are defined by taking the vector L in an appropriate form in (20), (22), (24) and (26). Each of these variance estimators is related to a specified correlation structure $R_i(\alpha)$.

First, the situation is observed, where the $R_i(\alpha)$ in the estimators of variances are correctly specified to a constant. Figure 1 shows the comparisons of $\hat{\sigma}^2_{model}(\hat{\beta}_1)$, $\hat{\sigma}^2_{sand}(\hat{\beta}_1)$, $\hat{\sigma}^2_{sand_u}(\hat{\beta}_1)$, and $\hat{\sigma}^2_{new}(\hat{\beta}_1)$ and the true variance (empirical variance) $\mathrm{var}(\hat{\beta}_1)$ over 1000 simulations, when the regression parameters are estimated by the GEE estimator. From Figure 1, it is found that the estimator $\hat{\sigma}^2_{new}$ of the variance is better than other three, since the biases are smaller, even for the clusters with small sample size.

The curves shown in Figure 1 are consistent with the property that all four estimators are asymptotically unbiased. Notice that, in all these plots, the sandwich estimator $\hat{\sigma}^2_{sand}(\hat{\beta}_1)$ has the biggest bias when the sample size is small. It corresponds to the fact that the sandwich estimator would be expected to underestimate the variance of $\hat{\beta}_1$. It is not surprising that the model based estimator $\hat{\sigma}^2_{model}(\hat{\beta}_1)$ performs better than the sandwich estimator because the model is correct (the $R_i(\alpha)$ is correctly specified except for the constant α).

When the model is mis-specified, for example, if $R_i(\alpha)$ is an identity matrix, the model based estimator $\hat{\sigma}^2_{model}(\hat{\beta}_1)$ is the worst one. Figure 2 shows that (i) estimators $\hat{\sigma}^2_{sand}(\hat{\beta}_{GEE})$, $\hat{\sigma}^2_{sand_u}(\hat{\beta}_{GEE})$, and $\hat{\sigma}^2_{new}(\hat{\beta}_{GEE})$ are asymptotically unbiased; (ii) the $\hat{\sigma}^2_{model}(\hat{\beta}_{GEE})$ is significantly biased; (iii) the new estimator $\hat{\sigma}^2_{new}(\hat{\beta}_1)$ of the variance is the best one to estimate the $\mathrm{var}(\hat{\beta}_1)$.

Now, the efficiency of the variance estimators is compared. For Figure 3, the study is based on 1000 simulations for each number of clusters being 10, 20, $\cdots$, 100 respectively. The variances are calculated by

$$\mathrm{var}(\hat{\sigma}^2_{estimator}) = s^2_{estimator} ,$$

where $s^2_{estimator}$ is sample variance of values of $\hat{\sigma}^2_{estimator}$ which is obtained from the formula in the last section for each simulation. The estimator can be “model”, “sand”, “$sand_u$” and “new” respectively. Figure 3 illustrates that the corrected sandwich variance estimator $\hat{\sigma}^2_{sand}(\hat{\beta}_1)$ has the biggest standard error even for large sample size.

When the correlation structure is correctly specified, the model based estimator $\hat{\sigma}^2_{model}(\hat{\beta}_1)$ could be better than the corrected sandwich variance estimator, especially, when the sample size is small. When the number of clusters is greater than 30, the simulation shows that new variance estimator is the most stable one. It follows from Figure 4 that these facts still hold when the correlation structure is mis-specified in the variance estimators in the manner of the example. Of course, the model based variance estimator should not be used in this case because it is biased, although its variance is the smallest one. If the sample size is small, the sandwich estimator performs well.

With variance estimators at hand, confidence intervals could be constructed with different variance estimators. It will be seen that the confidence intervals obtained by the new variance estimator perform better than the other three in terms of coverage probability. The problem of testing a null hypothesis $H_0 : \beta \in \vartheta_0$ will be considered. Essentially, confidence intervals are closely related with tests. The aim is to compare CI's which are related to the various estimators introduced in the third sections of this article. In the simulation study, the CI for β_1 corresponds to a test that $H_0 : \beta_1 = \beta_{10}$. The test statistic could be $T_{new} = (\hat{\beta}_1 - \beta_{10}) / \hat{\sigma}_{new}(\hat{\beta}_1)$ or other ones obtained by different variance estimators. It follows from Figure 5 that the coverage percentages with the new variance estimator are bigger; therefore, the confidence interval based on the new variance estimator is accurate for smaller sample sizes than other ones with the variance estimators 'model', 'sand' or 'sand_u'.

It appears to be better to use the new variance estimator to construct confidence intervals, especially when the sample size is small. In the example of a mis-specified correlation structure in the variance estimators, the new and adjusted sandwich estimators both give accurate confidence intervals (see Figure 6). Again, the model based variance estimator should not be used in this case.

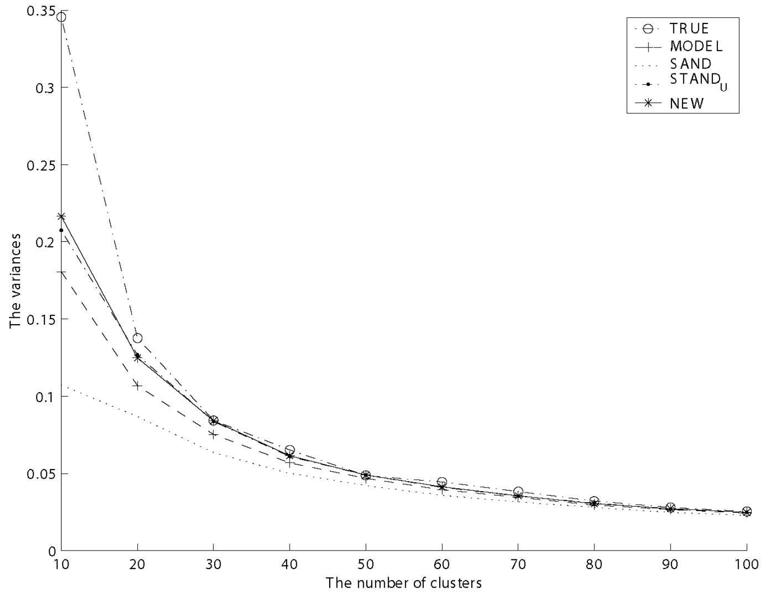

Figure 1

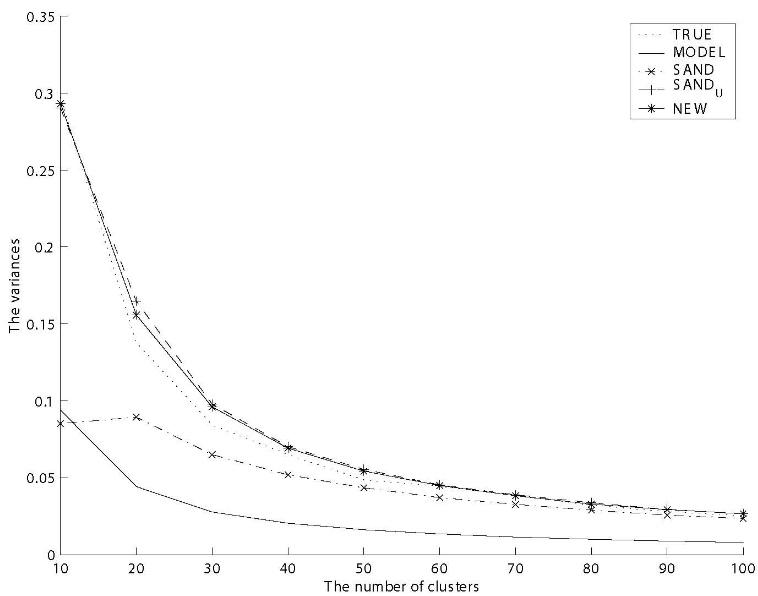

Figure 2

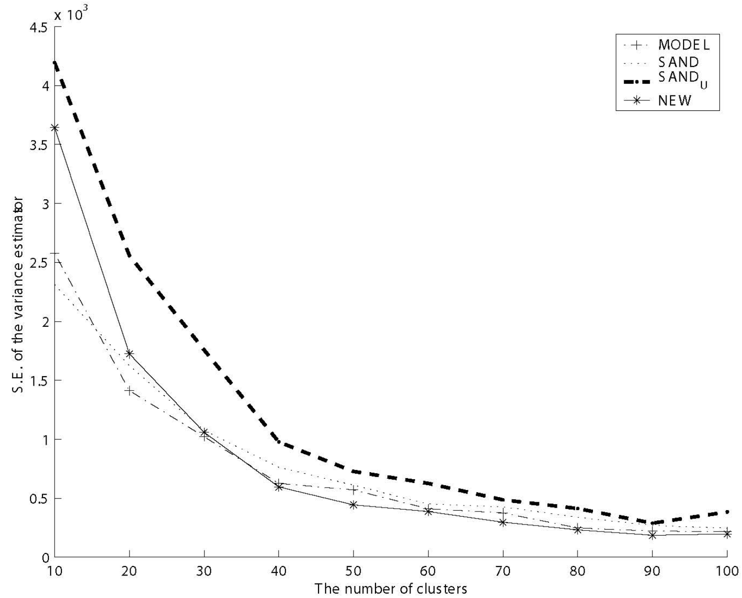

Figure 3

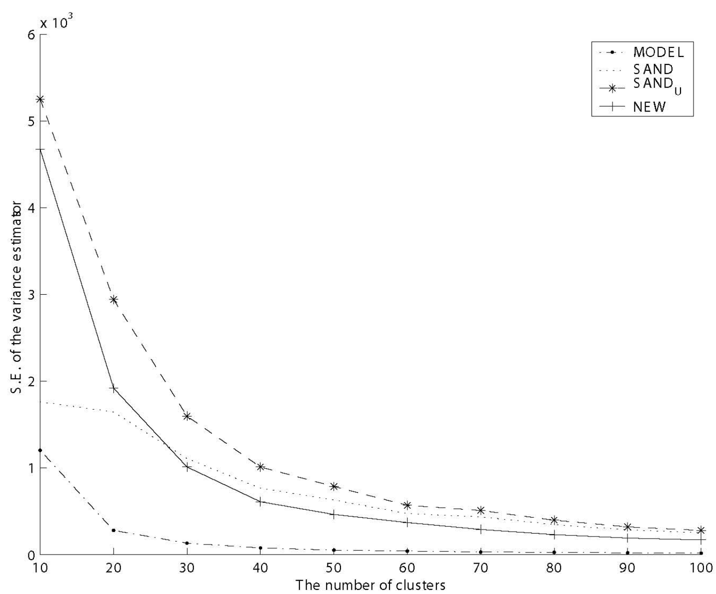

Figure 4

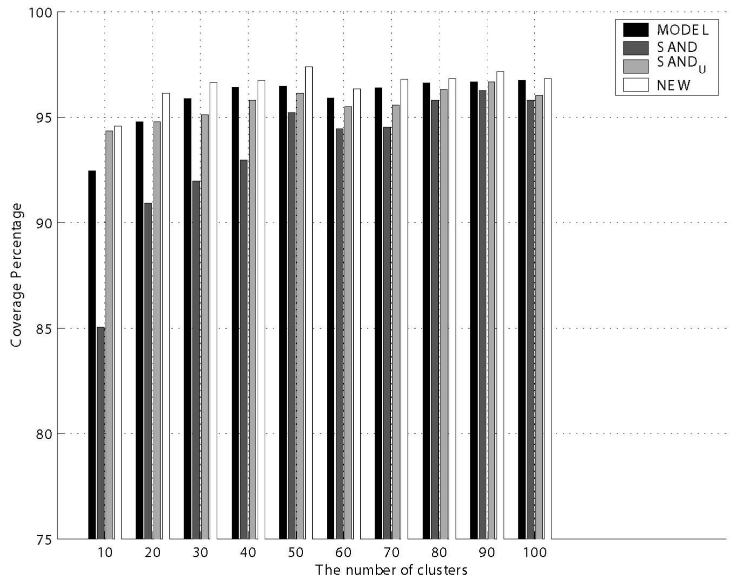

Figure 5

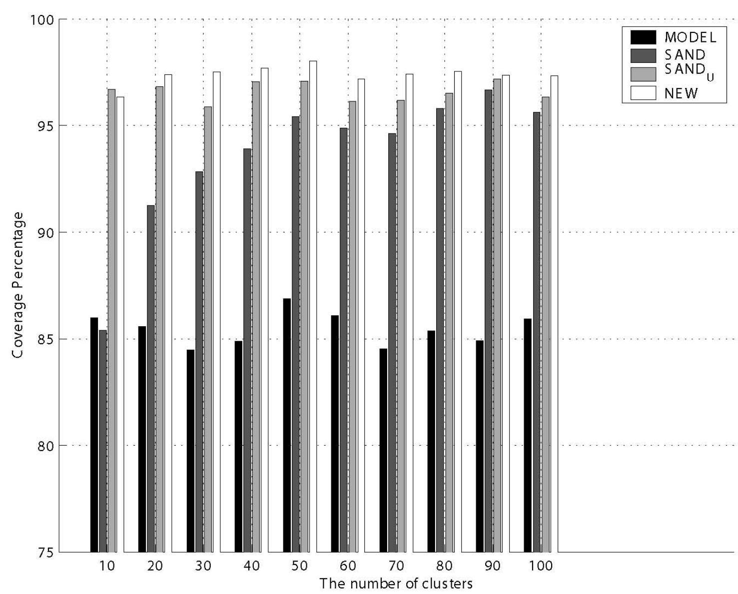

Figure 6

References

Chaganty, N. R. (1997). An alternative approach to the analysis of longitudinal data via generalized estimating equations. *Journal of Statistica Planning and Inference*, *63*, 39–54.

Diggle, P. J., Liang, K. Y., & Zeger, S. T. (1994). *Analysis of longitudinal data*. NY: Oxford Science Publications.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modeling based on gerealized linear models*. (2^{nd} ed.). NY: Springer.

Kauermann, G., & Carroll, R. J. (2001). *Journal of the American Statistical Association*, *96*, 1387-1396.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 12-22.

Mancl, L. A., & DeRouen, T. A. (2001). A covariance estimator for gee with improved small-sample properties. *Biometrics*, *57*, 126-134.

Pan, W. (2001). On the robust variance estimator in generalized estimating equations. *Biometrika*, *3*, 901-906.

Schott, J. R. (1997). *Matrix analysis for statistics*. NY: Wiley.

Zeger, S. L., & Liang, K. Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, *11*, 1825-1839.

Zhang, S. (2003). *Statistical analysis for clustered data*. Unpublished doctoral dissertation. Ontario, Canada: University of Waterloo.

Zhao, L. P., & Prentice, R. L. (1990). Correlated binary regression using a generalized quadratic models. *Biometrika*, *77,* 642-648.