Statistica Sinica **18**(2008), 121-133

EFFICIENT ESTIMATION METHODS FOR INFORMATIVE CLUSTER SIZE DATA

Chin-Tsang Chiang and Kuang-Yao Lee

National Taiwan University and Penn State University

Abstract: Based on clustered data with informative cluster size, two efficient estimation methods are proposed for marginal models. In our procedures, the information of within-cluster correlation and minimum cluster size is fully used; this is not the case with the within-cluster re-sampling (WCR) and cluster-weighted generalized estimating equation (CWGEE) methods. When the correlation model is valid and the minimum cluster size is greater than one, the proposed estimatiors further improve the efficiency of the WCR and CWGEE estimators. As with the WCR estimation procedure, our first estimation method is computationally intensive. To overcome this problem, a second estimation method is developed in which the estimator is asymptotically equivalent to the first one. Asymptotic properties of the estimators are derived. The finite sample properties of the second estimator are investigated through a Monte Carlo simulation; a comparison with the CWGEE estimator is made in the numerical study.

Key words and phrases: Cluster-weighted generalized estimating equation, efficient estimation, informative cluster size, within-cluster re-sampling.

1. Introduction

We consider clustered data of the form $\{((X_{i1}, Y_{i1}), \ldots, (X_{im_i}, Y_{im_i})) : i = 1, \ldots, n\}$, where Y_{ij} and X_{ij} are, respectively, the response and the covariate vector of the jth individual within the ith cluster, and m_i denotes the random cluster size. This type of data occurs in many biomedical and epidemiological studies in which the sampling units (clusters) include a number of individuals or repeated measurements. An appropriate parametric model, $E[Y_{ij}|x_{ij}] = \mu(x_{ij}; \beta)$, where $\mu(\cdot; \beta)$ is a known function and β is the parameter vector, is used to model the relationship between the response Y and the covariates X of each individual.

Under the validity of the model, an independent cluster size – for example in generalizes $E[Y_{ij}|x_{ij}, m_i] = E[Y_{ij}|x_{ij}]$ for all i, j – is usually assumed in estimation procedures, estimating equations (GEE). This assumption is impractical in some applications, evidenced by the periodontal study conducted by Gansky, Weinstraub and Multi-Pied Investigators (1998, 1999). From exploratory data analysis there, it was observed that people with poor dental health tend to possess fewer teeth. Since the cluster size is not predetermined and might be correlated with the measurements, two different analyses used in applications were studied. One of the estimation procedures used all individuals of the randomly selected clusters in the GEE. Another one used a randomly selected individual from each cluster in the GEE. When cluster size is informative, $E[Y_{ij}|x_{ij}, m_i] \neq E[Y_{ij}|x_{ij}]$ for some i, j, it can be verified that the first method results in an inconsistent estimator of β because of over-sampling. For example, in the periodontal study, people with higher dental health will tend to be over-sampled. Based on this biased sample, each tooth response is inappropriately given an equal weight in the traditional GEE. For informative cluster size data, Hoffman, Sen and Weinberg (2001) proposed the within-cluster resampling (WCR) method for the estimation of β. The proposed estimator takes the average of the estimators computed based on each possible sub-sample, in which one individual is independently drawn from each cluster. Generally, the WCR estimation procedure is computationally intensive. To overcome this problem, Williamson, Datta and Satten (2003) suggested the cluster-weighted generalized estimating equation (CWGEE) method. The authors used the averages of all individual score functions within each cluster in the GEE, and showed the estimators from the CWGEE and WCR methods to be asymptotically equivalent.

When cluster size is independent, the estimators from WCR and CWGEE are relatively inefficient compared with those from GEE with appropriate specification for the working matrices. However, if the cluster size is correlated with the measurements, estimators of the traditional GEE method will result in inconsistent estimators. As can be seen, the information of within-cluster correlation is not fully used in the WCR and CWGEE methods, although an estimation of correlation function was considered in the work of Williamson et al. (2003). In many empirical examples, the minimum cluster size $m = \min\{m_1, \cdots, m_n\}$ of n randomly selected clusters is often greater than one. When an appropriate correlation model, $Cov(Y_{ij_1}, Y_{ij_2}|x_{ij_1}, x_{ij_2}) = h(x_{ij_1}, x_{ij_2}; \alpha)$, $i = 1, \ldots, n$; $j_1, j_2 = 1, \ldots, m_i$, is specified, we propose two efficient estimation methods for β. The first (MWCR) is a modification of the WCR method, and uses m randomly selected individuals from each cluster and the corresponding estimated variance matrix of Williamson et al. (2003) in the estimating equations. As with the WCR procedure, MWCR is computationally intensive. To save computational cost, a second estimation method is proposed, in which the asymptotic expression of the MWCR estimator is used as the estimating equation. Thus, the estimator can be shown to be asymptotically equivalent to that computed by the MWCR method.

The contents of this paper are organized as follows. In Section 2, two efficient estimation methods for β, and a consistent estimator for the variance matrix of the estimators, are proposed. The asymptotic properties of the estimators are derived in Section 3. In Section 4, a Monte Carlo simulation is implemented to investigate the finite sample properties of the second estimator. A comparison with the CWGEE estimator is also made in this section. Finally, a brief discussion is provided in Section 5.

2. Estimation Methods

Let $X_i = (X_{i1}, \ldots, X_{im_i})^T$, $Y_i = (Y_{i1}, \ldots, Y_{im_i})^T$, $\mu_i = (\mu_{i1}, \ldots, \mu_{im_i})^T$, and $V_i = (v_{ij_1j_2})$, with $\mu_{ij} = \mu(x_{ij}; \beta)$ and $v_{ij_1j_2} = h(x_{ij_1}, x_{ij_2}; \alpha)$, $i = 1, \ldots, n$; j, j_1, $j_2 = 1, \ldots, m_i$. The minimum cluster size of $\{m_1, \ldots, m_n\}$ is denoted by m. To take account of within-cluster correlation structure, estimated variance-covariance matrices of any m individuals within each cluster are considered in our estimating equations.

In our first procedure (MWCR), m individuals are randomly selected from each cluster. Let $\{(X_{1q_{(m)}}, Y_{1q_{(m)}}), \ldots, (X_{nq_{(m)}}, Y_{nq_{(m)}})\}$ be the qth sub-sample of all $Q_m = \prod_{i=1}^n C_m^{m_i}$ sub-samples of $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, where $X_{iq_{(m)}} = (X_{i1q}, \ldots, X_{imq})^T$ and $Y_{iq_{(m)}} = (Y_{i1q}, \ldots, Y_{imq})^T$, $i = 1, \ldots, n$. Based on the qth sub-sample, the estimator $\widehat{\beta}_{q_{(m)}}$ of β is defined to be the solution of the estimating equation

$$S_{1q}^{(m)}(\beta) = \sum_{i=1}^{n} \Big(\frac{\partial \mu_{iq_{(m)}}}{\partial \beta}\Big) \widehat{V}_{iq_{(m)}}^{-1} (Y_{iq_{(m)}} - \mu_{iq_{(m)}}) = 0, \tag{1}$$

where $\mu_{iq_{(m)}} = (\mu_{i1q}, \cdots, \mu_{imq})^T$, and $\widehat{V}_{iq_{(m)}}$ is an estimator of $V_{iq_{(m)}} = (h(x_{ij_1q}, x_{ij_2q}; \alpha))$ with the estimator $\widehat{\alpha}$ of Williamson et al. (2003) being substituted for α, $i = 1, \ldots, n$. Note that $\widehat{\alpha}$ is obtained from an unbiased estimating equation in which β is substituted by the CWGEE estimator. By taking the average of the Q_m estimators $\{\widehat{\beta}_{1_{(m)}}, \ldots, \widehat{\beta}_{Q_{m(m)}}\}$, the proposed MWCR estimator $\widehat{\beta}_{1m} = Q_m^{-1} \sum_{q=1}^{Q_m} \widehat{\beta}_{q_{(m)}}$ is obtained. Since Q_m is extremely large in practice, a reasonable number of re-samplings is implemented. Parallelling the proof of Hoffman, Sen and Weinberg (2001) and Theorem 3.1 in the next section, the asymptotic equivalence between the estimators can be derived. When independent working matrices are used in the estimation procedure of MWCR, the proposed estimator $\widehat{\beta}_{1m}$ is shown to have the same asymptotic distribution as that of the WCR estimator $\widehat{\beta}_{wcr}$. Similarly if k individuals are drawn in the MWCR approach with independent working matrices, $2 \le k \le m$, the same asymptotic risks can be derived, as is shown in the next section.

Use of the MWCR procedure comes at great computational cost. The second estimation method is to be preferred. It is mainly motivated by writing $\widehat{\beta}_{1m}$ as

$$\widehat{\beta}_{1m}=\beta+\frac{1}{n}H_m^{-1}(\beta)\sum_{i=1}^{n}\frac{1}{C_m^{m_i}}\sum_{\{j_{(m)}\in\Omega_i^{(m)}\}}\Big(\frac{\partial\mu_{ij_{(m)}}}{\partial\beta}\Big)V_{ij_{(m)}}^{-1}(Y_{ij_{(m)}}-\mu_{ij_{(m)}})+o_p\Big(\frac{1}{\sqrt{n}}\Big), \tag{2}$$

where $\Omega_i^{(m)}=\{(j_1,\ldots,j_m):1\leq j_1<\cdots<j_m\leq m_i\}$, $H_m(\beta)=E[(\partial\mu_{ij_{(m)}}/\partial\beta)V_{ij_{(m)}}^{-1}(\partial\mu_{ij_{(m)}}/\partial\beta)^T]$. Then $\widehat{\beta}_{2m}$ is obtained from the following estimating equation:

$$S_2^{(m)}(\beta)=\sum_{i=1}^{n}\frac{1}{C_m^{m_i}}\sum_{\{j_{(m)}\in\Omega_i^{(m)}\}}\Big(\frac{\partial\mu_{ij_{(m)}}}{\partial\beta}\Big)\widehat{V}_{ij_{(m)}}^{-1}(Y_{ij_{(m)}}-\mu_{ij_{(m)}})=0. \tag{3}$$

For the estimation of the variance-covariance matrix Σ_m of $\widehat{\beta}_{2m}$, an easily computed sandwich estimator $\widehat{\Sigma}_m=\widehat{H}_m^{-1}\widehat{V}_m\widehat{H}_m^{-1}$ is suggested, where

$$\widehat{H}_m=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{C_m^{m_i}}\sum_{\{j_m\in\Omega_i^{(m)}\}}\Big(\frac{\partial\mu_{ij_{(m)}}}{\partial\beta}\Big)\widehat{V}_{ij_{(m)}}^{-1}\Big(\frac{\partial\mu_{ij_{(m)}}}{\partial\beta}\Big)^T\Big|_{\beta=\widehat{\beta}_{2m}},$$

$$\widehat{V}_m=\frac{1}{C_m^{m_i}}\sum_{\{j_{(m)}\in\Omega_i^{(m)}\}}\Big(\frac{\partial\mu_{ij_{(m)}}}{\partial\beta}\Big)\widehat{V}_{ij_{(m)}}^{-1}(Y_{ij_{(m)}}-\mu_{ij_{(m)}})(Y_{ij_{(m)}}-\mu_{ij_{(m)}})^T \times\widehat{V}_{ij_{(m)}}^{-1}\Big(\frac{\partial\mu_{ij_{(m)}}}{\partial\beta}\Big)^T\Big|_{\beta=\widehat{\beta}_{2m}}.$$

By using the asymptotic normality of $\widehat{\beta}_{2m}$ and a consistent estimator $\widehat{\Sigma}_m$ of Σ_m, a $100(1-\alpha)\%$ confidence interval for β_j, $0<\alpha<1$, can be constructed as

$$\widehat{\beta}_{2m_j}\pm z_{(1-\frac{\alpha}{2})}\widehat{\sigma}_{jm},\quad j=1,\ldots,p, \tag{4}$$

where $\widehat{\sigma}_{jm}$ is the jth diagonal element of $\widehat{\Sigma}_m$ and $z_{(1-(\alpha/2))}$ is the $(1-\alpha/2)$th percentile of a standard normal distribution.

Note that the WCR estimator $\widehat{\beta}_{wcr}$ and the CWGEE estimator $\widehat{\beta}_{cw}$ might be fully efficient in some cases even when informative cluster size data are considered. Under the validity of a generalized linear model with a canonical link function or a variance stabilizing function, Mancl and Leroux (1996) detected that, in the setting of independent and equal cluster size, the efficiency of an estimator from the estimating equation with an independent working matrix (IEE) relies on the distribution of covariates. When $x_{ij}=z_i$ or $\bar{x}_1=\cdots=\bar{x}_n$ with $\bar{x}_i=\sum_{j=1}^{m}x_{ij}/m$, $i=1,\cdots,n$; $j=1,\cdots,m$, the authors found that the estimator from the IEE has the same asymptotic variance as that from the GEE with the optimal working matrix. For example, under the validity of a linear model $\mu_{ij} = x_{ij}^T\beta$ with $x_{ij} = z_i$ and $V_i = \rho I_m + (1-\rho)1_m 1_m^T$, the score function $S_{gee}(\beta)$ for β is

$$\begin{aligned} S_{gee}(\beta) &= \sum_{i=1}^{n}(z_i 1_m^T)V_i^{-1}(Y_i - \mu_i) \\ &= \frac{1}{(1-\rho)(1+(m-1)\rho)}\sum_{i=1}^{n}(z_i 1_m^T)\Big((1+(m-1)\rho)I_m - \rho 1_m 1_m^T\Big)(Y_i - \mu_i) \\ &= \frac{1}{(1+(m-1)\rho)}\sum_{i=1}^{n}(z_i 1_m^T)(Y_i - \mu_i). \end{aligned} \tag{5}$$

This implies that $S_{gee}(\beta) = (1+(m-1)\rho)^{-1}\sum_{i=1}^{n}(z_i 1_m^T)(Y_i - \mu_i)$ in (5) is the score function of the IEE for β. Thus, the estimator computed from the GEE is the same as that from the IEE.

3. Asymptotic Properties

In this section, the asymptotic properties of the proposed estimators are derived. The conditions for the main theorems are as follows.

(A1) $H_1(\beta_0) = E[(\partial\mu_{ij}/\partial\beta)v_{ijj}^{-1}(\partial\mu_{ij}/\partial\beta)^T]|_{\beta=\beta_0}$ exists for β_0 in $B(\delta) = \{\beta_0 \in R^p : \|\beta_0 - \beta\| < \delta\}$ for some positive value δ, and $H_1(\beta)$ is a positive definite matrix.

(A2) $(1/n)E[v_{ijj} tr(G_{ij}(\beta)G_{ij}^T(\beta))] \to 0$ as $n\to\infty$, where $G_{ij}(\beta) = (\partial/\partial\beta)((\partial\mu_{ij}/\partial\beta) v_{ijj}^{-1})$.

(A3) $E[\sup_{\beta_0\in B(\delta)} \|(\partial\mu_{ij}/\partial\beta)v_{ijj}^{-1}(\partial\mu_{ij}/\partial\beta)^T|_{\beta=\beta_0} - (\partial\mu_{ij}/\partial\beta)v_{ijj}^{-1}(\partial\mu_{ij}/\partial\beta)^T\|_\infty] \to 0$, $E[\sup_{\beta_0\in B(\delta)} \|G_{ij}(\beta_0) - G_{ij}(\beta)\|_\infty] \to 0$, and $E[\sup_{\beta_0\in B(\delta)} |Y_{ij}|\|G_{ij}(\beta_0) - G_{ij}(\beta)\|_\infty] \to 0$ as $\delta \to 0$, where $\|A\|_\infty$ is the maximum eigenvalue of matrix A.

When independent working matrices are used in the estimation procedure of MWCR, the proposed estimator $\widehat{\beta}_{1m}$ will be shown to have the same asymptotic distribution as those of $\widehat{\beta}_{wcr}$ and $\widehat{\beta}_{cw}$ in the following theorem. Similarly, if k individuals are drawn in the MWCR approach with independent working matrices, $2 \le k \le m$, the same asymptotic equivalence can be derived. Let $\widehat{\beta}_{0k} = Q_k^{-1}\sum_{q=1}^{Q_k}\widehat{\beta}_{0q(k)}$ denote the corresponding estimator, where $Q_k = \prod_{i=1}^{n} C_k^{m_i}$ and the qth estimator $\widehat{\beta}_{0q(k)}$ is the root of the estimating equation $S_{10q}^{(k)}(\beta) = \sum_{i=1}^{n}(\partial\mu_{iq_{(k)}}/\partial\beta)(Y_{iq_{(k)}} - \mu_{iq_{(k)}}) = 0$.

Theorem 3.1. *Suppose* (A1)$-$(A3) *are satisfied. Then*

$$\widehat{\beta}_{0k} = \bar{\beta}_{wcr} + o_p(\frac{1}{\sqrt{n}}), \quad 2 \le k \le m. \tag{6}$$

Proof. From (A1)$-$(A3), $\widehat{\beta}_{0k}$ can be written as

$$\widehat{\beta}_{0k} = \beta + \frac{1}{n} H_1^{-1}(\beta) \sum_{q=1}^{Q_k} \frac{1}{kQ_k} S_{10q}^{(k)}(\beta) + o_p\Big(\frac{1}{\sqrt{n}}\Big). \tag{7}$$

Using the equality

$$\begin{aligned}
\frac{1}{kQ_k} \sum_{q=1}^{Q_k} S_{10q}^{(k)}(\beta) &= \sum_{i=1}^{n} \frac{1}{kQ_k} \sum_{q=1}^{Q_k} \Big(\frac{\partial \mu_{iq_{(k)}}}{\partial \beta}\Big)(Y_{iq_{(k)}} - \mu_{iq_{(k)}}) \\
&= \sum_{i=1}^{n} \frac{1}{Q_k} \sum_{q=1}^{Q_k} \frac{1}{k} \sum_{j=1}^{k} \Big(\frac{\partial \mu_{ijq_{(k)}}}{\partial \beta}\Big)(Y_{ijq_{(k)}} - \mu_{ijq_{(k)}}) \\
&= \sum_{i=1}^{n} \frac{1}{kQ_k} \Big(\prod_{l \ne i}^{n} C_k^{m_l}\Big) \frac{(C_k^{m_i} \times k)}{m_i} \sum_{j=1}^{m_i} \Big(\frac{\partial \mu_{ij}}{\partial \beta}\Big)(Y_{ij} - \mu_{ij}) \\
&= \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \Big(\frac{\partial \mu_{ij}}{\partial \beta}\Big)(Y_{ij} - \mu_{ij}) = S_{cw}(\beta),
\end{aligned} \tag{8}$$

where $S_{cw}(\beta)$ is the estimating equation of the CWGEE estimation procedure. Thus, $\widehat{\beta}_{0k}$ and $\widehat{\beta}_{cw}$ are asymptotically equivalent. Finally, using the asymptotic equivalence of $\widehat{\beta}_{wcr}$ and $\widehat{\beta}_{cw}$, derived by Williamson et al. (2003), (6) is obtained.

Since the estimators $\widehat{\beta}_{1m}$ and $\widehat{\beta}_{0m}$ can be expressed asymptotically as linear combinations of the responses, the Gauss-Markov Theorem for linear models can be used to show that the estimator $b^T\widehat{\beta}_{1m}$ is asymptotically more efficient than the estimator $b^T\widehat{\beta}_{0m}$ for all b. In addition, when k individuals are randomly selected from each cluster in the MWCR procedure, $2 \le k \le m-1$, one can show that $b^T\widehat{\beta}_{1m}$ is asymptotically more efficient than $b^T\widehat{\beta}_{1k}$. The asymptotic normalities of $\widehat{\beta}_{1m}$ and $\widehat{\beta}_{1k}$'s are derived first. Some natural extensions of assumptions (A1)$-$(A3) are also made below. Let (A1$'$)$-$(A3$'$) correspond to (A1)$-$(A3) with $H_k(\beta) = E[(\partial\mu_{ij_{(k)}}/\partial\beta)V_{ij_{(k)}}^{-1}(\partial\mu_{ij_{(k)}}/\partial\beta)^T]$ and $G_{ij_{(k)}} = (\partial/\partial\beta)((\partial\mu_{ij_{(k)}}/\partial\beta)v_{ij_{(k)}}^{-1})$ substituting separately for $H_1(\beta)$ and $G_{ij}(\beta)$, where $V_{ij_{(k)}} = \mathrm{Var}\,(Y_{ij_{(k)}}|\, x_{ij_{(k)}})$ and $\mu_{ij_{(k)}} = E[Y_{ij_{(k)}}|x_{ij_{(k)}}]$.

Theorem 3.2. *Suppose* (A1$'$) $-$ (A3$'$) *are satisfied. Then*

$$\sqrt{n}(\widehat{\beta}_{1k} - \beta) \xrightarrow{d} N_p(0, H_k^{-1}(\beta)), \quad 2 \le k \le m. \tag{9}$$

Proof. By using a Taylor expansion and the consistency of $\widehat{V}_{iq_{(k)}}$, one has

$$S_{1q}^{(k)}(\widehat{\beta}_{q(k)}) = S_{1q}^{(k)}(\beta) + \frac{\partial S_{1q}^{(k)}(\beta)}{\partial \beta}\Big|_{\beta=\beta^*}(\widehat{\beta}_{q(k)} - \beta)(1 + o_p(1)), \tag{10}$$

where β^* lies along the line segment joining $\widehat{\beta}_{q(k)}$ and β. From (A1′)−(A3′), and by the Law of Large Numbers, it can be seen that

$$\frac{1}{n}\frac{\partial S_{1q}^{(k)}(\beta)}{\partial \beta}\Big|_{\beta=\beta^*} \xrightarrow{p} H_k(\beta) \quad \text{as} \quad n \to \infty. \tag{11}$$

Thus, $\widehat{\beta}_{q(k)}$ can be expressed as

$$\widehat{\beta}_{q(k)} = \beta + H_k^{-1}(\beta) S_{1q}^{(k)}(\beta) + o_p(\frac{1}{\sqrt{n}}). \tag{12}$$

From (12) and $\widehat{\beta}_{1k} = Q_k^{-1}\sum_{q=1}^{Q_k}\widehat{\beta}_{q(k)}$, we get

$$\begin{aligned}\widehat{\beta}_{1k} &= \beta + \frac{1}{n}H_k^{-1}(\beta)\frac{1}{Q_k}\sum_{q=1}^{Q_k}\sum_{i=1}^{n}\Big(\frac{\partial \mu_{iq_{(k)}}}{\partial \beta}\Big)V_{iq_{(k)}}^{-1}(Y_{iq_{(k)}} - \mu_{iq_{(k)}}) + o_p(\frac{1}{\sqrt{n}}) \\ &= \beta + \frac{1}{n}H_k^{-1}(\beta)\sum_{i=1}^{n} U_i^{(k)}(\beta) + o_p(\frac{1}{\sqrt{n}}),\end{aligned} \tag{13}$$

where $U_i^{(k)}(\beta) = (C_k^{m_i})^{-1}\sum_{\{j_{(k)}\in\Omega_i^{(k)}\}}(\partial\mu_{ij_{(k)}}/\partial\beta)V_{ij_{(k)}}^{-1}(Y_{ij_{(k)}} - \mu_{ij_{(k)}})$ with $\Omega_i^{(k)} = \{(j_1,\cdots,j_k) : 1 \le j_1 < \cdots < j_k \le m_i\}$. Since the $U_i^{(k)}(\beta)'s$ are independent and identically distributed with zero mean and variance $H_k(\beta)$, by the Central Limit Theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i^{(k)}(\beta) \xrightarrow{d} N_p(0, H_k(\beta)). \tag{14}$$

Finally, (9) is obtained from (13) and (14).

Theorem 3.3. *Suppose* (A1′)−(A3′) *are satisfied. Then* (a) Var $(\widehat{\beta}_{0k})$−Var $(\widehat{\beta}_{1k})$, and (b) Var $(\widehat{\beta}_{1k})$ − Var $(\widehat{\beta}_{1m})$ are at least positive semi-definite for $2 \le k \le m$.

Proof. Using the Gauss-Markov Theorem, it is straightforward to derive that Var $(\widehat{\beta}_{0k})$ − Var $(\widehat{\beta}_{1k})$ is at least positive semi-definite. From (13), $\widehat{\beta}_{1k}$ is shown to be asymptotically equivalent to $\widetilde{\beta}_{1m}$, which is computed from the estimation procedure of MWCR with the weighting matrix

$$W_{ij_{(m)}} = \begin{pmatrix} V_{ij_{(k)}}^{-1} & 0 \\ 0 & 0 \end{pmatrix}. \tag{15}$$

Since $V_{ij_{(m)}}^{-1}$ is the optimal weighting matrix, $\widehat{\beta}_{1m}$ is asymptotically more efficient than $\widehat{\beta}_{1k}$ hence (b) is obtained.

4. Monte Carlo Simulations

In this section, a Monte Carlo simulation examines the finite sample properties of the proposed estimator $\widehat{\beta}_{2m}$ and the estimated variance matrix $\widehat{\Sigma}_m$. A comparison with the CWGEE estimator $\widehat{\beta}_{cw}$ is also made. The informative clustered data are generated from n independent clusters $\{(X_{ij}, Y_{ij}) : i = 1, \cdots, n;\ j = 1, \cdots, m_i\}$ with a sample size at 25, 50, 100, and 200, respectively. Here, the cluster sizes $m_1, \ldots, m_n$ are designed to be a random sample from a discrete uniform distribution with the support $\{m, \ldots, m+10\}$. The minimum cluster size is set separately to be $2, 5$, and 10. The random numbers $U_1, \cdots, U_n$ are first generated from a $Uniform(4, 12)$. Given each random number U_i, the covariates $X_{ij} = (X_{ij1}, X_{ij2})^T$'s are designed with $X_{ij1} = 1$ and the X_{ij2}'s being independently drawn from a $Uniform(U_i - 4, U_i + 4)$. As for the response vectors $Y_1, \ldots, Y_n$, conditioning on $(m_i, x_{i1}, \ldots, x_{im_i})$, Y_i is generated from a multivariate normal distribution with mean μ_i and variance matrix V_i, where $\mu_{ij} = 1.5x_{ij} + 0.5m_i$ and $v_{ij_1j_2} = (1.25\rho - 0.25) \cdot \text{Var}\ (m_i)1_{(j_1 \neq j_2)} + (1.25 - 1.25\rho) \cdot \text{Var}\ (m_i)1_{(j_1 = j_2)}$, with $\rho = Cor(Y_{ij}, Y_{ik}|x_{ij}, x_{ik})$ is 0.2, 0.5, or 0.8 for all i, j. Obviously then, cluster size is informative.

The informative cluster size data are repeatedly generated 1,000 times. For each simulated data set, the estimates $\widehat{\beta}_{cw}$, $\widehat{\beta}_{2m}$, and $\widehat{\Sigma}_m$ are computed. Moreover, the equi-correlation working matrix is chosen in the estimation of $\widehat{\beta}_{2m}$ and $\widehat{\Sigma}_m$. It is found in Tables 4.1 (a)$-$(b) that the averages of 1,000 estimates $\widehat{\beta}_{cw}$ and $\widehat{\beta}_{2m}$ are very close to the true values $(\beta_0, \beta_1) = (0.5E(m_i), 1.5)$ under different sample sizes, correlation values, and minimum cluster sizes. As for the relative efficiencies of $\widehat{\beta}_{cwi}$ versus $\widehat{\beta}_{2mi}$, say, $RE_i = \sigma^2(\widehat{\beta}_{2mi})/\sigma^2(\widetilde{\beta}_{cwi})$, $i = 0, 1$, the variances of both estimators are computed based on 1,000 estimates. From Figures 4.1 through 4.2, we can see that RE_i's decrease as the minimum cluster size and the correlation coefficient increase. However, the influence of the sample size is not apparent when the minimum cluster size and the correlation value are fixed. Moreover, one sees from Tables 4.2 (a)$-$(b) that the averages of 1,000 variance estimates $(\widehat{\sigma}_{1m}^2, \widehat{\sigma}_{2m}^2)$ approach the variances of 1,000 estimates $(\widehat{\beta}_{2m0}, \widehat{\beta}_{2m1})$ as the sample size increases. The minimum cluster size and the correlation value are not dominant factors in the accuracy of the variance estimates. Finally, the empirical coverage probabilities of β_0 and β_1 based on 1,000 95% confidence intervals in (4) are provided in Table 4.3. It is seen in this table that the empirical coverage probabilities are close to the 0.95 nominal level when the sample size is large enough. As for the influence of correlation and minimum cluster size, no apparent difference is found in the coverage probabilities.

Table 4.1(a). Averages on 1,000 estimates of $\widetilde{\beta}_{cw_0}$ and $\widehat{\beta}_{2m_0}$.

		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
	true	$Ave.(\widetilde{\beta}_{cw_0})$	$Ave.(\widehat{\beta}_{2m_0})$	$Ave.(\widetilde{\beta}_{cw_0})$	$Ave.(\widehat{\beta}_{2m_0})$	$Ave.(\widetilde{\beta}_{cw_0})$	$Ave.(\widehat{\beta}_{2m_0})$
$n = 25$							
$m = 2$	3.25	3.34	3.34	3.28	3.28	3.27	3.23
5	4.75	4.75	4.74	4.76	4.79	4.78	4.78
10	7.25	7.28	7.27	7.25	7.25	7.26	7.26
$n = 50$							
$m = 2$	3.25	3.25	3.25	3.24	3.24	3.27	3.27
5	4.75	4.72	4.72	4.77	4.78	4.73	4.74
10	7.25	7.27	7.26	7.25	7.26	7.29	7.26
$n = 100$							
$m = 2$	3.25	3.25	3.25	3.24	3.25	3.25	3.26
5	4.75	4.73	4.73	4.74	4.75	4.74	4.74
10	7.25	7.24	7.25	7.27	7.25	7.27	7.25
$n = 200$							
$m = 2$	3.25	3.24	3.24	3.26	3.25	3.23	3.24
5	4.75	4.75	4.75	4.74	4.74	4.74	4.75
10	7.25	7.24	7.24	7.25	7.25	7.26	7.26

$\widehat{\beta}_{2m_0}$ and $\widetilde{\beta}_{cw_0}$ = the estimates of $\widehat{\beta}_{2m}$ and $\widetilde{\beta}_{cw}$ for β_0.
Ave. = the average of 1,000 simulated data sets.

Table 4.1(b). Averages on 1,000 estimates of $\widetilde{\beta}_{cw_1}$ and $\widehat{\beta}_{2m_1}$.

		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
	true	$Ave.(\widetilde{\beta}_{cw_1})$	$Ave.(\widehat{\beta}_{2m_1})$	$Ave.(\widetilde{\beta}_{cw_1})$	$Ave.(\widehat{\beta}_{2m_1})$	$Ave.(\widetilde{\beta}_{cw_1})$	$Ave.(\widehat{\beta}_{2m_1})$
$n = 25$							
$m = 2$	1.5	1.48	1.49	1.49	1.49	1.49	1.50
5	1.5	1.50	1.50	1.49	1.49	1.49	1.49
10	1.5	1.49	1.49	1.50	1.50	1.49	1.49
$n = 50$							
$m = 2$	1.5	1.49	1.49	1.49	1.50	1.50	1.50
5	1.5	1.50	1.50	1.49	1.49	1.50	1.50
10	1.5	1.49	1.49	1.49	1.49	1.49	1.49
$n = 100$							
$m = 2$	1.5	1.49	1.49	1.50	1.50	1.49	1.49
5	1.5	1.50	1.50	1.50	1.50	1.50	1.50
10	1.5	1.49	1.49	1.49	1.50	1.49	1.49
$n = 200$							
$m = 2$	1.5	1.50	1.50	1.49	1.49	1.50	1.50
5	1.5	1.49	1.49	1.50	1.50	1.49	1.50
10	1.5	1.50	1.50	1.49	1.49	1.50	1.50

$\widehat{\beta}_{2m_1}$ and $\widetilde{\beta}_{cw_1}$ = the estimates of $\widehat{\beta}_{2m}$ and $\widetilde{\beta}_{cw}$ for β_1.
Ave. = the average of 1,000 simulated data sets.

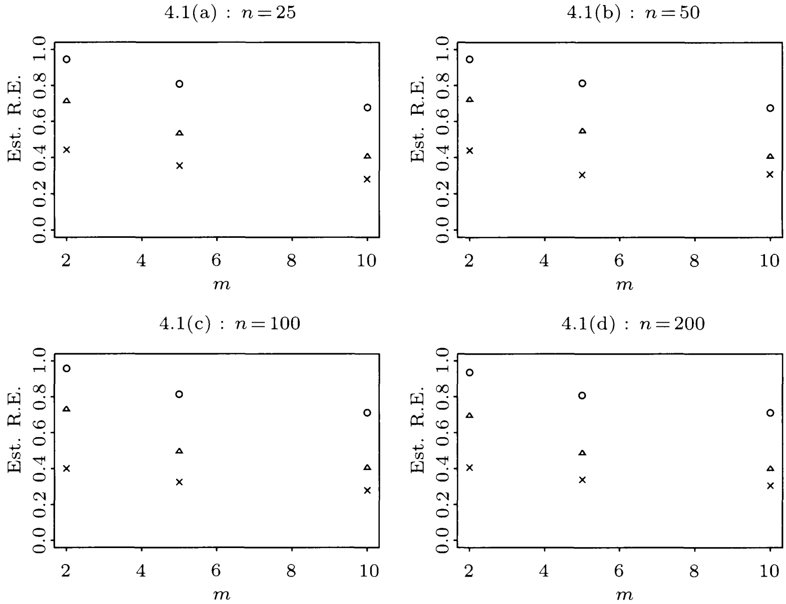

Figure 1. The O, $\triangle$, and $\times$ represent the estimated relative efficiency of $\widetilde{\beta}_{cw_0}$ versus $\widehat{\beta}_{2m_0}$ for ρ =0.2, 0.5 and 0.8, respectively.

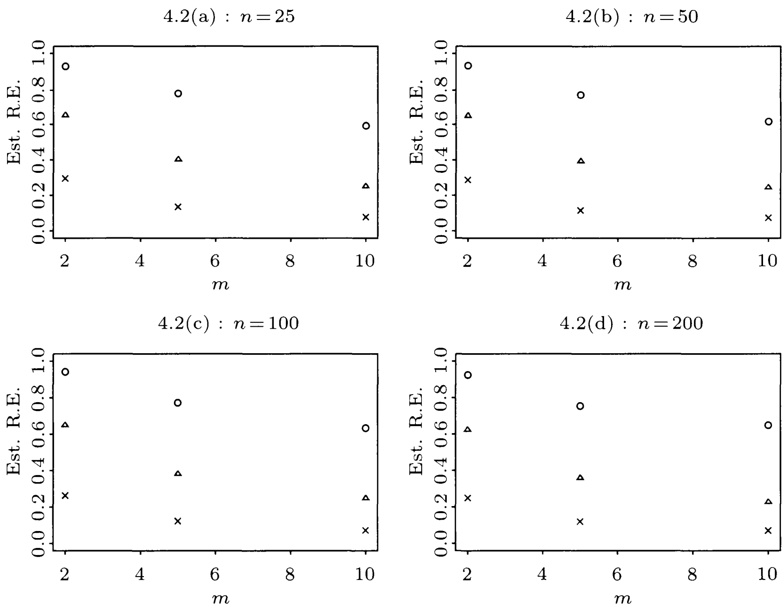

Figure 2. The O, $\triangle$, and $\times$ represent the estimated relative efficiency of $\widetilde{\beta}_{cw_1}$ versus $\widehat{\beta}_{2m_1}$ for ρ =0.2, 0.5 and 0.8, respectively.

Table 4.2(a). he variances of 1,000 estimates $\widehat{\beta}_{2m_0}$ and the averages of 1,000 variance estimates $\widehat{\sigma}^2_{1m}$.

	$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
	Var $(\widehat{\beta}_{2m_0})$	$Ave.(\widehat{\sigma}^2_{1m})$	Var $(\widehat{\beta}_{2m_0})$	$Ave.(\widehat{\sigma}^2_{1m})$	Var $(\widehat{\beta}_{2m_0})$	$Ave.(\widehat{\sigma}^2_{1m})$
$n = 25$						
$m = 2$	0.8154	0.7312	0.9044	0.8021	0.7448	0.6742
5	0.4858	0.4684	0.5625	0.5131	0.5170	0.4620
10	0.3540	0.3349	0.3758	0.3880	0.4045	0.3978
$n = 50$						
$m = 2$	0.4083	0.3856	0.4489	0.4221	0.3516	0.3452
5	0.2694	0.2489	0.2777	0.2633	0.2421	0.2351
10	0.1731	0.1732	0.1936	0.1991	0.2209	0.2027
$n = 100$						
$m = 2$	0.1962	0.1961	0.2237	0.2132	0.1697	0.1679
5	0.1310	0.1257	0.1289	0.1333	0.1183	0.1171
10	0.083	0.0869	0.0930	0.0990	0.1043	0.1020
$n = 200$						
$m = 2$	0.1018	0.0989	0.1096	0.1095	0.0858	0.0838
5	0.0640	0.0635	0.0645	0.0671	0.0640	0.0589
10	0.0433	0.0441	0.0497	0.0501	0.0548	0.0513

Table 4.2(b). The variances of 1,000 estimates $\widehat{\beta}_{2m_1}$ and the averages of 1,000 variance estimates $\widehat{\sigma}^2_{2m}$.

	$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
	Var $(\widehat{\beta}_{2m_1})$	$Ave.(\widehat{\sigma}^2_{2m})$	Var $(\widehat{\beta}_{2m_1})$	$Ave.(\widehat{\sigma}^2_{2m})$	Var $(\widehat{\beta}_{2m_1})$	$Ave.(\widehat{\sigma}^2_{2m})$
$n = 25$						
$m = 2$	0.0098	0.0092	0.0102	0.0090	0.0064	0.0053
5	0.0058	0.0055	0.0052	0.0046	0.0024	0.0021
10	0.0038	0.0036	0.0029	0.0027	0.0013	0.0011
$n = 50$						
$m = 2$	0.0052	0.0048	0.0049	0.0047	0.0027	0.0026
5	0.0031	0.0029	0.0026	0.0023	0.0010	0.0010
10	0.0020	0.0018	0.0014	0.0013	0.0006	0.0005
$n = 100$						
$m = 2$	0.0025	0.0024	0.0025	0.0024	0.0013	0.0012
5	0.0016	0.0014	0.0012	0.0011	0.0005	0.0005
10	0.0010	0.0009	0.0006	0.0006	0.0003	0.0002
$n = 200$						
$m = 2$	0.0012	0.0012	0.0012	0.0012	0.0006	0.0006
5	0.0007	0.0007	0.0006	0.0005	0.0003	0.0002
10	0.0005	0.0004	0.0003	0.0003	0.0001	0.0001

Table 4.3. he empirical coverage probabilities of β_0 and β_1 at 0.95 nominal level.

	β_0				β_1		
	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$		$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 25$							
$m = 2$	0.925	0.929	0.927	$m = 2$	0.938	0.922	0.904
5	0.932	0.92	0.924	5	0.937	0.911	0.913
10	0.932	0.934	0.936	10	0.933	0.924	0.92
$n = 50$							
$m = 2$	0.943	0.939	0.936	$m = 2$	0.937	0.933	0.948
5	0.93	0.933	0.942	5	0.936	0.925	0.939
10	0.945	0.959	0.941	10	0.931	0.947	0.943
$n = 100$							
$m = 2$	0.94	0.94	0.945	$m = 2$	0.937	0.938	0.939
5	0.942	0.953	0.943	5	0.929	0.946	0.941
10	0.952	0.950	0.942	10	0.943	0.956	0.935
$n = 200$							
$m = 2$	0.949	0.941	0.945	$m = 2$	0.953	0.951	0.934
5	0.946	0.954	0.938	5	0.948	0.946	0.943
10	0.946	0.952	0.934	10	0.951	0.958	0.944

5. Discussion

In this study we show that, with informative cluster size data, our estimators $\widehat{\beta}_{1m}$ and $\widehat{\beta}_{2m}$ are asymptotically more efficient than $\bar{\beta}_{wcr}$ and $\widetilde{\beta}_{cw}$ when the covariance matrix is available and the minimum cluster size is greater than one. However, $\widehat{\beta}_{1m}$ and $\widehat{\beta}_{2m}$ are not fully efficient under independent cluster size data. Since our estimators rely on appropriate specification for the covariance matrices, there are limitations and difficulties in some clustered data with complicated correlation structure. Generally, our proposed methods will be very useful for a homogeneous correlation function within each cluster.

Longitudinal data, frequently encountered in biomedical and epidemiological studies, is one type of clustered data. Here, the variables of each subject (cluster) are repeatedly measured at different time points (individuals), and the number of repeated measurements is treated as the cluster size. However, the measurements might depend not only on the number of repeated measurements but also on the counting process of occurring times. The extension of our proposed estimation methods to this data setting is an important issue.

Acknowledgements

The first author's research was partially supported by the National Science Council grant 94-2118-M-002-009 (Taiwan). We would like to thank the referee for valuable comments.

References

Gansky, S. A., Weinstraub, J. A. and the Multi-Pied Investigators (1998). Parental periodontal predictors of oral health in adult children of a community cohort. *J. Dental Research.* **77** (Special Issue B), 707.

Gansky, S. A., Weinstraub, J. A. and the Multi-Pied Investigators (1999). Family aggregation of periodontal status in a two-generation cohort. *J. Dental Research.* **78** (Special Issue B), 123.

Hoffman, E. B., Sen, P. K. and Weinberg, C. R. (2001). Within-cluster resampling. *Biometrika.* **81**, 1121-1134.

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

Williamson, J. M., Datta, S. and Satten, G. A. (2003). Marginal analyses of clustered data when cluster size is informative. *Biometrics.* **59**, 36-42.

Department of Mathematics, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan.

E-mail: chiangct@ntu.edu.tw

Department of Statistics, Penn State University, 326 Thomas Building, University Park, PA 16802, USA.

E-mail: kxl280@psu.edu

(Received June 2005; accepted May 2006)