Japanese Journal of Biometrics Vol. 32, No. 1, 1–12 (2011)

Original Article

Modifications of QIC and CIC for Selecting a Working Correlation Structure in the Generalized Estimating Equation Method

Masahiko Gosho[*1], Chikuma Hamada[*1] and Isao Yoshimura[*1]

[*1]Graduate School of Engineering, Tokyo University of Science

e-mail:m-gosyo@kowa.co.jp

The generalized estimating equation (GEE) method is a popular method for analyzing longitudinal data. An inappropriate specification of the working correlation structure reduces the efficiency of the GEE estimation. Pan (2001a) and Hin and Wang (2009) proposed a quasi-likelihood under the independence model criterion (QIC) and a correlation information criterion (CIC) for selecting a proper working correlation structure, respectively. In this study, we proposed modifications to the QIC and CIC using the variance estimators of the GEE with improved small-sample properties. In a simulation study, the performance of the modified QIC and CIC was better than that of the original QIC and CIC. The modified methods were illustrated using the data for an air pollution study.

Key words: CIC; Criterion; GEE; QIC; Working correlation structure.

1. Introduction

Correlated data are very commonly adopted in biomedical studies. A typical example of such a study is a longitudinal study in which each subject is followed over a period of time, and an outcome variable Y_{it} and relevant covariate x_{it} for subjects $i = 1, \dots, K$ at times $t = 1, \dots, n_i$ are repeatedly observed.

Liang and Zeger (1986) proposed the generalized estimating equation (GEE) method for analyzing longitudinal data. We need to specify the working correlation matrix as an assumption of the joint probability distribution when we apply the GEE method to analyze longitudinal data. An independent structure, an exchangeable structure, and a first-order autoregressive (AR(1)) structure are commonly used as the working correlation structure.

In addition, it is important to select a proper working correlation matrix when applying the GEE method to actual data, since an improper selection sometimes causes inefficient parameter estimation even for complete data, as pointed out by Fitzmaurice (1995), Mancl and Leroux (1996), and Sutradhar and Das (2000). These findings imply the importance of selecting an

Received October 2010. Revised November 2010. Accepted November 2010.

appropriate correlation structure; as such, some researchers have proposed new criteria for selecting a working correlation structure in the GEE method. For example, Pan (2001a) proposed a modification of Akaike's information criterion (AIC), called the "quasi-likelihood under the independence model criterion (QIC)." In addition, Hin and Wang (2009) proposed a correlation information criterion (CIC) that modifies the QIC and substantially improves its performance.

In this study, we proposed modifications to the QIC and CIC for selecting a proper working correlation structure in the GEE method. In addition, we investigated the performance of the modified QIC and CIC through a simulation study.

This paper is organized as follows. In Section 2, we summarize the GEE method. In Section 3, we provide the definitions of QIC and CIC and propose the modified criteria of QIC and CIC. In Section 4, we present the results of our simulation study. In Section 5, we apply the modified criteria to air pollution data. Finally, Section 6 provides a brief discussion.

2. GEE method

Assume that an $n_i \times p$ matrix of covariate values $X_i = (x_{i1}, \ldots, x_{in_i})^{\mathrm{T}}$ is adjoined to the outcome vector $Y_i = (Y_{i1}, \ldots, Y_{in_i})^{\mathrm{T}}$ on subjects $i = 1, \ldots, K$ and linked with parameters $\theta_{i1}, \theta_{i2}, \ldots,$ θ_{in_i} in the marginal density of Y_{it} in the form of Formula (1) as $\theta_{it} = h^*(x_{it}^{\mathrm{T}}\beta)$, where the superscript T denotes the transpose and h^* is a specified link function:

$$f(y_{it}) = \exp\left[\{y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})\}\phi\right]. \tag{1}$$

As is well known, the expected value $\mu_{it} = E(Y_{it})$ and variance $\sigma_{it}^2 = var(Y_{it})$ are given by $\mu_{it} = a'(\theta_{it})$ and $\sigma_{it}^2 = a''(\theta_{it})/\phi = v(\mu_{it})/\phi$, respectively. Let v denote a variance function to indicate the mean-variance relation.

Assume that Y_i $(i = 1, \ldots, K)$ are independently distributed with mean vector $\mu_i = (\mu_{i1}, \ldots,$ $\mu_{in_i})^{\mathrm{T}}$ and variance matrix Σ_i with diagonal elements $\sigma_{i1}^2, \ldots, \sigma_{in_i}^2$ and off-diagonal elements $\rho_{itt'}\sigma_{it}\sigma_{it'}$ $(t, t' = 1, \ldots, n_i;\ t \neq t')$. We define an $n_i \times n_i$ diagonal matrix $A_i = \mathrm{diag}\{a''(\theta_{it})\}$ and let R_i be the correlation matrix with off-diagonal elements $\rho_{itt'}$; then, $\Sigma_i = A_i^{\frac{1}{2}} R_i A_i^{\frac{1}{2}}/\phi$. Let the regression coefficient β ($p \times 1$ vector) be the parameter to be estimated and ϕ be a nuisance parameter. Further, let $R(\alpha)$ be an $n \times n$ symmetric matrix, which fulfills the requirement of being a correlation matrix, and α be an $s \times 1$ vector, which fully characterizes $R(\alpha)$, where s is a suitable positive integer. We refer to $R(\alpha)$ as a 'working' correlation matrix. An independent structure with an identity matrix, an exchangeable structure with $\rho_{itt'} = \alpha$, and an AR(1) structure with $\rho_{itt'} = \alpha^{|t-t'|}$ are commonly used as the working correlation structure.

The GEE method identifies the estimator $\hat{\beta}$ of parameter β as the solution to Equation (2), substituting ϕ with a $K^{\frac{1}{2}}$-consistent estimator $\hat{\phi}(Y, \beta)$ after replacing α with a $K^{\frac{1}{2}}$-consistent estimator $\hat{\alpha}(Y, \beta, \phi)$:

$$U(\beta, \alpha) \equiv \sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} S_i = 0, \tag{2}$$

where D_i is an $n_i \times p$ matrix defined by $D_i = \partial\mu_i/\partial\beta$, $V_i = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}}/\phi$, and $S_i = Y_i - \mu_i$. ϕ can be estimated by

$$\hat{\phi}^{-1} = \frac{1}{\left(\sum_{i=1}^{K} n_i - p\right)} \sum_{i=1}^{K}\sum_{t=1}^{n_i} \hat{r}_{it}^2,$$

where $\hat{r}_{it} = (Y_{it} - \hat{\mu}_{it})/\sqrt{v(\hat{\mu}_{it})}$. When exchangeable structure is specified as the working correlation structure, given ϕ, α can be estimated by follows:

$$\hat{\alpha} = \frac{\phi}{\left\{\frac{1}{2}\sum_{i=1}^{K} n_i(n_i - 1) - p\right\}} \sum_{i=1}^{K}\sum_{t<t'} \hat{r}_{it}\hat{r}_{it'}.$$

Also, when AR(1) structure is specified as the working correlation structure, given ϕ, α can be estimated by follows:

$$\hat{\alpha} = \frac{\phi}{\left\{\sum_{i=1}^{K} (n_i - 1) - p\right\}} \sum_{i=1}^{K}\sum_{t\le n_i-1} \hat{r}_{it}\hat{r}_{i,t+1}.$$

The variance matrix V_r of $\hat{\beta}$ from the GEE method, which is referred to as the robust variance, is given by Equation (3):

$$V_r = \left(\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} D_i\right)^{-1} \left\{\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} var(Y_i) V_i^{-1} D_i\right\} \left(\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} D_i\right)^{-1}. \quad (3)$$

The variance estimate $\hat{V}_r$ of $\hat{\beta}$ can be obtained by replacing $var(Y_i)$ with $S_i S_i^{\mathrm{T}}$ and β, ϕ, α with their estimates in Equation (3).

3. Criteria for Selecting the Working Correlation Structure

3.1 QIC

AIC is a well-known criterion for likelihood-based model selection. However, we cannot apply a criterion such as AIC to the GEE approach, since the GEE is not likelihood based. Pan (2001a) proposed a criterion based on quasi-likelihood, named QIC, to select the proper mean model or the working correlation structure.

Hardin and Hilbe (2003) defined the quasi-likelihood function as follows:

$$Q(\mu, \phi; y) = \int_y^{\mu} \frac{\phi(y - \mu^*)}{v(\mu^*)} d\mu^*.$$

According to Pan (2001a), under the assumptions that the subjects and time points are independent, the quasi-likelihood for the longitudinal data is, in fact, calculated by

$$Q(\mu, \phi) = \sum_{i=1}^{K}\sum_{t=1}^{n_i} Q(\mu, \phi; Y_{it}).$$

QIC, as proposed by Pan (2001a), can be expressed as

$$\mathrm{QIC}(R) = -2Q(\hat{\beta}, \hat{\phi}) + 2tr\left(\hat{\Omega}_I \hat{V}_r\right), \tag{4}$$

where tr refers to the sum of the diagonal elements of the matrix. The quasi-likelihood—the first term in Equation (4)—can be written as a function of $\hat{\beta}$ replacing $\hat{\mu}$. Ω_I is given by Equation (5):

$$\Omega_I = \sum_{i=1}^{K} D_i^{\mathrm{T}} A_i^{-1} D_i. \tag{5}$$

$\hat{\Omega}_I$ in Equation (4) can be obtained by replacing β, ϕ, α with their estimates in Equation (5). $\hat{V}_r$ in Equation (4) is the robust variance estimate corresponding with Equation (3). Pan (2001a) proposed a method for selecting the correlation structure that minimizes $\mathrm{QIC}(R)$ as defined by the working correlation structure expressed in Equation (4).

3.2 CIC

Hin and Wang (2009) introduced CIC as a modification of QIC to improve its performance and proposed a method for selecting the correlation structure that minimizes $\mathrm{CIC}(R)$ as defined by the working correlation structure expressed in Equation (6):

$$\mathrm{CIC}(R) = tr\left(\hat{\Omega}_I \hat{V}_r\right). \tag{6}$$

CIC is constructed using only the second term that represents the penalty of QIC in Equation (4). Much like when calculating QIC, $\hat{\Omega}_I$ in Equation (6) can be obtained by replacing β, ϕ, α with their estimates in Equation (5) and $\hat{V}_r$ in Equation (6) is the robust variance estimate corresponding with Equation (3).

3.3 Modified QIC and CIC

$\mathrm{QIC}(R)$ in Equation (4) and $\mathrm{CIC}(R)$ in Equation (6) involve the robust variance estimate $\hat{V}_r$ of the GEE estimation corresponding with Equation (3). The robust variance V_r is expected to underestimate the variance of $\hat{\beta}$ when the sample size is small, as pointed out by Mancl and DeRouen (2001). Mancl and DeRouen (2001), Kauermann and Carroll (2001), and Pan (2001b) proposed some alternative variance estimators for $\hat{\beta}$ to improve the small-sample bias.

Mancl and DeRouen (2001) provided a bias-corrected variance estimator for $\hat{\beta}$:

$$\begin{aligned}\hat{V}_{\mathrm{MD}} = &\left(\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} D_i\right)^{-1} \\ &\times \left\{\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} (I_i - H_i)^{-1} S_i S_i^{\mathrm{T}} (I_i - H_i^{\mathrm{T}})^{-1} V_i^{-1} D_i\right\} \left(\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} D_i\right)^{-1},\end{aligned} \tag{7}$$

where $H_i = D_i(\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} D_i)^{-1} D_i^{\mathrm{T}} V_i^{-1}$ and I_i is an identity matrix of the same dimensions as H_i.

Kauermann and Carroll (2001) provided another bias-corrected variance estimator for $\hat{\beta}$:

$$\hat{V}_{\mathrm{KC}} = \left(\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} D_i\right)^{-1} \times \left\{\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} (I_i - H_i)^{-1/2} S_i S_i^{\mathrm{T}} (I_i - H_i^{\mathrm{T}})^{-1/2} V_i^{-1} D_i\right\} \left(\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} D_i\right)^{-1}. \quad (8)$$

Pan (2001b) proposed a variance estimator for $\hat{\beta}$ in Equation (9) for small-sample adjustment:

$$\hat{V}_{\mathrm{PA}} = \left(\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} D_i\right)^{-1} \times \left\{\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} A_i^{1/2} \left(\frac{1}{K}\sum_{i=1}^{K} A_i^{-1/2} S_i S_i^{\mathrm{T}} A_i^{-1/2}\right) A_i^{1/2} V_i^{-1} D_i\right\} \left(\sum_{i=1}^{K} D_i^{\mathrm{T}} V_i^{-1} D_i\right)^{-1}. \quad (9)$$

In practice, the variance estimates $\hat{V}_{\mathrm{MD}}$, $\hat{V}_{\mathrm{KC}}$, and $\hat{V}_{\mathrm{PA}}$ of $\hat{\beta}$ can be obtained by substituting β, ϕ, α with their estimates in Equations (7) to (9), respectively.

We modify $\mathrm{QIC}(R)$ and $\mathrm{CIC}(R)$ using the three bias-corrected variance estimates for $\hat{\beta}$ shown by Equations (7) to (9). In fact, the robust variance estimate $\hat{V}_r$ in Equations (4) and (6) is replaced by the variance estimates $\hat{V}_{\mathrm{MD}}$, $\hat{V}_{\mathrm{KC}}$, and $\hat{V}_{\mathrm{PA}}$. Thus, the modifications of $\mathrm{QIC}(R)$ and $\mathrm{CIC}(R)$ are, respectively, expressed in Equations (10) and (11) as follows:

$$\begin{aligned}
\mathrm{QIC}_{\mathrm{MD}}(R) &= -2Q(\hat{\beta},\hat{\phi}) + 2tr\left(\hat{\Omega}_I \hat{V}_{\mathrm{MD}}\right),\\
\mathrm{QIC}_{\mathrm{KC}}(R) &= -2Q(\hat{\beta},\hat{\phi}) + 2tr\left(\hat{\Omega}_I \hat{V}_{\mathrm{KC}}\right),\\
\mathrm{QIC}_{\mathrm{PA}}(R) &= -2Q(\hat{\beta},\hat{\phi}) + 2tr\left(\hat{\Omega}_I \hat{V}_{\mathrm{PA}}\right),
\end{aligned} \quad (10)$$

$$\begin{aligned}
\mathrm{CIC}_{\mathrm{MD}}(R) &= tr\left(\hat{\Omega}_I \hat{V}_{\mathrm{MD}}\right),\\
\mathrm{CIC}_{\mathrm{KC}}(R) &= tr\left(\hat{\Omega}_I \hat{V}_{\mathrm{KC}}\right),\\
\mathrm{CIC}_{\mathrm{PA}}(R) &= tr\left(\hat{\Omega}_I \hat{V}_{\mathrm{PA}}\right).
\end{aligned} \quad (11)$$

In this section, we propose six methods for selecting the correlation structure that minimizes $\mathrm{QIC}_{\mathrm{MD}}(R)$, $\mathrm{QIC}_{\mathrm{KC}}(R)$, $\mathrm{QIC}_{\mathrm{PA}}(R)$, $\mathrm{CIC}_{\mathrm{MD}}(R)$, $\mathrm{CIC}_{\mathrm{KC}}(R)$, and $\mathrm{CIC}_{\mathrm{PA}}(R)$ as defined by the working correlation structure.

4. Simulation Study

4.1 Design

We conducted a simulation study to evaluate the performance of the proposed methods in terms of the proportion of selecting a true correlation structure. The performance of $\mathrm{QIC}_{\mathrm{MD}}(R)$, $\mathrm{QIC}_{\mathrm{KC}}(R)$, and $\mathrm{QIC}_{\mathrm{PA}}(R)$ in Equation (10), and of $\mathrm{CIC}_{\mathrm{MD}}(R)$, $\mathrm{CIC}_{\mathrm{KC}}(R)$, and $\mathrm{CIC}_{\mathrm{PA}}(R)$ in Equation (11) were compared with that of original $\mathrm{QIC}(R)$ in Equation (4) and original $\mathrm{CIC}(R)$ in Equation (6), respectively.

Simulation data were generated for the following scenario. We assumed a multivariate

Table 1. Proportion (%) of selecting the correlation structure when $n = 4$ for an exchangeable true correlation structure

		$K = 10$			$K = 30$			$K = 100$		
ρ	Criterion	IN	EX	AR	IN	EX	AR	IN	EX	AR
0.1	QIC(R)	35.8	31.9	32.3	25.3	40.8	33.9	20.8	51.0	28.2
	QIC$_{\mathrm{MD}}(R)$	35.2	31.9	33.0	20.3	45.6	34.0	17.3	54.9	27.8
	QIC$_{\mathrm{KC}}(R)$	35.5	31.7	32.8	22.6	43.4	34.1	18.9	53.0	28.1
	QIC$_{\mathrm{PA}}(R)$	35.7	32.3	32.0	16.4	47.9	35.7	11.8	59.7	28.5
	CIC(R)	32.2	33.9	33.9	17.0	46.5	36.6	9.9	61.8	28.3
	CIC$_{\mathrm{MD}}(R)$	32.7	33.0	34.4	13.8	50.5	35.6	7.9	65.1	27.1
	CIC$_{\mathrm{KC}}(R)$	32.5	33.5	34.0	15.3	48.7	36.1	8.9	63.6	27.6
	CIC$_{\mathrm{PA}}(R)$	32.6	33.6	33.9	5.6	54.7	39.7	0.6	73.1	26.4
0.3	QIC(R)	29.0	31.9	39.2	21.1	53.6	25.3	13.4	67.6	19.0
	QIC$_{\mathrm{MD}}(R)$	25.5	35.1	39.4	15.1	59.6	25.2	11.8	69.3	18.9
	QIC$_{\mathrm{KC}}(R)$	27.2	33.5	39.3	17.8	56.8	25.5	12.6	68.4	19.0
	QIC$_{\mathrm{PA}}(R)$	26.5	34.4	39.1	15.1	61.7	23.2	11.3	71.5	17.2
	CIC(R)	23.3	36.3	40.5	6.3	67.3	26.4	0.2	86.9	12.9
	CIC$_{\mathrm{MD}}(R)$	20.4	39.5	40.2	4.0	71.2	24.8	0.1	87.8	12.2
	CIC$_{\mathrm{KC}}(R)$	21.5	38.2	40.3	5.0	69.5	25.6	0.1	87.3	12.6
	CIC$_{\mathrm{PA}}(R)$	21.8	38.3	39.9	1.0	76.9	22.2	0.0	93.8	6.2
0.5	QIC(R)	32.4	35.6	32.0	17.7	60.7	21.6	13.6	71.1	15.3
	QIC$_{\mathrm{MD}}(R)$	28.9	37.7	33.4	13.4	64.8	21.9	12.4	72.4	15.2
	QIC$_{\mathrm{KC}}(R)$	30.6	36.7	32.7	15.3	62.9	21.8	12.9	71.7	15.4
	QIC$_{\mathrm{PA}}(R)$	31.9	36.1	32.0	15.1	66.3	18.6	13.3	73.5	13.2
	CIC(R)	22.2	40.8	37.0	1.3	76.2	22.5	0.0	92.1	7.9
	CIC$_{\mathrm{MD}}(R)$	17.6	44.6	37.8	0.9	77.9	21.2	0.0	92.5	7.5
	CIC$_{\mathrm{KC}}(R)$	19.7	42.7	37.5	1.1	77.2	21.8	0.0	92.4	7.6
	CIC$_{\mathrm{PA}}(R)$	21.6	42.5	35.9	0.1	84.3	15.6	0.0	97.7	2.4

IN, independent; EX, exchangeable; AR, AR(1).

binomial distribution with mean μ_{it} for the outcome variable Y_{it}. The mean model is $\mathrm{logit}(\mu_{it}) = \beta_0 + \beta_1 x_{it}$ with $\beta_0 = -0.7$ and $\beta_1 = 0.2$. The true correlation structures are exchangeable or AR(1) with $\rho_{itt'} = 0.1, 0.3, 0.5$. x_{it} is a covariate and is generated using a multivariate binomial distribution with mean 0.5. The covariate x_{it} for the same subject are correlated with each time point on the assumption that the between-subject and within-subject variations are equivalent. We assumed the sample size K to be 10, 30, and 100 and the number of time points n as 4 and 8. Data generation was repeated 10000 times.

4.2 Results

Tables 1 to 4 list the results of the simulation study in terms of the proportion (out of 10000 simulations) of selecting the true correlation structure from three given structures (independent, exchangeable, and AR(1)) by the proposed methods using QIC$_{\mathrm{MD}}(R)$, QIC$_{\mathrm{KC}}(R)$, QIC$_{\mathrm{PA}}(R)$, CIC$_{\mathrm{MD}}(R)$, CIC$_{\mathrm{KC}}(R)$, and CIC$_{\mathrm{PA}}(R)$, and by the existent methods using original QIC(R) and CIC(R).

Tables 1 and 2 show the results when the true correlation structure is exchangeable and the number of time points is 4 or 8, respectively. Tables 3 and 4 show the results when the true

Table 2. Proportion (%) of selecting the correlation structure when $n = 8$ for an exchangeable true correlation structure

		$K = 10$			$K = 30$			$K = 100$		
ρ	Criterion	IN	EX	AR	IN	EX	AR	IN	EX	AR
0.1	$\mathrm{QIC}(R)$	23.2	45.0	31.8	18.9	50.3	30.9	8.4	61.8	29.9
	$\mathrm{QIC}_{\mathrm{MD}}(R)$	19.3	49.8	30.9	13.5	58.3	28.2	6.9	65.4	27.7
	$\mathrm{QIC}_{\mathrm{KC}}(R)$	21.1	47.6	31.3	16.0	54.7	29.3	7.7	63.4	28.9
	$\mathrm{QIC}_{\mathrm{PA}}(R)$	20.0	47.2	32.8	10.3	57.4	32.3	4.1	68.0	27.9
	$\mathrm{CIC}(R)$	18.5	49.0	32.5	11.6	63.0	25.4	2.4	84.5	13.2
	$\mathrm{CIC}_{\mathrm{MD}}(R)$	16.6	52.7	30.6	8.4	69.5	22.1	1.7	86.9	11.4
	$\mathrm{CIC}_{\mathrm{KC}}(R)$	17.5	51.4	31.0	9.9	66.5	23.6	2.0	85.7	12.3
	$\mathrm{CIC}_{\mathrm{PA}}(R)$	14.9	52.1	33.0	2.8	75.2	21.9	0.0	95.9	4.1
0.3	$\mathrm{QIC}(R)$	26.4	33.0	40.7	11.3	59.3	29.4	7.5	65.0	27.5
	$\mathrm{QIC}_{\mathrm{MD}}(R)$	21.3	39.0	39.7	7.6	64.8	27.6	6.7	66.7	26.6
	$\mathrm{QIC}_{\mathrm{KC}}(R)$	23.6	36.3	40.1	9.3	62.1	28.6	7.1	65.9	27.0
	$\mathrm{QIC}_{\mathrm{PA}}(R)$	24.5	35.0	40.5	8.6	62.9	28.5	6.8	67.0	26.3
	$\mathrm{CIC}(R)$	18.9	42.0	39.1	1.1	83.4	15.5	0.0	97.3	2.7
	$\mathrm{CIC}_{\mathrm{MD}}(R)$	15.0	47.7	37.4	0.7	86.6	12.7	0.0	97.7	2.3
	$\mathrm{CIC}_{\mathrm{KC}}(R)$	16.9	44.9	38.2	0.8	85.1	14.1	0.0	97.5	2.5
	$\mathrm{CIC}_{\mathrm{PA}}(R)$	16.2	43.7	40.2	0.1	91.5	8.4	0.0	99.8	0.2
0.5	$\mathrm{QIC}(R)$	31.7	35.1	33.3	12.6	60.2	27.2	11.2	65.3	23.5
	$\mathrm{QIC}_{\mathrm{MD}}(R)$	27.4	37.9	34.7	9.5	64.4	26.1	10.3	66.6	23.1
	$\mathrm{QIC}_{\mathrm{KC}}(R)$	29.5	36.4	34.1	10.9	62.5	26.6	10.7	66.0	23.3
	$\mathrm{QIC}_{\mathrm{PA}}(R)$	30.7	35.4	33.9	11.8	63.5	24.7	11.1	66.9	22.0
	$\mathrm{CIC}(R)$	20.3	43.6	36.1	0.1	84.5	15.4	0.0	97.2	2.8
	$\mathrm{CIC}_{\mathrm{MD}}(R)$	15.8	48.7	35.5	0.1	86.5	13.4	0.0	97.5	2.5
	$\mathrm{CIC}_{\mathrm{KC}}(R)$	17.9	46.4	35.8	0.1	85.5	14.4	0.0	97.3	2.7
	$\mathrm{CIC}_{\mathrm{PA}}(R)$	19.3	44.5	36.1	0.0	91.6	8.4	0.0	99.8	0.2

IN, independent; EX, exchangeable; AR, AR(1).

correlation structure is AR(1) and the number of time points is 4 or 8, respectively.

The results shown in Tables 1 to 4 indicate that the proportion of selecting a true correlation structure by the proposed methods using $\mathrm{QIC}_{\mathrm{MD}}(R)$, $\mathrm{QIC}_{\mathrm{KC}}(R)$, and $\mathrm{QIC}_{\mathrm{PA}}(R)$ was generally higher than that by the existent method using $\mathrm{QIC}(R)$, regardless of ρ and the true correlation structure when $K \geq 30$. In addition, the proportion of selecting a true correlation structure by $\mathrm{QIC}_{\mathrm{PA}}(R)$ was the highest among the three modified QIC, regardless of ρ and the true correlation structure when $K \geq 30$. Likewise, the proportion of selecting a true correlation structure by the proposed methods using $\mathrm{CIC}_{\mathrm{MD}}(R)$, $\mathrm{CIC}_{\mathrm{KC}}(R)$, and $\mathrm{CIC}_{\mathrm{PA}}(R)$ was mostly higher than that by the existent method using $\mathrm{CIC}(R)$; further, the proportion of selecting a true correlation structure by $\mathrm{CIC}_{\mathrm{PA}}(R)$ was also the highest among the three modified CIC, regardless of ρ and the true correlation structure when $K \geq 30$.

As shown in Tables 1 to 4, the proportion of selecting a true correlation structure is quite low when the sample size and the outcome variable correlation are small, regardless of the applied criterion for selecting the working correlation structure. In addition, the proportion of selecting

Table 3. Proportion (%) of selecting the correlation structure when $n = 4$ for an AR(1) true correlation structure

		$K = 10$			$K = 30$			$K = 100$		
ρ	Criterion	IN	EX	AR	IN	EX	AR	IN	EX	AR
0.1	$\mathrm{QIC}(R)$	35.1	32.8	32.1	25.1	35.2	39.8	22.8	30.7	46.5
	$\mathrm{QIC}_{\mathrm{MD}}(R)$	34.6	32.6	32.9	21.0	36.3	42.7	19.0	32.4	48.6
	$\mathrm{QIC}_{\mathrm{KC}}(R)$	34.9	32.7	32.5	22.7	35.8	41.5	21.0	31.4	47.6
	$\mathrm{QIC}_{\mathrm{PA}}(R)$	35.2	33.1	31.7	15.5	41.7	42.8	12.5	34.5	53.0
	$\mathrm{CIC}(R)$	32.3	34.3	33.4	17.5	37.2	45.3	13.5	30.4	56.1
	$\mathrm{CIC}_{\mathrm{MD}}(R)$	33.3	32.0	34.7	15.2	37.9	46.9	11.3	31.7	57.1
	$\mathrm{CIC}_{\mathrm{KC}}(R)$	32.8	33.1	34.1	16.1	37.5	46.4	12.5	31.0	56.6
	$\mathrm{CIC}_{\mathrm{PA}}(R)$	32.7	34.3	33.0	6.3	42.8	50.9	1.8	32.1	66.1
0.3	$\mathrm{QIC}(R)$	32.0	32.5	35.5	21.1	25.2	53.7	12.8	23.2	64.0
	$\mathrm{QIC}_{\mathrm{MD}}(R)$	30.4	33.5	36.2	15.5	27.6	57.0	10.7	24.3	65.0
	$\mathrm{QIC}_{\mathrm{KC}}(R)$	31.1	33.2	35.7	18.1	26.4	55.5	11.8	23.7	64.5
	$\mathrm{QIC}_{\mathrm{PA}}(R)$	31.7	33.0	35.4	13.4	26.3	60.3	10.5	20.5	69.0
	$\mathrm{CIC}(R)$	27.6	32.6	39.8	8.6	21.7	69.7	0.8	11.7	87.5
	$\mathrm{CIC}_{\mathrm{MD}}(R)$	26.0	32.3	41.6	5.9	23.7	70.4	0.6	12.1	87.3
	$\mathrm{CIC}_{\mathrm{KC}}(R)$	27.2	32.3	40.6	7.2	22.7	70.2	0.7	11.9	87.4
	$\mathrm{CIC}_{\mathrm{PA}}(R)$	27.6	32.2	40.2	1.5	18.0	80.5	0.0	4.7	95.3
0.5	$\mathrm{QIC}(R)$	31.6	34.2	34.2	15.2	24.2	60.7	10.7	21.8	67.5
	$\mathrm{QIC}_{\mathrm{MD}}(R)$	28.4	34.9	36.7	11.4	25.7	62.9	9.6	22.1	68.4
	$\mathrm{QIC}_{\mathrm{KC}}(R)$	30.3	34.4	35.3	13.2	25.0	61.9	10.1	21.9	68.0
	$\mathrm{QIC}_{\mathrm{PA}}(R)$	31.4	34.3	34.3	12.6	22.1	65.4	10.1	19.8	70.1
	$\mathrm{CIC}(R)$	22.2	32.1	45.8	1.6	15.9	82.5	0.0	4.2	95.8
	$\mathrm{CIC}_{\mathrm{MD}}(R)$	18.9	32.8	48.3	1.0	17.3	81.7	0.0	4.5	95.5
	$\mathrm{CIC}_{\mathrm{KC}}(R)$	20.5	32.4	47.1	1.3	16.6	82.1	0.0	4.3	95.7
	$\mathrm{CIC}_{\mathrm{PA}}(R)$	22.2	31.0	46.8	0.1	8.7	91.2	0.0	0.5	99.5

IN, independent; EX, exchangeable; AR, AR(1).

a true correlation structure by all criteria generally increased with K, n, and ρ. Furthermore, the proportion of selecting a true correlation structure by $\mathrm{QIC}(R)$, $\mathrm{QIC}_{\mathrm{MD}}(R)$, $\mathrm{QIC}_{\mathrm{KC}}(R)$, and $\mathrm{QIC}_{\mathrm{PA}}(R)$ was lower than that by $\mathrm{CIC}(R)$, $\mathrm{CIC}_{\mathrm{MD}}(R)$, $\mathrm{CIC}_{\mathrm{KC}}(R)$, and $\mathrm{CIC}_{\mathrm{PA}}(R)$, respectively, under most simulation conditions.

We thus conclude that the frequency of selecting a true correlation structure by the modified QIC and CIC is usually higher than that by the original QIC and CIC, respectively. In particular, the performance of $\mathrm{CIC}_{\mathrm{PA}}(R)$ was generally the best among all the criteria.

5. Application

In order to examine the practicability of the proposed criteria, we applied the proposed criteria to an air pollution data set that was reported in Stokes, Davis, and Koch (2000). The data studied the effect of air pollution on the health of 25 children. The outcome variable was a binary indicator of whether or not the subject wheezed, and was measured consistently four times yearly at ages 8, 9, 10, and 11.

Table 4. Proportion (%) of selecting a correlation structure when $n = 8$ for an AR(1) true correlation structure

		$K = 10$			$K = 30$			$K = 100$		
ρ	Criterion	IN	EX	AR	IN	EX	AR	IN	EX	AR
0.1	$\text{QIC}(R)$	33.3	32.6	34.2	24.7	35.8	39.5	23.2	28.9	48.0
	$\text{QIC}_{\text{MD}}(R)$	32.4	31.6	36.0	20.8	36.6	42.6	19.9	30.4	49.7
	$\text{QIC}_{\text{KC}}(R)$	32.9	32.1	35.1	22.6	36.3	41.1	21.5	29.5	49.0
	$\text{QIC}_{\text{PA}}(R)$	31.0	36.4	32.6	16.4	45.0	38.7	12.3	33.0	54.8
	$\text{CIC}(R)$	29.8	36.3	33.8	18.3	38.0	43.7	14.1	28.5	57.5
	$\text{CIC}_{\text{MD}}(R)$	30.6	33.1	36.3	15.3	38.2	46.5	11.9	29.3	58.8
	$\text{CIC}_{\text{KC}}(R)$	30.3	34.9	34.9	16.7	38.1	45.2	12.9	28.9	58.2
	$\text{CIC}_{\text{PA}}(R)$	28.2	40.5	31.3	8.0	46.2	45.8	2.3	28.8	69.0
0.3	$\text{QIC}(R)$	24.7	28.1	47.2	19.3	25.1	55.6	11.5	21.1	67.4
	$\text{QIC}_{\text{MD}}(R)$	18.6	30.5	50.9	14.0	27.2	58.9	10.0	21.5	68.6
	$\text{QIC}_{\text{KC}}(R)$	21.1	29.9	49.0	16.4	26.2	57.4	10.8	21.2	68.0
	$\text{QIC}_{\text{PA}}(R)$	20.0	32.2	47.9	13.0	24.2	62.8	8.9	17.7	73.4
	$\text{CIC}(R)$	17.0	30.7	52.3	7.6	18.6	73.8	0.7	6.6	92.7
	$\text{CIC}_{\text{MD}}(R)$	14.5	30.2	55.3	5.3	19.8	75.0	0.6	6.7	92.7
	$\text{CIC}_{\text{KC}}(R)$	15.7	30.5	53.8	6.4	19.3	74.4	0.7	6.6	92.7
	$\text{CIC}_{\text{PA}}(R)$	11.9	33.3	54.9	1.6	10.5	88.0	0.0	0.6	99.4
0.5	$\text{QIC}(R)$	21.8	29.1	49.2	13.3	22.7	64.0	9.3	18.7	72.0
	$\text{QIC}_{\text{MD}}(R)$	14.6	34.7	50.7	9.4	23.2	67.4	8.4	18.6	72.9
	$\text{QIC}_{\text{KC}}(R)$	18.1	31.9	50.0	11.1	23.1	65.8	8.9	18.7	72.5
	$\text{QIC}_{\text{PA}}(R)$	18.4	31.7	50.0	11.0	19.0	70.0	8.8	17.7	73.6
	$\text{CIC}(R)$	13.5	28.5	58.1	0.8	8.9	90.3	0.0	0.4	99.6
	$\text{CIC}_{\text{MD}}(R)$	9.1	31.2	59.7	0.5	9.2	90.3	0.0	0.4	99.6
	$\text{CIC}_{\text{KC}}(R)$	11.1	29.9	59.0	0.6	9.1	90.3	0.0	0.4	99.6
	$\text{CIC}_{\text{PA}}(R)$	10.1	30.0	59.8	0.1	1.6	98.3	0.0	0.0	100.0

IN, independent; EX, exchangeable; AR, AR(1).

We fitted the following logistic model to the data:

$$\text{logit}\,\{E(Y_{it})\} = \beta_0 + \beta_1\ \text{City}_i + \beta_2\ \text{Age}_{it} + \beta_3\ \text{Smoke}_{it},$$

where Y_{it} is the binary indicator of whether or not subject i wheezed at time t; $\text{City}_i = 0, 1$ indicates whether the child is a resident of Green Hills or Steel City, respectively; $\text{Age}_{it} = 8, 9, 10, 11$ denotes the child's age; and $\text{Smoke}_{it} = 0, 1, 2$ indicates a passive smoking index that reflected the degree of smoking in the home. Age_{it} and Smoke_{it} are the time-varying covariates. Three structures—independent, exchangeable, and AR(1)—are adopted as candidates for the working correlation structure.

The estimates of the regression parameters; square roots of the variances (standard errors) based on Equations (3), (7) to (9); p values of the z-test corresponding with each variance estimator; and the values of $\text{QIC}(R)$, $\text{QIC}_{\text{MD}}(R)$, $\text{QIC}_{\text{KC}}(R)$, $\text{QIC}_{\text{PA}}(R)$, $\text{CIC}(R)$, $\text{CIC}_{\text{MD}}(R)$, $\text{CIC}_{\text{KC}}(R)$, and $\text{CIC}_{\text{PA}}(R)$ obtained using each working correlation structure are presented in Table 5.

According to Table 5, the estimates of β_1 and β_2, which are related to city and age, respectively, were not significantly different among the three structures. Meanwhile, the estimates

Table 5. Parameter estimates, standard errors, p values, and criteria values for each working correlation structure in wheeze data

Covariate	Variance estimator	Independent Estimate SE (p value)	Exchangeable Estimate SE (p value)	AR(1) Estimate SE (p value)
Intercept		2.418	2.303	2.419
	OR	1.812 (0.182)	1.813 (0.204)	1.832 (0.187)
	MD	1.909 (0.205)	1.910 (0.228)	1.927 (0.209)
	KC	1.860 (0.194)	1.861 (0.216)	1.878 (0.198)
	PA	1.809 (0.181)	1.811 (0.203)	1.833 (0.187)
City		−0.002	−0.042	0.017
	OR	0.532 (0.997)	0.543 (0.939)	0.528 (0.974)
	MD	0.599 (0.998)	0.609 (0.945)	0.593 (0.977)
	KC	0.564 (0.998)	0.575 (0.942)	0.559 (0.976)
	PA	0.485 (0.997)	0.496 (0.933)	0.484 (0.972)
Age		−0.328	−0.320	−0.333
	OR	0.189 (0.083)	0.188 (0.089)	0.191 (0.081)
	MD	0.199 (0.099)	0.199 (0.107)	0.201 (0.098)
	KC	0.194 (0.091)	0.193 (0.098)	0.196 (0.089)
	PA	0.183 (0.072)	0.182 (0.079)	0.185 (0.073)
Smoking		0.560	0.651	0.599
	OR	0.291 (0.055)	0.282 (0.021)	0.279 (0.031)
	MD	0.333 (0.093)	0.318 (0.041)	0.315 (0.057)
	KC	0.311 (0.072)	0.299 (0.030)	0.296 (0.043)
	PA	0.315 (0.075)	0.318 (0.041)	0.314 (0.056)

Criterion	Independent	Exchangeable	AR(1)
$\text{QIC}(R)$	137.05	137.14	136.83
$\text{QIC}_{\text{MD}}(R)$	139.13	139.11	138.77
$\text{QIC}_{\text{KC}}(R)$	138.02	138.07	137.74
$\text{QIC}_{\text{PA}}(R)$	137.26	137.50	137.21
$\text{CIC}(R)$	4.46	4.46	4.34
$\text{CIC}_{\text{MD}}(R)$	5.51	5.45	5.31
$\text{CIC}_{\text{KC}}(R)$	4.95	4.93	4.80
$\text{CIC}_{\text{PA}}(R)$	4.57	4.64	4.53

SE refers to standard error. OR, MD, KC, and PA are variance estimators based on Equations (3), (7) to (9), respectively.

of β_3, which are related to smoking, somewhat varied between the specified structures. The standard errors for the effects of city and age were quite similar for all structures regardless of the variance estimator; however, the standard errors for the effect of smoking were slightly dependent on the specified structure. The p values for the effects of city and age were quite similar for all structures; on the other hand, the p values for the effect of smoking were dependent on the specified structure. In fact, the effect of smoking under the exchangeable structure was statistically significant regardless of the variance estimator, but the same effect under the independent structure was not. Further, the significance of the same effect under the AR(1) structure depended on the variance estimator.

The realized values of the modified QIC and CIC and the original QIC and CIC were minimized when AR(1) was specified. According to the results of this analysis, AR(1) was the most appropriate structure.

6. Discussion

Through the results of a simulation study, it was found that our proposed modifications to QIC and CIC based on the three bias-corrected variance estimators were considerably more accurate when it came to selecting the working correlation structure in the GEE method. In particular, $\mathrm{CIC}_{\mathrm{PA}}(R)$ generally had the best performance among all modified criteria. Unfortunately, much like the original CIC, the modified CIC cannot be used to select the appropriate covariate in a mean model. On the other hand, the modified QIC and the original QIC can be used not only for the selection of the correlation structures but also for covariate selection in the GEE method. The applicability of the original QIC as a criterion for selecting the covariate was evaluated by Pan (2001a) and Cui and Qian (2007). The performance of the modified QIC for model selection should be investigated in future researches.

REFERENCES

Cui, J. and Qian, G. (2007). Selection of working correlation structure and best model in GEE analyses of longitudinal data. *Communications in Statistics—Simulation and Computation* **36**, 987–996.

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

Hardin, J. and Hilbe, J. (2003). *Generalized Estimating Equations.* London: Chapman and Hall.

Hin, L. Y. and Wang, Y. G. (2009). Working-correlation-structure identification in generalized estimating equations. *Statistics in Medicine* **28**, 642–658.

Kauermann, G. and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* **96**, 1387–1396.

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

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

Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for GEE with improved small sample properties. *Biometrics* **57**, 126–134.

Pan, W. (2001a). Akaike's information criterion in generalized estimating equations. *Biometrics* **57**, 120–125.

Pan, W. (2001b). On the robust variance estimator in generalised estimating equations. *Biometrika* **88**, 901–906.

Stokes, M. E., Davis, C. S., and Koch, G. G. (2000). *Categorical Data Analysis Using The SAS System*, 2nd ed. North Carolina: SAS Institute.

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