J. R. Statist. Soc. B (2004)
66, *Part* 4, *pp.* 851–860

Efficiency of generalized estimating equations for binary responses

N. Rao Chaganty

Old Dominion University, Norfolk, USA

and Harry Joe

University of British Columbia, Vancouver, Canada

[Received October 2003. Revised February 2004]

Summary. Using standard correlation bounds, we show that in generalized estimation equations (GEEs) the so-called 'working correlation matrix' $\mathbf{R}(\alpha)$ for analysing binary data cannot in general be the true correlation matrix of the data. Methods for estimating the correlation parameter in current GEE software for binary responses disregard these bounds. To show that the GEE applied on binary data has high efficiency, we use a multivariate binary model so that the covariance matrix from estimating equation theory can be compared with the inverse Fisher information matrix. But $\mathbf{R}(\alpha)$ should be viewed as the weight matrix, and it should not be confused with the correlation matrix of the binary responses. We also do a comparison with more general weighted estimating equations by using a matrix Cauchy–Schwarz inequality. Our analysis leads to simple rules for the choice of α in an exchangeable or autoregressive AR(1) weight matrix $\mathbf{R}(\alpha)$, based on the strength of dependence between the binary variables. An example is given to illustrate the assessment of dependence and choice of α.

Keywords: Generalized estimating equations; Multivariate binary data; Odds ratio; Quasi-least squares; Repeated measurements

1. Introduction

The theoretical study of the method of generalized estimating equations (GEEs) for binary response data is inadequate partly because of the confusing meaning of the term 'working correlation matrix' that was introduced by Liang and Zeger (1986) in their seminal paper. Crowder (1995) pointed out that this matrix lacks a proper definition when the true correlation is misspecified, thus causing a breakdown of the asymptotic properties of the estimation procedure. Liang and Zeger (1986) have assumed that the working correlation matrix is the correlation matrix of the response vector $\mathbf{y}$, which is constant over the possible covariate vectors $\mathbf{x}$. However, for a non-normal random vector, in particular a binary random vector, this may be impossible. For dependent non-normal random variables, the range of correlation depends on the univariate marginals. The lower and upper bounds on the correlation come from maximal negative and positive dependence: the Fréchet bounds. Bernoulli random variables can be strongly positively dependent without the correlation coefficient being very high (Joe (1997), chapter 7).

Owing to the violation of the correlation bounds, Sutradhar and Das (1999) have invalid values for the efficiencies of GEEs applied to binary data under a misspecification of the

Address for correspondence: N. Rao Chaganty, Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529-0077, USA.
E-mail: rchagant@odu.edu

 1369–7412/04/66851

correlation structures. Moreover, comparisons of efficiency between GEE and maximum likelihood methods for binary responses have used likelihoods that are based on questionable assumptions. For example, Lipsitz *et al.* (1995), pages 563–564, considered likelihood for the binary responses based on the Bahadur representation with the rth-order parameters to be constant depending on r, whereas Liang *et al.* (1992), page 14, assumed that all the third- and higher order parameters are equal and constant. Diggle *et al.* (2002), page 145, acknowledged that

> 'if we assume that the marginal means (for binary responses) depend on covariates, $\mathbf{x}$, it may not be correct to assume that the correlations and higher-order moments are independent of $\mathbf{x}$, as would be convenient'.

In this paper we use a latent variable multivariate binary model for efficiency calculations. Such direct comparisons of efficiency have not previously been made in the literature. The score equations for the latent variable binary model *do not* look like the GEE, and analytically it is difficult to assess the efficiency of GEEs for this model. We use simulations to make the comparisons of efficiency.

The efficiency calculations show that the GEE method with the so-called working correlation matrix $\mathbf{R}(\alpha)$ does have good efficiency relative to a likelihood approach using a multivariate probit model. We believe that this result, in favour of the GEE, is new. However, $\mathbf{R}(\alpha)$ should be considered as a weight matrix in which the parameter α should be chosen to be larger if there is stronger dependence in the binary data. Considering a set of estimating equations based on more general weight matrices than those employed in GEEs, we use a matrix Cauchy–Schwarz inequality to determine the optimal weight matrices. The best choice of a constant $\mathbf{R}(\alpha)$ matrix, to approximate the efficiency from the optimal weight matrices, is one that roughly approximates the average correlation matrix over the data $\mathbf{y}_1, \ldots, \mathbf{y}_n$.

Our efficiency calculations suggest simple rules for choosing α for using GEEs with binary responses. In particular, it is not necessary to use an estimating equation for α. Existing methods for estimating α sometimes run into problems as there is no guarantee that the estimated value ensures that $\mathbf{R}(\alpha)$ is positive definite (Chaganty, 1997a; Shults and Chaganty, 1998).

In Section 2 we present the correlation bounds, and in Section 3 we study the asymptotic covariance matrices of estimators from various sets of weighted estimating equations. A multivariate binary model is presented in Section 4. In Section 5 we discuss our simulation model and results. Several comparisons of efficiency in the literature that show that GEEs are better than independent estimating equations (IEEs) or vice versa are incomplete because there has not been a check on how the efficiency varies as a function of the regression coefficients. As part of our analysis (Section 6) we show when IEEs can have poor efficiency. Section 7 has some simple guidelines on the choice of α for applying GEEs to binary data and Section 8 has an example. We conclude the paper with some discussion.

2. Correlation bounds for binary variables

If $\mathbf{y} = (y_1, \ldots, y_d)'$ is a Bernoulli random vector with marginal probabilities p_j with $q_j = 1 - p_j$ for $j = 1, \ldots, d$, and constant correlation ρ between any two pairs, then

$$\max_{j \neq k} \left\{ -\sqrt{\left(\frac{p_j p_k}{q_j q_k}\right)}, -\sqrt{\left(\frac{q_j q_k}{p_j p_k}\right)} \right\} \leqslant \rho \leqslant \min_{j \neq k} \left\{ \sqrt{\left(\frac{p_j q_k}{q_j p_k}\right)}, \sqrt{\left(\frac{q_j p_k}{p_j q_k}\right)} \right\}. \tag{1}$$

These inequalities are well known; see McDonald (1993), page 393, for example. If $p_j(\mathbf{x})$ is a function of a covariate vector $\mathbf{x}$, then a constant correlation matrix over all $\mathbf{x}$ would imply that the constant correlation must lie in the interval

$$\max_{\mathbf{x}}\left(\max_{j\neq k}\left[-\sqrt{\left\{\frac{p_j(\mathbf{x})\,p_k(\mathbf{x})}{q_j(\mathbf{x})\,q_k(\mathbf{x})}\right\}},-\sqrt{\left\{\frac{q_j(\mathbf{x})\,q_k(\mathbf{x})}{p_j(\mathbf{x})\,p_k(\mathbf{x})}\right\}}\right]\right)\leqslant\rho$$
$$\leqslant\min_{\mathbf{x}}\left(\min_{j\neq k}\left[\sqrt{\left\{\frac{p_j(\mathbf{x})\,q_k(\mathbf{x})}{q_j(\mathbf{x})\,p_k(\mathbf{x})}\right\}},\sqrt{\left\{\frac{q_j(\mathbf{x})\,p_k(\mathbf{x})}{p_j(\mathbf{x})\,q_k(\mathbf{x})}\right\}}\right]\right). \tag{2}$$

Inequality (2) shows that the constraints on ρ will depend on the covariates. These restrictions are necessary, but current GEE software ignores them. When the range of $\mathbf{x}$ is wide, the interval in inequality (2) can be quite narrow. For example, when $\mathbf{x}$ is normally distributed, as assumed by some researchers for comparisons of efficiency (e.g. Sutradhar and Das (2000), equation (2.1)), the interval for ρ given by inequality (2) reduces to a single point 0, thus defeating the purpose of modelling correlated binary data. Hence the assumption of a constant correlation over covariates is unreasonable for dependent binary variables. A proper analysis of the efficiency of GEEs applied to binary responses involves the correlation of each pair of $\mathbf{y}_i$s, which cannot be assumed to be constant over covariates for a proper model with a wide range of dependence.

3. Weighted estimating equations

Suppose that n independent binary vectors $\mathbf{y}_i=(y_{i1},\ldots,y_{id})'$ are observed, with y_{ij} distributed as Bernoulli(μ_{ij}), $\mu_{ij}=F(\boldsymbol{\beta}'\mathbf{x}_{ij})$, where F is the latent variable distribution; for example, F is the standard normal cumulative distribution function for a probit model, and a standard logistic cumulative distribution function for a logit model. Suppose that the dimension of $\boldsymbol{\beta}$ and $\mathbf{x}_{ij}$ is $p\times 1$ (the intercept terms with $x_{i1}=1$ could be included). If $y_{i1},\ldots,y_{id}$ are independent for each i, then the log-likelihood is

$$l(\boldsymbol{\beta})=\sum_{i=1}^{n}\sum_{j=1}^{d}\{y_{ij}\log(\mu_{ij})+(1-y_{ij})\log(1-\mu_{ij})\},$$

and the likelihood equations are

$$\frac{\partial l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{n}\sum_{j=1}^{d}\frac{y_{ij}-\mu_{ij}}{\mu_{ij}(1-\mu_{ij})}\frac{\partial\mu_{ij}}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{n}\frac{\partial\boldsymbol{\mu}_i'}{\partial\boldsymbol{\beta}}\mathbf{A}_i^{-1}(\mathbf{y}_i-\boldsymbol{\mu}_i)=0, \tag{3}$$

where $\boldsymbol{\mu}_i=(\mu_{i1},\ldots,\mu_{id})'$ and $\mathbf{A}_i=\mathrm{diag}\{\mu_{i1}(1-\mu_{i1}),\ldots,\mu_{id}(1-\mu_{id})\}$. We refer to equations (3) as IEEs. With $f=F'$, expression (3) reduces to

$$\frac{\partial l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{n}\mathbf{X}_i'\boldsymbol{\Delta}_i\mathbf{A}_i^{-1}(\mathbf{y}_i-\boldsymbol{\mu}_i)=0, \tag{4}$$

since $\partial\boldsymbol{\mu}_i'/\partial\boldsymbol{\beta}=\mathbf{X}_i'\boldsymbol{\Delta}_i$ where $\mathbf{X}_i'=(\mathbf{x}_{i1},\ldots,\mathbf{x}_{id})$ and $\boldsymbol{\Delta}_i=\mathrm{diag}\{f(\boldsymbol{\beta}'\mathbf{x}_{i1}),\ldots,f(\boldsymbol{\beta}'\mathbf{x}_{id})\}$. Furthermore, if F is the standard logistic cumulative distribution function we have $\boldsymbol{\Delta}_i=\mathbf{A}_i$ and equation (4) reduces to

$$\frac{\partial l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{n}\mathbf{X}_i'(\mathbf{y}_i-\boldsymbol{\mu}_i)=0.$$

The idea in Liang and Zeger (1986) is that this set of estimating equations or generalized versions of them can be used to estimate $\boldsymbol{\beta}$, even for dependent data. The GEEs for dependent data are obtained by replacing $\mathbf{A}_i$ in expression (3) with a symmetric weight matrix $\mathbf{W}_i$ which has the same diagonal elements. The GEE method uses specific forms for the $\mathbf{W}_i$.

Consider the weighted estimating equations

$$\psi = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \mathbf{W}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = 0. \tag{5}$$

These are unbiased estimating equations for any set of $\mathbf{W}_i$ since $E(\mathbf{y}_i) = \boldsymbol{\mu}_i$. If $\hat{\boldsymbol{\beta}}$ is the solution of equation (5) then from the theory of estimating equations (Godambe, 1991) we have $\hat{\boldsymbol{\beta}}$ is AN$(\boldsymbol{\beta}, \mathbf{V})$ where $\mathbf{V} = (-\mathbf{D}_{\psi}^{-1})\mathbf{M}_{\psi}(-\mathbf{D}_{\psi}^{-1})'$, $\mathbf{M}_{\psi} = \mathrm{cov}(\psi)$ and $\mathbf{D}_{\psi} = E(\partial\psi/\partial\boldsymbol{\beta}')$. Here AN denotes 'asymptotically normal'. Suppose that $\mathrm{cov}(\mathbf{y}_i) = \boldsymbol{\Sigma}_i$; then

$$\mathbf{M}_{\psi} = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \mathbf{W}_i^{-1} \boldsymbol{\Sigma}_i \mathbf{W}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'},$$

$$-\mathbf{D}_{\psi} = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \mathbf{W}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'},$$

since only one of the three derivative terms has a non-zero expectation. Chaganty's (1997b) matrix Cauchy–Schwarz inequality is that

$$\left(\sum_{i=1}^{n} B_i' C_i\right)^{-1} \left(\sum_{i=1}^{n} B_i' \boldsymbol{\Sigma}_i B_i\right) \left(\sum_{i=1}^{n} C_i' B_i\right)^{-1} - \left(\sum_{i=1}^{n} C_i' \boldsymbol{\Sigma}_i^{-1} C_i\right)^{-1} \tag{6}$$

is non-negative definite for any B_i and C_i of appropriate dimensions. With $C_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}'$, $B_i = \mathbf{W}_i^{-1}\partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}'$, the first term in expression (6) is $\mathbf{V}$ and the second term is

$$\left(\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'}\right)^{-1}$$

so the optimal choice of $\mathbf{W}_i$ is $\boldsymbol{\Sigma}_i$. This choice of $\mathbf{W}_i$ depends on the unknown $\boldsymbol{\beta}$ and on the dependence parameters. With a given multivariate binary model such as the probit model, $\boldsymbol{\Sigma}_i$ can be easily computed from bivariate marginal probabilities. For GEEs, consider weight matrices of the form $\mathbf{W}_i = \mathbf{A}_i^{1/2}\mathbf{R}(\alpha)\mathbf{A}_i^{1/2}$ where $\mathbf{R}(\alpha)$ is an exchangeable or autoregressive AR(1) correlation matrix. A conjecture based on the above inequality (6) would be that the best choice in this class is with the value of α and a structured correlation matrix $\mathbf{R}(\alpha)$ that is close to the average correlation matrix over the $\mathbf{y}_i$s, i.e.

$$\mathbf{R}(\alpha) \approx \frac{1}{n} \sum_{i=1}^{n} \mathbf{A}_i^{-1/2} \boldsymbol{\Sigma}_i \mathbf{A}_i^{-1/2}. \tag{7}$$

Approximation (7) is loose, as there is no reason to expect it to be equally good for all elements of the matrix. Indeed the right-hand side of approximation (7) may not resemble a structured correlation matrix. We can compare $\mathbf{V}$ for various choices of $\mathbf{W}_i$ with different $\mathbf{R}(\alpha)$ and also compare $\mathbf{V}$ with the optimal choice of $\mathbf{W}_i$. To understand the effect of α in $\mathbf{R}(\alpha)$ and the optimal $\mathbf{W}_i$, we compute $\mathbf{V}$ for these various cases as well as the inverse Fisher information matrix (the asymptotic covariance matrix for the maximum likelihood estimator). For calculating $\boldsymbol{\Sigma}_i$ and the information matrix, we need a multivariate binary model, such as the model that is given in the next section.

4. Likelihood approach for correlated binary data

We choose the multivariate probit model for comparisons of efficiency because it is a commonly used model for multivariate binary data. It is widely used in psychometrics as a latent variable model (Muthén, 1978; Maydeu-Olivares, 2001); in genetics (Mendell and Elston, 1974; Szudek

et al., 2002), the latent correlations are of primary interest for genetic hypotheses. Suppose that $\mathbf{y} = (y_1, \ldots, y_d)'$ is a multivariate binary vector with covariate matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_d)'$. The multivariate probit model (Ashford and Sowden (1970) and Joe (1997), chapter 7) has the stochastic representation

$$y_j = I(z_j \leqslant \nu_j = \boldsymbol{\beta}'\mathbf{x}_j), \qquad j = 1, \ldots, d,$$

where I is the indicator function and $\mathbf{Z} = (z_1, \ldots, z_d)'$ is a latent variable distributed as multivariate normal with mean 0 and covariance matrix $\boldsymbol{\Omega}(\gamma)$, where γ is the latent correlation. For an AR(1) probit model, $\boldsymbol{\Omega}(\gamma) = (\gamma^{|j-k|})$ and, for an exchangeable probit model, $\boldsymbol{\Omega}(\gamma) = (1-\gamma)\mathbf{I} + \gamma\mathbf{J}$ where $\mathbf{I}$ is the identity and $\mathbf{J}$ is a matrix of 1s. If $\Phi_2(\omega_1, \omega_2; \gamma)$ and $\Phi(\omega)$ denote the standardized bivariate normal with correlation γ and the univariate standard normal distribution functions respectively, then

$$\operatorname{corr}(y_j, y_k) = \frac{\Phi_2(\nu_j, \nu_k; \gamma) - \Phi(\nu_j)\,\Phi(\nu_k)}{[\Phi(\nu_j)\{1 - \Phi(\nu_j)\}\,\Phi(\nu_k)\{1 - \Phi(\nu_k)\}]^{1/2}}. \tag{8}$$

When $\boldsymbol{\beta} = 0$, we have $\nu_j = 0$ for all j and equation (8) reduces to $2\sin^{-1}(\gamma)/\pi$, independent of j and k; in general, $2\sin^{-1}(\gamma)/\pi$ is an upper bound on the correlation. This relationship could be used as a guideline between the correlation of the binary variables and the latent correlation γ. The correlation between the binary variables is always less than the latent correlation γ in the range $(0, 1)$; for $\gamma = 0.1, 0.3, 0.5, 0.7, 0.9$, the values of $2\sin^{-1}(\gamma)/\pi$ are 0.064, 0.194, 0.333, 0.494 and 0.713 respectively. Let

$$\pi_d(\mathbf{y}; \mathbf{X}, \boldsymbol{\beta}) = \Pr(\mathbf{Y} = \mathbf{y}; \mathbf{X}, \boldsymbol{\beta}), \qquad \mathbf{y} \in \{0, 1\}^d,$$

be the probabilities based on the multivariate probit model. Calculation of these probabilities involve multivariate normal rectangle probabilities, which can be reduced to one-dimensional numerical integrals when $\boldsymbol{\Omega}(\gamma)$ is an exchangeable correlation matrix and $0 < \gamma < 1$; see page 134 of Kotz *et al.* (2000). When γ is known, on the basis of a sample of n observations $\mathbf{y}_i$ and covariates $\mathbf{X}_i$, the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_L$ is obtained by maximizing the likelihood function

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi_d(\mathbf{y}_i; \mathbf{X}_i, \boldsymbol{\beta})$$

with respect to $\boldsymbol{\beta}$. For known γ, the Fisher information matrix is calculated from a sample of n observations by using the formula

$$\mathcal{I} = \sum_{i=1}^{n} \sum_{\mathbf{y}} \frac{\partial \pi_d(\mathbf{y}; \mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \pi_d(\mathbf{y}; \mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \Big/ \pi_d(\mathbf{y}; \mathbf{X}_i, \boldsymbol{\beta})$$

where the inner sum is taken over the 2^d possible vectors $\mathbf{y}$. The inverse of $\mathcal{I}$ gives the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_L$ (Lehmann (1998), page 545). The extension is straightforward when γ is unknown or when $\boldsymbol{\Omega}(\gamma)$ is AR(1) or an unknown unstructured correlation matrix. Plackett's identity (Kotz *et al.* (2000), page 259) is useful to derive the Fisher information for the correlation parameters when the unstructured correlation matrix $\boldsymbol{\Omega}$ is estimated along with the regression parameter $\boldsymbol{\beta}$ by maximum likelihood.

5. Findings on efficiency

For studying the performance of GEEs, we took $p = 2$, $\mathbf{x}_{ij} = (1, x_{ij})'$ where x_{ij} are taken as uniform random variables in the interval $[-1, 1]$. We selected values for $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ to obtain a range of slopes and average marginal probabilities. We chose $\beta_0 = 0$, 0.42, 1.1 for marginal

Table 1. Diagonal elements scaled by n and efficiencies of the inverse Fisher information matrix for the exchangeable probit model and of $\mathbf{V}$ with optimal weights and exchangeable $\mathbf{R}(\alpha)$ for $\alpha = 0.0, 0.1, 0.2, \ldots, 0.9$†

Method	$n\,V(\beta_0)$	$n\,V(\beta_1)$
Maximum likelihood	0.735 (1.000)	0.807 (1.000)
Optimal	0.739 (0.994)	0.828 (0.974)
$\alpha = 0.0$	0.739 (0.994)	1.033 (0.781)
$\alpha = 0.1$	0.739 (0.994)	0.884 (0.912)
$\alpha = 0.2$	0.739 (0.994)	0.839 (0.962)
$\alpha = 0.3$	0.739 (0.994)	0.829 (0.974)
$\alpha = 0.4$	0.740 (0.994)	0.832 (0.969)
$\alpha = 0.5$	0.740 (0.993)	0.841 (0.959)
$\alpha = 0.6$	0.741 (0.992)	0.852 (0.947)
$\alpha = 0.7$	0.743 (0.989)	0.863 (0.936)
$\alpha = 0.8$	0.751 (0.979)	0.873 (0.924)
$\alpha = 0.9$	0.801 (0.918)	0.883 (0.914)

†The parameter values are $d = 5$, $\gamma = 0.5$, $\beta_0 = 0.0$, $\beta_1 = 0.5$ and $n = 500$; efficiencies are given in parentheses.

Table 2. Diagonal elements scaled by n and efficiencies of the inverse Fisher information matrix for the exchangeable probit model and of $\mathbf{V}$ with optimal weights and exchangeable $\mathbf{R}(\alpha)$ for $\alpha = 0.0, 0.1, 0.2, \ldots, 0.9$†

Method	$n\,V(\beta_0)$	$n\,V(\beta_1)$
Maximum likelihood	1.423 (1.000)	1.355 (1.000)
Optimal	1.427 (0.997)	1.405 (0.964)
$\alpha = 0.0$	1.428 (0.997)	2.581 (0.525)
$\alpha = 0.1$	1.428 (0.997)	2.258 (0.600)
$\alpha = 0.2$	1.427 (0.997)	1.987 (0.682)
$\alpha = 0.3$	1.427 (0.997)	1.768 (0.766)
$\alpha = 0.4$	1.427 (0.997)	1.602 (0.846)
$\alpha = 0.5$	1.427 (0.997)	1.487 (0.911)
$\alpha = 0.6$	1.427 (0.997)	1.423 (0.952)
$\alpha = 0.7$	1.427 (0.997)	1.411 (0.960)
$\alpha = 0.8$	1.427 (0.997)	1.450 (0.935)
$\alpha = 0.9$	1.428 (0.997)	1.539 (0.880)

†The parameter values are $d = 2$, $\gamma = 0.9$, $\beta_0 = 0.42$, $\beta_1 = 0.25$ and $n = 500$; efficiencies are given in parentheses.

'average' probabilities of 0.5, 0.7 and 0.9 (the results are symmetric for probabilities that are less than 0.5), $\beta_1 = 0.25, 0.5, 2.5$ for a range of variation in marginal probabilities as x_{ij} vary, $\gamma = 0.1$, 0.5, 0.9 for a range of exchangeable or AR(1) dependence and $d = 2, \ldots, 5$. If n is sufficiently large, the efficiencies will not depend much on the random x_{ij}. Some typical results are shown in Tables 1 and 2. Conclusions from the computations in the above design are the following.

(a) For exchangeable and AR(1) probit models with the corresponding form for $\mathbf{R}(\alpha)$ in $\mathbf{W}_i$, the best α-value, with maximum efficiency, increases as the latent correlation γ increases. The best α-value is in the range 0–0.1 for $\gamma = 0.1$, 0.2–0.3 for $\gamma = 0.5$ and 0.4–0.7 for $\gamma = 0.9$.

(b) Small latent correlation means that the IEE method ($\alpha = 0$) does about as well as optimal weight matrices. The value α should be larger for stronger dependence.
(c) The best choice α in a family that is AR(1) or exchangeable is such that $\mathbf{R}(\alpha)$ is close to the average correlation matrix (see approximation (7)) over the $\mathbf{y}_i$s.
(d) The use of optimal weight matrices in (5) is almost as good as maximum likelihood.
(e) The diagonal elements of $\mathbf{V}$ are roughly constant near their minimum values (interval of α of lengths 0.2 or more), but not necessarily the same interval for different regression coefficients, i.e. the efficiency function is quite flat near the 'optimal' α.
(f) A bad choice of α in $\mathbf{R}(\alpha)$ can lead to low efficiency. The worse choices for β_1 are small α near 0 for strong dependence, and large α near 1 for weak dependence. In particular, the choice of $\alpha = 0$ is bad if there is stronger dependence and the regression coefficient β_1 is small relative to the range of the x_{ij}s. This result is explained theoretically in the next section. The worse choice for β_0 is generally a large α such as $\alpha > 0.9$.

6. Analysis of when independent estimating equations do poorly

In this section we use the Fréchet upper bound to give an indication of when IEEs can be expected to do poorly relative to maximum likelihood and the proper use of GEEs improves efficiency for strongly dependent binary data.

Suppose that the range of the x_{ij} is a fixed interval such as $[-1, 1]$. As $\beta_1 \to 0$, the IEE method becomes inefficient for strong dependence. This follows from the case of $\beta_1 = 0$, with the Fréchet upper bound. In this case, since the covariate value has no effect, the marginal Bernoulli probability is $\Phi(\beta_0)$ for $j = 1, \ldots, d$. The Fréchet upper bound assigns total probability to the two vectors $(0, \ldots, 0)$ and $(1, \ldots, 1)$. Thus, we could observe $y_{i1} = \ldots = y_{id}$ for all i and conclude that $\beta_1 = 0$ from maximum likelihood estimation. But, with IEEs, the dependence information is ignored.

As an example to illustrate this point, consider the set-up in Section 5, with $d = 5$, $\gamma = 0.95$, $\beta_0 = 1.1$, $\beta_1 = 0.01$ and $n = 500$ observations with the x_{ij}s uniform in $[-1, 1]$. In one computation, the diagonals (scaled by n) of the inverse Fisher information matrix, $\mathbf{V}$ with $\mathbf{R}(\alpha = 0.8)$ and $\mathbf{V}$ with IEEs, are (1.8828, 0.4297), (2.0078, 0.4321) and (2.0078, 1.4994) respectively. Note that the $\mathbf{V}$-matrix with a good α-value is not far from the inverse Fisher information matrix, but the variance of the slope that is obtained by using IEEs is much larger and therefore less efficient.

7. Guidelines for selecting R(α) for binary data

The efficiency calculations for GEEs using binary data, compared with the inverse Fisher information matrix, show that the best choice $\mathbf{R}(\alpha)$ in an exchangeable or AR(1) family is one for which α is such that $\mathbf{R}(\alpha)$ approximates the average correlation matrix given in equation (7) of the binary random vectors $\mathbf{y}_i$. Because of the bounds on the correlation, the α for the best approximating matrix of form $\mathbf{R}(\alpha)$ cannot be large if $\boldsymbol{\beta}'\mathbf{x}_{ij}$ varies greatly. This is quite different from the case of a continuous response variable. Therefore, we suggest the following procedure for selecting $\mathbf{R}(\alpha)$ and α.

An initial data analysis with tabulations and odds ratios can be used to assess the strength of dependence. If the covariate vectors have just a few values, then an initial analysis consists of tabulating the frequencies of the d-dimensional binary vectors for each case of the covariate vector, and computing empirical odds ratios and correlations for each bivariate margin. The

tabulation can also be done for the combined case (ignoring covariate information). If all or most d-dimensional binary vectors occur, then the dependence is not strong; the strongest dependence is indicated if the frequencies concentrate near the vectors of all 0s and all 1s. Otherwise the odds ratios and correlations will suggest whether the dependence is weak or moderate, and whether an exchangeable or AR(1) structure is better for the weight matrix. If odds ratios are used, then a latent correlation of $\gamma = 0.3$ is about the same as an odds ratio of 2.5 in Plackett's copula (Joe (1997), pages 143–144, Table 5.2); similarly $\gamma = 0.5$ is about the same as odds ratios of 5, and $\gamma = 0.7$ corresponds to odds ratios greater than 12. Thus a value of the odds ratio between 1 and 3 is considered weak, between 3 and 10 is an indication of moderate dependence and a value exceeding 10 might be considered as strong dependence. If there are many possible values for the covariate vectors or if the covariates are continuous, then the covariate vectors can be categorized into a few cases to do the initial data analysis that was mentioned above.

Generally we can use an exchangeable matrix $\mathbf{R}(\alpha)$ for cluster-type samples, and an AR(1) matrix for longitudinal data; choose $\alpha \approx 0$ or use IEEs for weakly dependent binary data, and α in the range 0.2–0.3 for moderately dependent binary data, and α in the range 0.4–0.7 for strongly dependent binary data. The estimates of the regression parameters could be obtained easily by using existing GEE software, since they allow a user-specified correlation matrix. Alternatively, apply IEEs first, and then check the bounds in inequality (2) for each pair from $\{1, \ldots, d\}$ and decide on an appropriate value for α in the midrange of the bounds. We emphasize that α should be regarded simply as a parameter to determine the weights in equations (5).

With these simple rules, there is no need to estimate α for binary data on the basis of the procedures that are described in Liang and Zeger (1986), GEE1 in Prentice (1988) and related methods. Since for multivariate binary distributions the correlation between the binary variables is not a constant, the assumption of constancy that is used in most of these α estimation procedures is invalid.

8. An example

In this section we give an example to illustrate the methods that we proposed in choosing $\mathbf{R}(\alpha)$ and α. Consider a subset of the data from the six cities study, a longitudinal study on the health effects of air pollution, that has previously been analysed by Fitzmaurice and Laird (1993) and others. The data that are given in Table 1 of Fitzmaurice and Laird (1993) contain repeated binary measures on the wheezing status (yes, 1; no, 0) for each of 537 children from Steubenville, Ohio, at ages 7, 8, 9 and 10 years. The goal was to model the probability of wheezing status as a function of the child's age and a binary variable representing the mother's smoking habit during the first year of study.

An examination of the data reveals that all 16 binary quadruple vectors occur with a non-zero frequency and the dependence does not appear to be strong. For both levels of maternal smoking, the pairwise odds ratios for the various time points are in the range 5.5–11.4 with the strongest dependence for ages 8 and 9 years; the correlations are in the range 0.29–0.46. The pattern of dependence is closer to exchangeable than to AR(1). Since the overall dependence is moderate, the choice will be an exchangeable $\mathbf{R}(\alpha)$ with $\alpha = 0.3$.

We also model the marginal probability of wheeze status over time as a probit model using the covariates age recentred at 9, maternal smoking (0 or 1) and their interaction. Table 3 contains estimates, standard errors and p-values for the regression parameters by using GEEs with $\alpha = 0.3$ and the results of fitting a multivariate probit model with latent correlation $\gamma = 0.6$. The results are very similar. Furthermore, the conclusions are also similar to the results of fitting a

Table 3. Six cities study: regression analysis of the wheezing status using GEEs with exchangeable $\mathbf{R}(\alpha)$ and a multivariate probit model with exchangeable $\Omega(\gamma)$

Parameter	*Results from the GEE method* ($\alpha = 0.3$)			*Results from the multivariate probit model* ($\gamma = 0.6$)		
	Estimate	*Standard error*	*p-value*	*Estimate*	*Standard error*	*p-value*
Intercept	−1.126	0.063	0.0	−1.119	0.061	0.0
Age	−0.077	0.031	0.014	−0.078	0.030	0.010
Maternal	0.171	0.103	0.097	0.161	0.100	0.109
Age * Maternal	0.037	0.049	0.450	0.038	0.049	0.435

saturated multivariate probit model to each level of maternal smoking; there is a slight maternal smoking effect on the rate of wheezing and the age effect is a little clearer for the category of no maternal smoking.

9. Discussion

A routine use of the currently available GEE software for binary data could lead to incorrect analysis, because there is no check on the dependence of the correlation range as a function of the covariates. In this paper, we have done a more detailed analysis of GEEs for binary responses with the so-called working correlation matrix $\mathbf{R}(\alpha)$. Identifying $\mathbf{R}(\alpha)$ as the weight matrix, our efficiency analysis shows that GEEs with an appropriate α have good efficiency for binary responses when compared with a proper likelihood model. We gave simple rules for the choice of the weight matrix $\mathbf{R}(\alpha)$ and the parameter α.

A similar analysis can be done for count data, and, as with the binary case, ignoring the correlation bounds could give misleading results. With continuous data, the efficiency analysis for GEE and related methods are not in error like they are for binary data. An efficient and robust estimate of $\mathbf{R}(\alpha)$ in the GEE method for continuous responses is the (bias-corrected) quasi-least-squares estimate that was described in Chaganty and Shults (1999). See Chaganty (2003), for large sample and robust properties of the quasi-least-squares correlation parameter estimates.

Acknowledgements

This material is based on work that was supported in part by grants from the US Army Research Office and the Natural Sciences and Engineering Research Council. We are grateful to the Joint Editor, an Associate Editor and a referee for comments leading to an improved presentation. We thank Deepak Mav for assistance with some programming.

References

Ashford, J. R. and Sowden, R. R. (1970) Multivariate probit analysis. *Biometrics*, **26**, 535–546.

Chaganty, N. R. (1997a) An alternative approach to the analysis of longitudinal data via generalized estimating equations. *J. Statist. Planng Inf.*, **63**, 39–54.

Chaganty, N. R. (1997b) Loss in efficiency due to misspecification of the correlation structure in GEE. In *Proc. 51st Sess. International Statistical Institute, Istanbul*, vol. 2, pp. 127–128. Voorburg: International Statistical Institute.

Chaganty, N. R. (2003) Analysis of growth curves with patterned correlation matrices using quasi-least squares. *J. Statist. Planng Inf.*, **117**, 123–139.
Chaganty, N. R. and Shults, J. (1999) On eliminating the asymptotic bias in the quasi-least squares estimate of the correlation parameter. *J. Statist. Planng Inf.*, **76**, 145–161.
Crowder, M. (1995) On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika*, **82**, 407–410.
Diggle, P. J., Heagerty, P., Liang, K.-Y. and Zeger, S. L. (2002) *Analysis of Longitudinal Data*, 2nd edn. Oxford: Oxford University Press.
Fitzmaurice, G. M. and Laird, N. M. (1993) A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, **80**, 141–151.
Godambe, V. P. (ed.) (1991) *Estimating Functions*. Oxford: Oxford University Press.
Joe, H. (1997) *Multivariate Models and Dependence Concepts*. London: Chapman and Hall.
Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000) *Continuous Multivariate Distributions*, vol. 1, *Models and Applications*. New York: Wiley.
Lehmann, E. L. (1998) *Elements of Large Sample Theory*. New York: Springer.
Liang, K.-Y. and Zeger, S. L. (1986) Longitudinal data analysis using generalised linear models. *Biometrika*, **73**, 13–22.
Liang, K.-Y., Zeger, S. L. and Qaqish, B. (1992) Multivariate regression analyses for categorical data (with discussion). *J. R. Statist. Soc.* B, **54**, 3–40.
Lipsitz, S. R., Fitzmaurice, G. M., Sleeper, L. and Zhao, L. P. (1995) Estimation methods for the joint distribution of repeated binary observations. *Biometrics*, **51**, 562–570.
Maydeu-Olivares, A. (2001) Multidimensional item response theory modeling of binary data: large sample properties of NOHARM estimates. *J. Educ. Behav. Statist.*, **26**, 49–69.
McDonald, B. W. (1993) Estimating logistic regression parameters for bivariate binary data. *J. R. Statist. Soc.* B, **55**, 391–397.
Mendell, N. and Elston, R. (1974) Multifactorial qualitative traits: genetic analysis and prediction of recurrence risks. *Biometrics*, **30**, 41–57.
Muthén, B. (1978) Contributions to factor analysis of dichotomous variables. *Psychometrika*, **43**, 551–560.
Prentice, R. L. (1988) Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1033–1048.
Shults, J. and Chaganty, N. R. (1998) The analysis of serially correlated data using quasi-least squares. *Biometrics*, **54**, 1622–1630.
Sutradhar, B. C. and Das, K. (1999) On the efficiency of regression estimators in generalised linear models for longitudinal data. *Biometrika*, **86**, 459–465.
Sutradhar, B. C. and Das, K. (2000) On the accuracy of efficiency of estimating equation approach. *Biometrics*, **56**, 622–625.
Szudek, J., Joe, H. and Friedman, J. M. (2002) Analysis of intra-familial phenotypic variation in neurofibromatosis 1 (Nf1). *Genet. Epidem.*, **23**, 150–164.