In this paper we develop a Gaussian estimation procedure involving a working correlation matrix for the estimation of the regression parameters in longitudinal binary response data. A Newton-Raphson algorithm is derived for estimating the regression parameters from the Gaussian likelihood estimating equations for known correlation parameters. The correlation parameters are estimated by the method of moments. A two-step iterative procedure is suggested for the joint estimation of the regression parameters and the correlation parameters. Asymptotic variances of the Gaussian estimates of the regression parameters are obtained. Some new results regarding the consistency property of the estimates of the regression parameters and the correlation parameters of the working correlation matrices are given. A simulation study is conducted to compare efficiency properties of twelve estimators of the regression parameters. Simulations show that the Gaussian estimates of the regression parameters using the unstructured correlation matrix of the responses for a subject are, in general, more efficient than those by the other eleven methods. The next best seems to be the Gaussian estimates using the general autocorrelation structure. A data set is analyzed and a discussion is given.

Key Words: Gaussian estimation; Generalized estimating equations; Longitudinal binary data.

1 Introduction

Correlated binary response data arise in many longitudinal studies in which the main purpose is to study the effects of the covariates on the correlated binary responses. For example, in the Six Cities study of the health effects of air pollution, analyzed by Fitzmaurice and Laird (1993), one of the purposes is to determine whether the maternal smoking significantly affects the wheezing status of children.

One method of analyzing binary longitudinal response data is by the method of generalized estimating equations (GEE) proposed by Liang and Zeger (1986) in which a working correlation matrix for the responses for each individual is used (see, for example, Prentice, 1988 and Fitzmaurice, Laird and Rotnitzky, 1993). However, there is a history of controversy over choosing the working correlation structure \( R(\rho) \) in GEE. For example, Crowder (1995) finds...
that in some cases the parameters involved in the working correlation matrix are subject to an uncertainty of definition which can lead to a breakdown of asymptotic properties of the estimators (see also Crowder, 2001). Further, the miss-specification of the correlation structure can result in loss of efficiency of the regression parameters (Wang and Carey, 2003).

Likelihood based methods are also available. For example, Lipsitz, Fitzmaurice, Sleeper and Zhao (1995) use a likelihood for the binary responses based on the Bahadur representation and Fitzmaurice and Laird (1993) use an exponential likelihood based on odds ratios. The likelihood approaches are rather complicated except in some special cases, such as, the analysis of paired binary data (Prentice, 1988).

Stefanescu and Turnbull (2005) use the likelihood approach based on a multivariate probit (MP) model for the analysis of longitudinal binary response data. Chaganty and Joe (2004) show that the GEE method with the working correlation matrix $R(\rho)$ has good efficiency relative to the likelihood approach using a MP model. However, they recommend that $R(\rho)$ should be a weight matrix rather than a correlation matrix of binary responses and they suggest a method of choosing this weight matrix.

Whittle (1961) introduces the Gaussian estimation procedure which uses the normal log-likelihood, without assuming that the data are normally distributed. The purpose of this paper is to develop and investigate Gaussian estimation procedure for the estimation of regression parameters in correlated (longitudinal) binary response data and compare this method with the GEE method and the weighted GEE method of Chaganty and Joe (2004). The motivation of this comes from the good properties of the Gaussian estimation procedure in other applications. For example, Crowder (1985) shows by simulation that Gaussian estimate of the correlation parameter of equi-correlated clustered binary data has high efficiency. Paul and Islam (1998), again, by simulation, show that Gaussian estimator of the overdispersion parameter in clustered binomial data has best efficiency in comparison to likelihood, quasi-likelihood and extended quasi-likelihood estimates. Wang and Zhao (2007) use Gaussian estimation for the analysis of longitudinal data when the covariance function is modelled by additional parameters to the mean parameters (see Wang and Zhao, 2007 for more details). See also Hand and Crowder (1996) for more applications.

As in the GEE we use a working correlation matrix for the responses of each individual. A Newton-Raphson algorithm is derived for estimating the regression parameters from the Gaussian likelihood estimating equations for known correlation parameters. The correlation parameters are estimated by the method of moments. A two-step iterative procedure is suggested for the
joint estimation of the regression parameters and the correlation parameters. Asymptotic variances of the Gaussian estimates of the regression parameters are obtained. Some new results regarding the consistency property of the estimates of the regression parameters and the correlation parameters of the working correlation matrices are given. Efficiency properties of twelve estimators of the regression parameters, namely, the maximum likelihood estimates using a multivariate probit (MP) model, four versions of the Gaussian estimates, five versions of the generalized estimating equations (GEE) and two versions from a recent weighted GEE by Chaganty and Joe (2004), are compared by simulation. Efficiency results are obtained for all the methods using four different data sets generated from the MP model with latent correlation structures (i) exchangeable, (ii) AR(1), (iii) general autocorrelation and (iv) unstructured.

The Gaussian estimation procedure is developed and some theoretical results are obtained in Section 2. A simulation study is conducted in Section 3. A data set is analyzed in Section 4 and a discussion follows in Section 5.

2 Gaussian Estimation of the Regression Parameters

2.1 Estimation of the regression parameters

Let \( y_i = (y_{i1}, \ldots, y_{id})^T \) be the vector of binary responses with a \( d \times p \) design matrix \( X_i = (x_{i1}, \ldots, x_{id})^T \) for the \( i \)th subject, \( i = 1, \ldots, K \). Assume that the \( K \) subjects are independent while the repeated measurements \( y_{ij} \) taken on each subject are correlated. Define \( \mu_i = E(y_i|X_i) = (\mu_{i1}, \ldots, \mu_{id})^T \) to be the expectation of \( y_i \) conditional on \( X_i \) and suppose \( \mu_i = F(X_i\beta) \), where \( \beta \) is a \( p \times 1 \) vector of regression parameters of interest and \( F^{-1} \) is the link function. For the binary response data we consider the logit and probit link functions. The variance of \( y_{ij} \) is given by \( \mu_{ij}(1 - \mu_{ij}) \).

Let \( R(\rho) \) be a working correlation matrix completely specified by the parameter vector \( \rho \) of length \( q \) and \( W_i = A_i^{1/2}R(\rho)A_i^{1/2} \) be the corresponding working covariance matrix, where \( A_i(\beta) = \text{diag}\{\mu_{ij}(1 - \mu_{ij})\}, j = 1, \ldots, d, i = 1, \ldots, K \). Further, let \( I_d \) be an identity matrix of dimension \( d \), \( \Delta_i = \text{diag}(f(x_{i1}^T\beta), \ldots, f(x_{id}^T\beta)) \) with \( f = F' \).

Then, the Gaussian log-likelihood is given by

\[
l = \sum_{i=1}^{K} l_i = \sum_{i=1}^{K} \{-\frac{1}{2} \log |W_i| - \frac{1}{2} Q_i\},
\]

where \( Q_i = (y_i - \mu_i)^T W_i^{-1}(y_i - \mu_i) \) and for given values of \( \rho \), the maximum Gaussian likelihood estimates of the regression parameters are obtained by
solving the system of $p$ estimating equations

$$-2 \frac{\partial l}{\partial \beta} = \sum_{i=1}^{K} \left( \frac{\partial \log |W_i|}{\partial \mu_i} + \frac{\partial Q_i}{\partial \mu_i} \right) \Delta_i X_i = 0,$$  

(2)

with \( \frac{\partial \log |W_i|}{\partial \mu_i} = \text{vec}^T(W_i^{-1}) \frac{\partial W_i}{\partial \mu_i} \), \( \frac{\partial Q_i}{\partial \mu_i} = [y_i^T \otimes y_i^T - 2(\mu_i^T \otimes \mu_i^T) + \mu_i^T \otimes \mu_i^T] \)

\( \frac{\partial W_i^{-1}}{\partial \mu_i} + 2(\mu_i^T - y_i^T)W_i^{-1}, \frac{\partial W_i}{\partial \mu_i} = \left[A_i^{1/2} R(\rho) \otimes I_d + I_d \otimes A_i^{1/2} R(\rho) \right] \frac{\partial A_i^{1/2}}{\partial \mu_i} \) and \( \frac{\partial W_i^{-1}}{\partial \mu_i} = \left[A_i^{-1/2} R^{-1}(\rho) \otimes I_d + I_d \otimes A_i^{-1/2} R^{-1}(\rho) \right] \frac{\partial A_i^{-1/2}}{\partial \mu_i} \). The derivative matrices \( \frac{\partial A_i^{1/2}}{\partial \mu_i} \) and \( \frac{\partial A_i^{-1/2}}{\partial \mu_i} \) are \( d^2 \times d \) sparse matrices with non-zero quantities \( \frac{1}{2}[\mu_{ij}(1-\mu_{ij})]^{-1/2}(1-2\mu_{ij}) \) and \( \frac{1}{2}[\mu_{ij}(1-\mu_{ij})]^{-3/2}(1-2\mu_{ij}) \) respectively in the \([j-1]d + j, j\)th term, \( j = 1, \ldots, d \).

To solve equations (2), we use the Newton-Raphson method for which we need \( \frac{\partial \beta_j}{\partial \beta} = \sum_{i=1}^{K} \frac{\partial \beta_j}{\partial \beta} \). The derivation can be obtained from the authors.

Based on the Newton-Raphson Method, the Gaussian estimates are updated according to

$$\hat{\beta}_{j+1} = \hat{\beta}_{j} + \left[ \frac{\partial^2 l}{\partial \beta^2} \right]^{-1} \frac{\partial l}{\partial \beta} \bigg|_{\hat{\beta}_{j}}, \quad j = 1, 2, \ldots$$  

(3)

Note that the Newton-Raphson procedure given above for estimating the regression parameter \( \beta \) is based on the assumption that the correlation parameters involved in the covariance matrix \( W_i = A_i^{1/2}(\beta)R(\rho)A_i^{1/2}(\beta) \) are known. In what follows, we consider the general autocorrelation and the unstructured correlation structure. These two correlation structures are robust in the sense that they include the exchangeable and AR(1) structures (Sutradhar, 2003). The general autocorrelation structure is

$$R(\rho_1, \ldots, \rho_{d-1}) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{d-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{d-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \rho_{d-1} & \rho_{d-2} & \rho_{d-3} & \cdots & 1 \end{bmatrix}.$$  

and the unstructured correlation matrix (Liang and Zeger, 1986) is

$$R = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1,d-1} \\ \rho_{12} & 1 & \rho_{23} & \cdots & \rho_{2,d-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \rho_{1,d-1} & \rho_{2,d-2} & \rho_{3,d-3} & \cdots & 1 \end{bmatrix}. \quad (4)$$

Submitted to Stat 2011 Canada/IMST 2011-FIM XX on April 29, 2011
Let \( y_{ij}^* = (y_{ij} - \hat{\mu}_{ij})/\sqrt{\hat{\mu}_{ij}(1-\hat{\mu}_{ij})} \). Then, the method of moments estimate of the correlation parameter \( \rho_l \) in \( R(\rho_1, \ldots, \rho_{d-1}) \) is

\[
\hat{\rho}_l = \frac{\sum_{i=1}^{K} \sum_{j=1}^{d-l} y_{ij}^* y_{i,j+l}/(d-l)}{\sum_{i=1}^{K} \sum_{j=1}^{d} y_{ij}^* y_{ij}^2/d}, \quad l = 1, \ldots, d - 1,
\]

and the estimate of the unstructured correlation matrix is given by

\[
\hat{R} = \sum_{i=1}^{K} \hat{A}_i^{-1/2} S_i S_i^T \hat{A}_i^{-1/2} / K, \quad \text{where } S_i = y_i - \hat{\mu}_i, i = 1, \ldots, K.
\]

The Newton-Raphson iterative procedure for estimating the regression parameters and the method of moments estimates of the correlation parameters are combined in a two-step iterative procedure which is described in what follows.

Step 1: For given initial values \( \beta^0 \) of \( \beta \) and \( \rho^0 \), where \( \rho \) is the vector of correlation parameters (depending on the structure of the working correlation matrix chosen), estimate \( \beta \) via the formula (3). Denote this by \( \beta^1 \).

Step 2: Obtain the elements of \( \rho \) by the method of moments described above using \( \beta^1 \). Denote this estimate of \( \rho \) by \( \rho^1 \).

Iterate between step 1 and step 2 until convergence.

### 2.2 Variance of \( \hat{\beta} \)

In the Appendix we have shown that the estimating equations (2) are asymptotically, as \( K \to \infty \), unbiased. So, by the general theory of unbiased estimating functions (Crowder, 1986 and Liang and Zeger, 1995), the estimator \( \hat{\beta} \) by (2) is consistent and has asymptotic multivariate normal distribution \( MVN(\beta, V_\beta) \), where \( V_\beta \) is given by

\[
V_\beta = D^{-1} V(D^{-1})^T, \tag{5}
\]

where \( D = E \left( \frac{\partial^2 l}{\partial \beta^2} \right) \) and \( V = \text{cov} \left( \frac{\partial l}{\partial \beta} \right) \). It can be shown that

\[
D = 1 - \sum_{i=1}^{K} \left\{ \left( \frac{\partial W_i}{\partial \beta} \right)^T \frac{\partial W_i^{-1}}{\partial \beta} + \left[ I_p \otimes \text{vec}^T(W_i^{-1}) \right] \frac{\partial^2 W_i}{\partial \beta^2} \right. \\
+ E \left( \frac{\partial}{\partial \mu_i} \left[ \frac{\partial Q_i}{\partial \mu_i} \Delta_i X_i \right] \Delta_i X_i \right) \}
\]
and

\[ V = \frac{1}{4} \sum_{i=1}^{K} [A \text{cov}(y_i \otimes y_i)AT + B \Sigma_i B^T + C + C^T], \]

where \( A = X_i^T \Delta_i \left( \frac{\partial W^{-1}}{\partial \mu} \right)^T \), \( B = -2X_i^T \Delta_i \left( \frac{\partial W^{-1}}{\partial \mu} \right)^T (\mu_i \otimes I_d) + W_i^{-1} \) and \( C = \text{cov} (A(y_i \otimes y_i), By_i) \). Note that the dimensions of the matrices \( A, B \) and \( C \) are \( p \times d^2, p \times d \) and \( p \times p \) respectively and that of \( \text{cov}(y_i \otimes y_i) \) is \( d^2 \times d^2 \).

To evaluate \( D \) and \( V \), we need to find \( E \left( \frac{\partial}{\partial \mu_i} \left[ \frac{\partial Q}{\partial \mu_i} \Delta_i X_i \Delta_i X_i \right] \right), \text{cov}(y_i \otimes y_i) \) and \( C \). These require extensive calculations which can be obtained from the authors. In (5), the true covariance matrix \( \Sigma_i \) is estimated by \( \hat{\Sigma}_i = \hat{A}_i^{1/2} R \hat{A}_i^{1/2} \). The variance \( V_\beta \) of \( \hat{\beta} \) is estimated by replacing \( \beta \) and \( \Sigma \) with their estimates \( \hat{\beta} \) and \( \hat{\Sigma}_i \) respectively.

### 2.3 Consistency of the Estimates of the parameters

We showed in the Appendix that if the estimate \( R(\hat{\rho}) \) of the working correlation \( R(\rho) \) converges to the true correlation matrix \( C(\rho) \) in probability, then the estimating equations (2) are asymptotically, as \( K \to \infty \), unbiased and therefore the estimator \( \hat{\beta} \) obtained by solving the system of equations given by (2) is consistent. It then remains to show that \( R(\hat{\rho}) \) is consistent.

The moment estimates of the correlation parameters of the unstructured correlation matrix are consistent whatever is the true correlation structure: unstructured, general autocorrelation, AR(1) or exchangeable. The proof is omitted here which can be obtained from the authors. However, the reverse is not true. For example, the moment estimate of the correlation parameter \( \rho \) of the exchangeable correlation structure is not consistent when the true correlation structure is any of the other three. Similarly, the moment estimates of the correlation parameters of the general autocorrelation structure are consistent when the true correlation structure is general autocorrelation, AR(1) or exchangeable. In this sense the unstructured correlation matrix is most robust against misspecification by other correlation structures. The next robust, of course, is the general autocorrelation structure.

Now, there is some circularity in the proofs, in that consistency of \( \hat{\beta} \) requires consistency of \( \hat{\rho} \) and vice versa. However, this problem can be overcome by using consistent estimate of, for example, \( \beta \) at the initial stage of the iterative procedure. That is, overall consistency of \( \hat{\beta} \) and \( \hat{\rho} \) are obtained if consistent initial estimates of \( \beta \), such as the GEEs, are used at step 1 of the two step iterative procedure described at the end of section 2.1.
3 Simulations

In this section we compare, by simulations, twelve estimators of the regression parameters, namely, the maximum likelihood estimates using a MP model, four versions of the Gaussian estimates, five versions of the GEE and two versions of the weighted GEE.

Following Chaganty and Joe (2004), we use the multivariate probit (MP) model as a data generation mechanism. The MP model is a commonly used model for multivariate binary data. It assumes that the binary response is the indicator of the event that an unobserved latent variable exceeds a given threshold. Let \( y_i = (y_{i1}, \ldots, y_{id})^T \) be the \( d \)-dimensional vector of binary responses on the \( i \)th subject, \( i = 1, \ldots, K \). Let \( x_i = (x_{i1}, \ldots, x_{id})^T \) be a \( d \times p \) covariate matrix. Let \( Z_i = (z_{i1}, \ldots, z_{id})^T \) be a \( d \)-dimensional vector of latent variables such that \( Z_i = x_i \beta + \epsilon_i, i = 1, \ldots, K \). The latent variable \( Z_i \) is assumed to follow a multivariate normal distribution with mean \( x_i \beta \) and covariance \( \Omega(\gamma) \), where \( \gamma \) is the latent correlation. The relationship between \( z_{ij} \) and \( y_{ij} \) in the MP model is given by

\[
y_{ij} = \begin{cases} 1, & \text{if } z_{ij} > 0; \\ 0, & \text{otherwise.} \end{cases} \quad j = 1, \ldots, d.
\]

So that \( P(y_{ij} = 1|X_i) = P(z_{ij} > 0) = \Phi(\beta' x_{ij}) \), where \( \Phi \) is the standard normal distribution function. It can be seen that the correlation between any two binary responses \( y_{ij} \) and \( y_{ik} \) is given by

\[
\text{corr}(y_{ij}, y_{ik}) = \frac{\Phi_2(v_j, v_k; \gamma) - \Phi(v_j)\Phi(v_k)}{[\Phi(v_j)\{1 - \Phi(v_j)\}\Phi(v_k)\{1 - \Phi(v_k)\}]^{1/2}},
\]

where \( \Phi_2(\omega_1, \omega_2; \gamma) \) is the bivariate normal distribution function with correlation \( \gamma \), \( v_j = \beta' x_{ij} \) and \( v_k = \beta' x_{ik} \).

For simulating data from the MP model, we use the latent covariance matrix \( \Omega(\gamma) \). For example, for generating binary data with exchangeable \( R(\rho) \), we use the exchangeable correlation matrix \( \Omega(\gamma) \). Note that the correlation \( \rho \) of the binary variables is always less than the latent correlation \( \gamma \) as shown in Chaganty and Joe (2004). Efficiencies of the estimates of the regression parameters are compared for all the methods using four different data sets generated from the MP model with latent correlation structures (i) exchangeable, (ii) AR(1), (iii) general autocorrelation and (iv) unstructured.

For exchangeable or AR(1) model, we choose \( \gamma = 0.5 \). For general auto-
correlation structure we use $A$ for $\Omega(\gamma)$, where
\[
A = \begin{pmatrix}
1.0 & 0.5 & 0.4 & 0.3 & 0.2 \\
0.5 & 1.0 & 0.5 & 0.4 & 0.3 \\
0.4 & 0.5 & 1.0 & 0.5 & 0.4 \\
0.3 & 0.4 & 0.5 & 1.0 & 0.5 \\
0.2 & 0.3 & 0.4 & 0.5 & 1.0
\end{pmatrix}.
\]
For unstructured correlation, we use the following positive definite correlation matrix
\[
U = \begin{pmatrix}
1.00 & 0.12 & 0.52 & 0.06 & 0.38 \\
0.12 & 1.00 & 0.63 & 0.16 & 0.78 \\
0.52 & 0.63 & 1.00 & 0.10 & 0.90 \\
0.06 & 0.16 & 0.10 & 1.00 & 0.15 \\
0.38 & 0.78 & 0.90 & 0.15 & 1.00
\end{pmatrix}.
\]
For all correlation structures we choose probit link, $d = 5$, $p = 2$, $x_{ij} = (1, x_{ij})^T$, where $x_{ij}$ are taken as uniform random variables in the interval $[-1, 1]$, $\beta = (0.0, 0.5)$ and $N = 50, 80, 150$.

For each $N$, we simulated 500 samples and obtained the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ for each sample. We then calculated $N \times \text{average estimated variance } (\sum_{i=1}^{500} \text{var}(\hat{\theta}_i))/500$, where $\hat{\theta}_i$ is either $\hat{\beta}_0$ or $\hat{\beta}_1$ for the $i$th sample.

We first compare the Gaussian estimation procedures with the four correlation structures discussed earlier, namely the exchangeable, the AR(1), the general autocorrelation and the unstructured correlation. The results are given in Table 1.

Table 1. $N \times$ average estimated variance for $\hat{\beta}_0$ and $\hat{\beta}_1$ by Gaussian estimation procedure using the four working correlation structures: data generated from MP model with latent (i) exchangeable $R(0.5)$; (ii) AR(1) $R(0.5)$; (iii) general autocorrelation matrix $A$ and (iv) unstructured covariance matrix $U$; $x_{ij} \sim \text{uniform(-1,1)}$; $p = 2$, $\beta_0 = 0.0$, $\beta_1 = 0.5$; observation times $d = 5$; based on 500 iterations.

<table>
<thead>
<tr>
<th>$N$</th>
<th>Method</th>
<th>$\text{(i) } N \times \text{var}(\beta_0)$</th>
<th>$\text{(ii) } N \times \text{var}(\beta_0)$</th>
<th>$\text{(iii) } N \times \text{var}(\beta_0)$</th>
<th>$\text{(iv) } N \times \text{var}(\beta_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>Gaussian-Exch</td>
<td>(0.783, 0.701) (0.621, 0.850) (0.705, 0.750) (0.831, 0.839)</td>
<td>(0.419, 0.853) (0.476, 0.831) (0.512, 0.765) (0.383, 0.1030)</td>
<td>(0.770, 0.661) (0.587, 0.722) (0.681, 0.710) (0.810, 0.724)</td>
<td>(0.721, 0.591) (0.552, 0.656) (0.649, 0.636) (0.602, 0.473)</td>
</tr>
<tr>
<td>80</td>
<td>Gaussian-AR(1)</td>
<td>(0.775, 0.700) (0.616, 0.855) (0.691, 0.793) (0.823, 0.845)</td>
<td>(0.408, 0.866) (0.466, 0.835) (0.499, 0.774) (0.371, 1.052)</td>
<td>(0.768, 0.675) (0.584, 0.736) (0.670, 0.730) (0.804, 0.745)</td>
<td>(0.740, 0.633) (0.561, 0.689) (0.643, 0.681) (0.610, 0.493)</td>
</tr>
<tr>
<td>150</td>
<td>Gaussian-Exch</td>
<td>(0.766, 0.710) (0.668, 0.857) (0.681, 0.793) (0.809, 0.855)</td>
<td>(0.396, 0.883) (0.455, 0.845) (0.491, 0.774) (0.359, 1.070)</td>
<td>(0.762, 0.696) (0.581, 0.754) (0.664, 0.744) (0.793, 0.759)</td>
<td>(0.747, 0.672) (0.570, 0.729) (0.651, 0.718) (0.617, 0.518)</td>
</tr>
</tbody>
</table>
We see from the results in Table 1 that the Gaussian estimation procedure using AR(1) correlation structure produces smallest variance estimates for $\beta_0$ and largest variance estimates for $\beta_1$ irrespective of the data generation procedure. Among the three other methods, in general, Gaussian estimation using exchangeable correlation structure produces largest variance estimates for both $\beta_0$ and $\beta_1$. The other two estimation procedures, in general, produce smallest estimated variance, although the Gaussian estimation procedure using the unstructured correlation produces smallest variance estimates among these three methods.

We now compare the two Gaussian estimation procedures using the general autocorrelation structure (Gaussian-Autocorr) and the unstructured correlation matrix (Gaussian-Unstr) with the maximum likelihood (ML) estimates based on the MP model and GEE approaches. We consider GEE-independence (GEE-I), GEE-AR(1), GEE-exchangeable (GEE-ex), GEE-general autocorrelation (GEE-Autocorr), GEE-unstructured (GEE-un) and weighted GEE-exchangeable by Chaganty and Joe (GEE-CJ).

For all data sets, ML estimates were obtained using the exchangeable correlation structure. Results are similar for AR(1) correlation structure. For the estimation using Chaganty and Joe’s method, we use the exchangeable correlation structure and the AR(1) correlation structure, both with $\rho = 0.3$ (following their guidelines for choosing $\rho$). Thus, we use two versions of GEE-CJ, henceforth named as GEE-CJ(EX) and GEE-CJ(AR(1)). Note that data were generated using latent correlation $\gamma = 0.5$. According to the recommendations of Chaganty and Joe (2004), the value of $\rho$ to be taken for the estimation of the regression parameters should be less than 0.5. We examined efficiency results of the above two methods using other values of $\rho$ which satisfies this requirement, such as, $\rho = 0.2$ and the results are found to be similar. Results of $N \times$ average estimated variance for $\hat{\beta}_0$ and $\hat{\beta}_1$ are given in Table 2.

Results in Table 2 show that the performance of GEE-I is the worst, at least in terms of $\hat{\beta}_1$, producing the largest variance irrespective of the data generation mechanism. Again, irrespective of the data generation mechanism, in terms of $\hat{\beta}_1$, Gaussian-Unstr performs the best, producing the smallest variance and the next best is Gaussian-Autocorr. Only when the data are generated using unstructured covariance matrix GEE-un has slight edge over Gaussian-Autocorr. Estimate of the variance of $\hat{\beta}_0$ does not seem to differ much irrespective of the data generation mechanism and the method of estimation, although, Gaussian-Autocorr seems to produce larger variance estimate when data are generated using unstructured covariance matrix.

The simulation study was extended to compare bias and MSE. Again,
Table 2. $N \times$ average estimated variance for $\hat{\beta}_0$ and $\hat{\beta}_1$ by ML, Gaussian-Autocorr, Gaussian-Unstr and GEE methods: data generated from MP model with latent (i) exchangeable $R(0.5)$; (ii) AR(1) $R(0.5)$; (iii) general autocorrelation matrix $A$ and (iv) unstructured covariance matrix $U$; $x_{ij} \sim \text{uniform}(-1,1)$; $p = 2$, $\beta_0 = 0$, $\beta_1 = 0.5$; observation times $d = 5$; based on 500 iterations.

<table>
<thead>
<tr>
<th>N</th>
<th>Method</th>
<th>(i) $N \times \text{var}(\hat{\beta}_0)$</th>
<th>(ii) $N \times \text{var}(\hat{\beta}_1)$</th>
<th>(iii) $N \times \text{var}(\hat{\beta}_0)$</th>
<th>(iv) $N \times \text{var}(\hat{\beta}_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>ML</td>
<td>(0.774, 0.650)</td>
<td>(0.581, 0.720)</td>
<td>(0.677, 0.714)</td>
<td>(0.811, 0.718)</td>
</tr>
<tr>
<td></td>
<td>Gaussian-Autocorr</td>
<td>(0.725, 0.586)</td>
<td>(0.554, 0.637)</td>
<td>(0.638, 0.643)</td>
<td>(0.616, 0.459)</td>
</tr>
<tr>
<td></td>
<td>Gaussian-Unstr</td>
<td>(0.740, 1.032)</td>
<td>(0.572, 1.013)</td>
<td>(0.654, 0.999)</td>
<td>(0.644, 1.023)</td>
</tr>
<tr>
<td></td>
<td>GEE-I</td>
<td>(0.749, 0.901)</td>
<td>(0.563, 0.869)</td>
<td>(0.651, 0.870)</td>
<td>(0.649, 0.986)</td>
</tr>
<tr>
<td></td>
<td>GEE-ex</td>
<td>(0.742, 0.830)</td>
<td>(0.574, 0.923)</td>
<td>(0.657, 0.860)</td>
<td>(0.646, 0.887)</td>
</tr>
<tr>
<td></td>
<td>GEE-Autocorr</td>
<td>(0.731, 0.801)</td>
<td>(0.558, 0.841)</td>
<td>(0.645, 0.821)</td>
<td>(0.644, 0.839)</td>
</tr>
<tr>
<td></td>
<td>GEE-un</td>
<td>(0.713, 0.765)</td>
<td>(0.540, 0.799)</td>
<td>(0.628, 0.779)</td>
<td>(0.553, 0.686)</td>
</tr>
<tr>
<td></td>
<td>GEE-CJ(EX)</td>
<td>(0.741, 0.832)</td>
<td>(0.575, 0.936)</td>
<td>(0.652, 0.884)</td>
<td>(0.646, 0.893)</td>
</tr>
<tr>
<td></td>
<td>GEE-CJ(AR(1))</td>
<td>(0.748, 0.905)</td>
<td>(0.564, 0.874)</td>
<td>(0.646, 0.885)</td>
<td>(0.661, 1.043)</td>
</tr>
<tr>
<td>80</td>
<td>ML</td>
<td>(0.722, 0.825)</td>
<td>(0.547, 0.856)</td>
<td>(0.570, 0.848)</td>
<td>(0.632, 0.875)</td>
</tr>
<tr>
<td></td>
<td>Gaussian-Autocorr</td>
<td>(0.761, 0.685)</td>
<td>(0.586, 0.734)</td>
<td>(0.666, 0.734)</td>
<td>(0.803, 0.733)</td>
</tr>
<tr>
<td></td>
<td>Gaussian-Unstr</td>
<td>(0.740, 0.634)</td>
<td>(0.561, 0.696)</td>
<td>(0.643, 0.688)</td>
<td>(0.613, 0.497)</td>
</tr>
<tr>
<td></td>
<td>GEE-I</td>
<td>(0.731, 0.335)</td>
<td>(0.571, 1.003)</td>
<td>(0.649, 1.014)</td>
<td>(0.648, 1.026)</td>
</tr>
<tr>
<td></td>
<td>GEE-ex</td>
<td>(0.733, 0.908)</td>
<td>(0.564, 0.862)</td>
<td>(0.645, 0.879)</td>
<td>(0.653, 0.995)</td>
</tr>
<tr>
<td></td>
<td>GEE-Autocorr</td>
<td>(0.733, 0.835)</td>
<td>(0.575, 0.913)</td>
<td>(0.650, 0.878)</td>
<td>(0.649, 0.893)</td>
</tr>
<tr>
<td></td>
<td>GEE-un</td>
<td>(0.735, 0.808)</td>
<td>(0.560, 0.849)</td>
<td>(0.640, 0.841)</td>
<td>(0.644, 0.845)</td>
</tr>
<tr>
<td></td>
<td>GEE-CJ(EX)</td>
<td>(0.733, 0.836)</td>
<td>(0.575, 0.923)</td>
<td>(0.653, 0.882)</td>
<td>(0.650, 0.898)</td>
</tr>
<tr>
<td></td>
<td>GEE-CJ(AR(1))</td>
<td>(0.738, 0.910)</td>
<td>(0.565, 0.864)</td>
<td>(0.646, 0.881)</td>
<td>(0.664, 1.055)</td>
</tr>
<tr>
<td>150</td>
<td>ML</td>
<td>(0.730, 0.826)</td>
<td>(0.542, 0.806)</td>
<td>(0.565, 0.838)</td>
<td>(0.632, 0.850)</td>
</tr>
<tr>
<td></td>
<td>Gaussian-Autocorr</td>
<td>(0.762, 0.699)</td>
<td>(0.581, 0.757)</td>
<td>(0.667, 0.743)</td>
<td>(0.794, 0.768)</td>
</tr>
<tr>
<td></td>
<td>Gaussian-Unstr</td>
<td>(0.752, 0.669)</td>
<td>(0.570, 0.726)</td>
<td>(0.655, 0.718)</td>
<td>(0.615, 0.530)</td>
</tr>
<tr>
<td></td>
<td>GEE-I</td>
<td>(0.738, 0.107)</td>
<td>(0.574, 1.005)</td>
<td>(0.652, 1.011)</td>
<td>(0.656, 1.032)</td>
</tr>
<tr>
<td></td>
<td>GEE-ex</td>
<td>(0.746, 0.888)</td>
<td>(0.564, 0.865)</td>
<td>(0.648, 0.883)</td>
<td>(0.661, 1.001)</td>
</tr>
<tr>
<td></td>
<td>GEE-Autocorr</td>
<td>(0.738, 0.814)</td>
<td>(0.574, 0.914)</td>
<td>(0.652, 0.874)</td>
<td>(0.656, 0.893)</td>
</tr>
<tr>
<td></td>
<td>GEE-un</td>
<td>(0.737, 0.819)</td>
<td>(0.564, 0.857)</td>
<td>(0.645, 0.846)</td>
<td>(0.646, 0.856)</td>
</tr>
<tr>
<td></td>
<td>GEE-CJ(EX)</td>
<td>(0.738, 0.815)</td>
<td>(0.574, 0.924)</td>
<td>(0.654, 0.871)</td>
<td>(0.656, 0.897)</td>
</tr>
<tr>
<td></td>
<td>GEE-CJ(AR(1))</td>
<td>(0.745, 0.888)</td>
<td>(0.565, 0.866)</td>
<td>(0.649, 0.879)</td>
<td>(0.671, 1.053)</td>
</tr>
</tbody>
</table>

Based on 500 simulated samples, we obtained (a) average bias($\hat{\theta}_i$) = $\sum_{i=1}^{500}(\hat{\theta}_i - \theta_i)/500$ and (b) $N \times$ average MSE ($\sum_{i=1}^{500}(\hat{\theta}_i - \theta_i)^2/500$). To save space we only summarize the results (not given here) of the simulation which can be obtained from the authors.

Our simulations show that biases of the estimates by all procedures compared are small. In terms of the MSE, the performance of GEE-I is the worst in general. When data are simulated using unstructured correlation structure, Gaussian-Unstr is the best for the estimation of $\beta_1$, agreeing with the results shown in terms of estimated variances. For the estimation of $\beta_0$, no method seems to perform better than any other. For data with other correlation structures there does not seem to be any significant differences in efficiency both for the estimation of $\beta_0$ and $\beta_1$. 

Submitted to Stat 2011 Canada/IMST 2011-FIM XX on April 29, 2011 10
We also conducted a further simulation study to compare these twelve estimators by generating correlated binary data with specified marginal means and correlations (Qaqish, 2003). Simulation results not reported in this paper show similar conclusions.

4 An Example

We consider a subset of data from the Six Cities study, a longitudinal study of the health effects of air pollution that was analyzed by Fitzmaurice and Laird (1993). The data set contains complete records on 537 children from Steubenville, Ohio, each of whom was examined annually at ages 7 through 10. The repeated binary response is the wheezing status (1=yes, 0=no) of a child at each occasion. The purpose of the study is to model the probability of the wheezing status as a function of the child’s age, his/her mother’s maternal smoking habit (a binary variable MS with 1 if the mother smoked regularly and 0 otherwise) and their interactions. We consider the same marginal model used by Fitzmaurice and Laird (1993) with a probit link

$$\text{probit}(\mu) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{MS} + \beta_3 \text{Age}^*\text{MS},$$

where ‘age’ is the age in years since the child’s 9th birthday.

Estimates of the regression parameters of model (6) and their standard errors by all the methods discussed are given in Table 3. Estimates of the correlation parameters by all the methods are given in Table 4.

Table 3. Results of the regression analysis of the wheezing status data; estimates of $\beta_0$, $\beta_1$, $\beta_2$ and $\beta_3$ of the model (6) with standard errors in parenthesis using maximum likelihood method based on the MP model, four Gaussian estimation methods and six GEE procedures; with probit link.

<table>
<thead>
<tr>
<th>Method</th>
<th>Intercept</th>
<th>Age</th>
<th>MS</th>
<th>Age*MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML-Exch</td>
<td>-1.1195(0.0611)</td>
<td>-0.0778(0.0307)</td>
<td>0.1610(0.1069)</td>
<td>0.0385(0.0500)</td>
</tr>
<tr>
<td>ML-AR(1)</td>
<td>-1.1296(0.0590)</td>
<td>-0.0819(0.0368)</td>
<td>0.1573(0.1072)</td>
<td>0.0438(0.0595)</td>
</tr>
<tr>
<td>Gaussian-Exch</td>
<td>-1.1255(0.0648)</td>
<td>-0.0829(0.0273)</td>
<td>0.1614(0.1056)</td>
<td>0.0391(0.0448)</td>
</tr>
<tr>
<td>Gaussian-AR(1)</td>
<td>-1.1562(0.0643)</td>
<td>-0.0839(0.0354)</td>
<td>0.1645(0.1034)</td>
<td>0.0408(0.0592)</td>
</tr>
<tr>
<td>Gaussian-Autocorr</td>
<td>-1.1252(0.0650)</td>
<td>-0.0846(0.0294)</td>
<td>0.1632(0.1059)</td>
<td>0.0410(0.0485)</td>
</tr>
<tr>
<td>Gaussian-Unstr</td>
<td>-1.1228(0.0649)</td>
<td>-0.0818(0.0289)</td>
<td>0.1598(0.1059)</td>
<td>0.0381(0.0477)</td>
</tr>
<tr>
<td>GEE-I</td>
<td>-1.1259(0.0634)</td>
<td>-0.0768(0.0313)</td>
<td>0.1709(0.1028)</td>
<td>0.0367(0.0486)</td>
</tr>
<tr>
<td>GEE-ex</td>
<td>-1.1258(0.0634)</td>
<td>-0.0768(0.0313)</td>
<td>0.1708(0.1028)</td>
<td>0.0367(0.0486)</td>
</tr>
<tr>
<td>GEE-AR(1)</td>
<td>-1.1259(0.0638)</td>
<td>-0.0708(0.0318)</td>
<td>0.1709(0.1035)</td>
<td>0.0426(0.0497)</td>
</tr>
<tr>
<td>GEE-Autocorr</td>
<td>-1.1289(0.0634)</td>
<td>-0.0780(0.0314)</td>
<td>0.1679(0.1028)</td>
<td>0.0390(0.0488)</td>
</tr>
<tr>
<td>GEE-un</td>
<td>-1.1299(0.0634)</td>
<td>-0.0771(0.0314)</td>
<td>0.1638(0.1030)</td>
<td>0.0354(0.0490)</td>
</tr>
<tr>
<td>GEE-CJ(EX)</td>
<td>-1.1258(0.0634)</td>
<td>-0.0768(0.0313)</td>
<td>0.1708(0.1028)</td>
<td>0.0367(0.0486)</td>
</tr>
<tr>
<td>GEE-CJ(AR(1))</td>
<td>-1.1331(0.0636)</td>
<td>-0.0792(0.0316)</td>
<td>0.1634(0.1031)</td>
<td>0.0413(0.0492)</td>
</tr>
</tbody>
</table>
Table 4. Estimates of the correlation parameters by different methods for the example.

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML-Exch</td>
<td>0.60</td>
</tr>
<tr>
<td>ML-AR(1)</td>
<td>0.67</td>
</tr>
<tr>
<td>Gaussian-Exch</td>
<td>0.35</td>
</tr>
<tr>
<td>Gaussian-AR(1)</td>
<td>0.30</td>
</tr>
<tr>
<td>Gaussian-Autocorr</td>
<td>(0.40, 0.31, 0.30)</td>
</tr>
</tbody>
</table>
| Gaussian-Unstr       | \[
|                      | \begin{pmatrix} 1 & 0.35 & 0.31 & 0.30 \\ 1 & 0.47 & 0.32 \\ 1 & 0.38 & 1 \end{pmatrix} \]
| GEE-ex               | 0.35                       |
| GEE-AR(1)            | 0.40                       |
| GEE-Autocorr         | (0.40, 0.31, 0.30)         |
| GEE-pararr           | \[
|                      | \begin{pmatrix} 1 & 0.35 & 0.31 & 0.30 \\ 1 & 0.47 & 0.32 \\ 1 & 0.38 & 1 \end{pmatrix} \]
| GEE-un               | 0.3                        |
| GEE-CJ(EX)           | 0.3                        |
| GEE-CJ(AR(1))        | 0.3                        |

The standard errors of the estimates of $\beta_0$ and $\beta_2$ by the maximum likelihood method are the smallest and there does not appear to be a lot difference among the standard errors of the estimates by all other methods. For $\beta_1$ and $\beta_3$, it appears that the estimates by the Gaussian estimation procedures, except Gaussian-AR(1), produce the smallest standard errors, providing some support that the estimates of the regression parameters by the Gaussian estimation procedure using the general autocorrelation structure and unstructured correlation have the highest efficiency.

5 Discussion

We develop a Gaussian estimation procedure involving a working correlation matrix for the estimation of the regression parameters in longitudinal binary response data. We show that given consistent estimates of the correlation parameters of the working correlation matrix the estimates of the regression parameters are consistent. Further we show that given consistent estimates of the regression parameters, the moment estimates of the correlation parameters of the unstructured correlation matrix are consistent whatever the true correlation structure is: unstructured, general autocorrelation, AR(1) or exchangeable. Similarly, the moment estimates of the correlation parameters of the general autocorrelation structure are consistent when the true correlation
structure is general autocorrelation, AR(1) or exchangeable. In this sense the unstructured correlation matrix is most robust against misspecification by other correlation structures. The next robust is the general autocorrelation structure.

A two-step iterative procedure is suggested for the joint estimation of the regression parameters and the correlation parameters. Consistent estimates of the regression parameters as well as the correlation parameters are obtained if consistent estimates of \( \beta \), such as the GEEs, at step 1 of the two step iterative procedure are used.

Twelve estimators of the regression parameters consisting of the maximum likelihood estimates based on the multivariate probit (MP) model, four Gaussian estimates, five GEE estimates and two weighted GEE estimates are compared by simulations. Efficiencies of the estimates of the regression parameters are compared for all the methods using four different data sets generated from the MP model with latent correlation structures (i) exchangeable, (ii) AR(1), (iii) general autocorrelation and (iv) unstructured.

When compared in terms of the estimated asymptotic variances, simulations show that Gaussian estimates of the regression parameters, using the unstructured correlation matrix of the responses for a subject, are, in general, more efficient than those by the other eleven methods irrespective of the data generation method. This shows some evidence of the robustness of this method even if the correlation structure is not unstructured, but one of the other three: exchangeable, AR(1) and general autocorrelation.

All estimation methods show small biases of the estimates of the regression parameters.

Results, in terms of the MSE, show that when data are simulated using unstructured correlation structure, Gaussian-Unstr is the best for the estimation of \( \beta_1 \), agreeing with the results shown in terms of estimated variances. For the estimation of \( \beta_0 \), no method seems to perform better than any other. For data with other correlation structures there does not seem to be any significant differences in efficiency both for the estimation of \( \beta_0 \) and \( \beta_1 \).

We have written a SAS macro \%Gaussian which can be implemented to estimate the regression parameters, the parameters of the working correlation matrix and the variances and standard errors of the estimates of the regression parameters.

Acknowledgments

This research was partially supported by the Natural Science and Engineering Research Council of Canada.
Appendix: Proof of asymptotic unbiasedness of equation (2)

Since \( W_iW_i^{-1} = I_d \), using the product rule we have \( (W_i^{-1} \otimes I_d) \frac{\partial W_i}{\partial \mu} + (I_d \otimes W_i) \frac{\partial W_i^{-1}}{\partial \mu} = 0 \). Thus, \( \frac{\partial W_i^{-1}}{\partial \mu} = -(W_i^{-1} \otimes W_i^{-1}) \frac{\partial W_i}{\partial \mu} \). Further, it is easy to see that \( E(y_i^T \otimes y_i^T) = \{vec(\Sigma_i + \mu_i \mu_i^T)\}^T = vec^T(\Sigma_i) + \mu_i^T \otimes \mu_i^T \). Therefore, we obtain \( E \left( \frac{\partial l}{\partial \beta} \right) = vec^T(\Sigma_i) \frac{\partial W_i^{-1}}{\partial \mu_i} = -\{I_d \otimes \Sigma_i vec(I_d)\}^T(W_i^{-1} \otimes W_i^{-1}) \frac{\partial W_i}{\partial \mu_i} = -vec^T(I_d)(W_i^{-1} \otimes \Sigma_iW_i^{-1}) \frac{\partial W_i}{\partial \mu_i} \). Now suppose the estimate \( \hat{\rho} \) of the working correlation converges to the true correlation matrix \( C(\rho) \) in probability. Then asymptotically, as \( K \to \infty \), \( \Sigma_iW_i^{-1} = A_i^{1/2}C(\rho)A_i^{1/2} A_i^{-1/2}R^{-1}(\hat{\rho})A_i^{-1/2} = A_i^{1/2}C(\rho)R^{-1}(\hat{\rho})A_i^{-1/2} = I_d \). Using this and the fact that \( vecA = (A^T \otimes I_m)vecI_m \), where \( A \) is an \( m \times n \) matrix, we obtain

\[
-2 E \left( \frac{\partial l}{\partial \beta} \right) = \left( vec^T(W_i^{-1}) \frac{\partial W_i}{\partial \mu_i} + vec^T(\Sigma_i) \frac{\partial W_i^{-1}}{\partial \mu_i} \right) \frac{\partial \mu_i}{\partial \beta} \\
= (vec^T(W_i^{-1}) - vec^T(I_d)(W_i^{-1} \otimes I_d)) \frac{\partial W_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta} \\
= \left\{ [(W_i^{-1} \otimes I_d)vec(I_d)]^T - vec^T(I_d)(W_i^{-1} \otimes I_d) \right\} \frac{\partial W_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta} = 0.
\]

Thus, \( E \left( \frac{\partial l}{\partial \beta} \right) = \sum_{i=1}^K E \left( \frac{\partial l}{\partial \beta} \right) = 0 \), so that the estimating equations (2) are asymptotically unbiased.

References


