Biometrika (2006), **93,** 4, *pp.* 927–941

Printed in Great Britain

Modelling of covariance structures in generalised estimating equations for longitudinal data

By HUAJUN YE and JIANXIN PAN

School of Mathematics, The University of Manchester, PO Box 88, Manchester, M60 1QD, U.K.

yhuajun@maths.man.ac.uk jpan@maths.man.ac.uk

Summary

When used for modelling longitudinal data generalised estimating equations specify a working structure for the within-subject covariance matrices, aiming to produce efficient parameter estimators. However, misspecification of the working covariance structure may lead to a large loss of efficiency of the estimators of the mean parameters. In this paper we propose an approach for joint modelling of the mean and covariance structures of longitudinal data within the framework of generalised estimating equations. The resulting estimators for the mean and covariance parameters are shown to be consistent and asymptotically Normally distributed. Real data analysis and simulation studies show that the proposed approach yields efficient estimators for both the mean and covariance parameters.

Some key words: Cholesky decomposition; Efficiency; Generalised estimating equation; Longitudinal data; Misspecification of covariance structure; Modelling of mean and covariance structures.

1. Introduction

The technique of generalised estimating equations (Liang & Zeger, 1986) for the modelling of longitudinal data specifies a working structure for the within-subject covariance matrices, aiming to produce efficient estimators for the mean parameters. It is well known that the resulting estimators of the mean parameters are consistent even though the working covariance structure is misspecified. However, this may lead to a great loss of efficiency of the mean parameter estimators (Wang & Carey, 2003). Also, if longitudinal data contain certain missing values and/or are not Normally distributed, the mean parameter estimators may be biased when the working covariance structure is misspecified (Daniels & Zhao, 2003). A good covariance modelling approach improves statistical inference of the mean of interest. Furthermore, the covariance structure itself may be of scientific interest (Diggle & Verbyla, 1998).

However, modelling of covariance structures is challenging because there are many parameters in covariance matrices and the estimated covariance matrices should be positive definite. So-called sandwich-type methods have been proposed. For example, Prentice & Zhao (1991) and Liang et al. (1992) use the products of all pairwise observations to build new generalised estimating equations for the correlation parameters. Carey et al. (1993) addressed the issue specifically for longitudinal binary and ordinal outcomes. Wang & Carey (2003) studied the effects of misspecification of covariance structures on efficiency of the mean parameters. Wang & Carey (2004) further proposed unbiased estimating equations for the correlation parameters by differentiating the Cholesky decomposition of the working correlation matrix.

Instead of using the sandwich kind of working covariance structure, we use the modified Cholesky decomposition (Pourahmadi, 1999) to decompose the within-subject covariance matrices and then parsimoniously model the within-subject correlation and variation in terms of simple regression models. We propose an approach for joint modelling of the mean and covariance structures of longitudinal data within the framework of generalised estimating equations. The resulting estimators for the mean and covariance parameters are shown to be consistent and asymptotically Normally distributed. In comparison with Pourahmadi's (1999, 2000) method, the proposed approach does not require the Normal distributional assumption and only assumes the existence of the first four moments of responses. We extend Pourahmadi's work from modelling balanced longitudinal data to unbalanced data but within a broader framework, that is generalised estimating equations. Compared to the conventional generalised estimating equations, on the other hand, the proposed approach produces more efficient estimators for the mean parameters.

2. Joint mean-covariance generalised linear models

2·1. *Modified Cholesky decomposition*

Let y_{ij} be the jth of m_i measurements on the ith of n subjects. Assume that t_{ij} is the time at which the measurement y_{ij} is made. Denote by $y_i=(y_{i1}, y_{i2}, \ldots, y_{im_i})'$ and $t_i=(t_{i1}, t_{i2}, \ldots, t_{im_i})'$ the $(m_i \times 1)$ vectors of responses and measuring time points of the ith subject. Suppose $E(y_i)=\mu_i=(\mu_{i1}, \mu_{i2}, \ldots, \mu_{im_i})'$ and $\text{var}(y_i)=\Sigma_i$ are the $(m_i \times 1)$ mean vector and $(m_i \times m_i)$ covariance matrix of y_i, respectively. Without loss of generality, the matrices Σ_i are assumed to be positive definite.

Accordingly, there exists a unique lower triangular matrix T_i with 1's as diagonal entries and a unique diagonal matrix D_i with positive diagonals such that $T_i\Sigma_i T_i'=D_i$. This modified Cholesky decomposition has a clear statistical interpretation: the below-diagonal entries of T_i are the negatives of the autoregressive coefficients, ϕ_{ijk}, in the autoregressive model

$$\hat{y}_{ij}=\mu_{ij}+\sum_{k=1}^{j-1}\phi_{ijk}(y_{ik}-\mu_{ik}), \tag{1}$$

that is the linear least squares predictor of y_{ij} based on its predecessors $y_{i(j-1)}, \ldots, y_{i1}$. On the other hand, we can show that the diagonal entries of D_i are the prediction error/innovation variances $\sigma_{ij}^2=\text{var}(\varepsilon_{ij})$, where $\varepsilon_{ij}=y_{ij}-\hat{y}_{ij}$ and $\hat{y}_{ij}$ are given in (1) $(1 \leqslant j \leqslant m_i; 1 \leqslant i \leqslant n)$ (Pourahmadi, 1999). Obviously, we have $\text{cov}(\varepsilon_{ij}, \varepsilon_{ik})=0$ if $j \neq k$. Throughout this paper we refer to ϕ_{ijk} as generalised autoregressive parameters and to σ_{ij}^2 as innovation variances. It follows immediately that $\Sigma_i^{-1}=T_i'D_i^{-1}T_i$.

2·2. *Generalised linear models*

In the spirit of Pourahmadi (1999), we propose three generalised linear models for modelling the mean, generalised autoregressive parameters and innovation variances:

$$g(\mu_{ij})=x_{ij}'\beta, \quad \phi_{ijk}=z_{ijk}'\gamma, \quad \log\sigma_{ij}^2=z_{ij}'\lambda, \tag{2}$$

where x_{ij}, z_{ijk} and z_{ij} are $(p \times 1)$, $(q \times 1)$ and $(d \times 1)$ vectors of covariates and β, γ and λ are the associated parameters. The link function $g(.)$ is assumed to be monotone and differentiable. The covariates x_{ij}, z_{ijk} and z_{ij} may contain baseline covariates, polynomials in time and their interactions as well. For example, when we use polynomials in time to model the mean, generalised autoregressive parameters and innovation variances the covariates may take the forms

$$\begin{aligned} x_{ij} &= (1, t_{ij}, t_{ij}^2, \ldots, t_{ij}^{p-1})', \\ z_{ijk} &= (1, (t_{ij} - t_{ik}), (t_{ij} - t_{ik})^2, \ldots, (t_{ij} - t_{ik})^{q-1})', \\ z_{ij} &= (1, t_{ij}, t_{ij}^2, t_{ij}^{d-1})', \end{aligned} \tag{3}$$

provided that the within-subject correlation only depends on the elapsed time $(i = 1, \ldots, n;\ j = 1, \ldots, m_i)$. An advantage of the use of the generalised linear models (2) is that the resulting covariance matrices are positive definite (Pourahmadi, 1999). Moreover, in the generalised linear models (2) the assumption of homogeneous covariances across subjects becomes testable (Pan & MacKenzie, 2003).

3. Generalised estimating equations estimation

3·1. *Generalised estimating equations*

To obtain generalised estimating equations, rather than specifying a working covariance structure, we model jointly the mean and covariance structures of responses in terms of the generalised linear models (2). We therefore propose the following generalised estimating equations for the mean, generalised autoregressive parameters and innovation variances, respectively:

$$\begin{gathered} S_1(\beta) = \sum_{i=1}^{n} \left(\frac{\partial \mu_i'}{\partial \beta}\right) \Sigma_i^{-1}(y_i - \mu_i), \quad S_2(\gamma) = \sum_{i=1}^{n} \left(\frac{\partial \hat{r}_i'}{\partial \gamma}\right) D_i^{-1}(r_i - \hat{r}_i), \\ S_3(\lambda) = \sum_{i=1}^{n} \left(\frac{\partial \sigma_i^{2'}}{\partial \lambda}\right) W_i^{-1}(\varepsilon_i^2 - \sigma_i^2), \end{gathered} \tag{4}$$

where r_i and $\hat{r}_i$ in $S_2(\gamma)$ are the $(m_i \times 1)$ vectors with jth components $r_{ij} = y_{ij} - \mu_{ij}$ and $\hat{r}_{ij} = E(r_{ij} | r_{i1}, \ldots, r_{i(j-1)}) = \sum_{k=1}^{j-1} \phi_{ijk} r_{ik}$ $(j = 1, \ldots, m_i)$, respectively. Note that when $j = 1$ the notation $\sum_{k=1}^{0}$ means zero throughout this paper. We can show that $D_i = \text{diag}(\sigma_{i1}^2, \ldots, \sigma_{im_i}^2)$ in $S_2(\gamma)$ is actually the covariance matrix of $r_i - \hat{r}_i$. On the other hand, ε_i^2 and σ_i^2 in $S_3(\lambda)$ are the $(m_i \times 1)$ vectors with jth components ε_{ij}^2 and σ_{ij}^2 $(j = 1, \ldots, m_i)$, respectively, where $\varepsilon_{ij} = y_{ij} - \hat{y}_{ij}$ and $\hat{y}_{ij}$ are given in (1). Obviously, we have $E(\varepsilon_i^2) = \sigma_i^2$. In addition, W_i is the covariance matrix of ε_i^2; that is, $W_i = \text{var}(\varepsilon_i^2)$. The solutions of the generalised estimating equations, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\lambda}$ say, are termed the generalised estimating equation estimators of β, γ and λ.

In (4), $\partial \mu_i' / \partial \beta$ is the $(p \times m_i)$ matrix with jth column $\partial \mu_{ij} / \partial \beta = \dot{g}^{-1}(x_{ij}\beta) x_{ij}$, where $\dot{g}^{-1}(.)$ is the derivative of the inverse function $g^{-1}(.)$, $\partial \hat{r}_i' / \partial \gamma$ is the $(q \times m_i)$ matrix with jth column $\partial \hat{r}_{ij} / \partial \gamma = \sum_{k=1}^{j-1} r_{ik} z_{ijk}$, and $\partial \sigma_i^{2'} / \partial \lambda$ is the $(d \times m_i)$ matrix with jth column $\partial \sigma_{ij}^2 / \partial \lambda = \sigma_{ij}^2 z_{ij}$ $(i = 1, \ldots, n;\ j = 1, \ldots, m_i)$. The idea behind (4) is to treat the generalised autoregressive parameters and innovation variances as being as important as the mean when modelling longitudinal data. In contrast to the mean equation $S_1(\beta)$, r_i in $S_2(\gamma)$ and ε_i^2 in $S_3(\lambda)$

play a role similar to that of responses in the estimation of γ and λ, which can be viewed as working responses when modelling the generalised autoregressive parameters and innovation variances.

Note that $S_3(\lambda)$ in (4) involves the covariance matrices W_i of ε_i^2. When the responses y_i are Normally distributed, it is obvious that $\varepsilon_i \sim N_{m_i}(0, D_i)$. Since D_i is diagonal, ε_{ij} and ε_{ik} $(j \neq k)$ are independent and so ε_{ij}^2 and ε_{ik}^2 $(j \neq k)$ are independent as well. On the other hand, since $\varepsilon_{ij}^2/\sigma_{ij}^2 \sim \chi_1^2$ it is obvious that $\text{var}(\varepsilon_{ij}^2) = 2\sigma_{ij}^4$. Therefore, if $y_i \sim N_{m_i}(\mu_i, \Sigma_i)$ we then have $W_i = 2\,\text{diag}(\sigma_{i1}^4, \ldots, \sigma_{im_i}^4)$. In this case the generalised estimating equations (4) reduce to Pourahmadi's (2000) Normal score equations, though he only considered modelling of balanced longitudinal data there. We will explain this equivalence in more detail in § 3·3. When the responses y_i are not Normally distributed, the covariance matrices W_i of ε_i^2 are in general no longer diagonal and remain unknown. In fact, it is very troublesome to calculate the W_i's in this case and their analytically explicit forms may not exist even for some special family of distributions such as the exponential family of distributions or the quadratic exponential family of distributions (Prentice & Zhao, 1991). In the spirit of the conventional generalised estimating equations, we propose to use a sandwich 'working' covariance structure $W_i = A_i^{\frac{1}{2}} R_i(\rho) A_i^{\frac{1}{2}}$ to approximate the true W_i's, where $A_i = 2\,\text{diag}(\sigma_{i1}^4, \ldots, \sigma_{im_i}^4)$ and $R_i(\rho)$ mimic the correlation between ε_{ij}^2 and ε_{ik}^2 $(i \neq k)$ by introducing a new parameter ρ. Typical structures for $R_i(\rho)$ include compound symmetry and AR(1). As with the conventional generalised estimating equations for the mean, the parameter ρ may have little effect on the estimates of γ and λ. Our real data analysis and simulation studies reported in later sections confirm this point very well. It implies that the resulting estimators of parameters in the mean, generalised autoregressive parameters and innovation variances are robust against misspecification of $R_i(\rho)$, as does the efficiency of the estimators of the mean parameters of interest.

3·2. *Estimators of parameters*

The estimators of β, γ and λ satisfy the equations

$$S_1(\beta) = 0, \quad S_2(\gamma) = 0, \quad S_3(\lambda) = 0, \tag{5}$$

where $S_1(\beta)$, $S_2(\gamma)$ and $S_3(\lambda)$ are given in (4). In general numerical solution is necessary, and in this paper we use a quasi-Fisher scoring algorithm. We first calculate the quasi-Fisher information matrix $\mathscr{I}_\theta$, defined by the expectation of minus the derivative of the score function $S'(\theta) = (S_1'(\beta), S_2'(\gamma), S_3'(\lambda))$ with respect to θ, where $\theta = (\beta', \gamma', \lambda')'$. It can be shown that $\mathscr{I}_\theta$ is block diagonal; that is, $\mathscr{I}_\theta = \text{diag}(\mathscr{I}_\beta, \mathscr{I}_\gamma, \mathscr{I}_\lambda)$ with

$$\begin{gathered} \mathscr{I}_\beta = \sum_{i=1}^{n} \left(\frac{\partial \mu_i'}{\partial \beta}\right) \Sigma_i^{-1} \left(\frac{\partial \mu_i'}{\partial \beta}\right)', \quad \mathscr{I}_\gamma = \sum_{i=1}^{n} E\left\{\left(\frac{\partial \hat{r}_i'}{\partial \gamma}\right) D_i^{-1} \left(\frac{\partial \hat{r}_i'}{\partial \gamma}\right)'\right\}, \\ \mathscr{I}_\lambda = \sum_{i=1}^{n} \left(\frac{\partial \sigma_i^{2'}}{\partial \lambda}\right) W_i^{-1} \left(\frac{\partial \sigma_i^{2'}}{\partial \lambda}\right)', \end{gathered} \tag{6}$$

where $\mathscr{I}_\gamma$ can be further simplified into

$$\mathscr{I}_\gamma = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sigma_{ij}^{-2} \sum_{k=1}^{j-1} \sum_{l=1}^{j-1} \sigma_{ikl} z_{ijk} z_{ijl}' \tag{7}$$

and σ_{ikl} is the (k, l)th element of the covariance matrix Σ_i. Therefore, given Σ_i, the estimators of the mean parameters β can be updated by

$$\hat{\beta} = \left\{ \sum_{i=1}^{n} \left(\frac{\partial \mu_i'}{\partial \beta} \right) \Sigma_i^{-1} \left(\frac{\partial \mu_i'}{\partial \beta} \right)' \right\}^{-1} \left\{ \sum_{i=1}^{n} \left(\frac{\partial \mu_i'}{\partial \beta} \right) \Sigma_i^{-1} \tilde{y}_i \right\}, \tag{8}$$

where $\tilde{y}_i = (y_i - \mu_i) + (\partial \mu_i'/\partial \beta)'\beta$. When $g(.)$ is the identity link, we have $\tilde{y}_i \equiv y_i$. On the other hand, given β and λ the values for the generalised autoregressive parameters γ can be updated approximately through

$$\hat{\gamma} = \left[\sum_{i=1}^{n} E\left\{ \left(\frac{\partial \hat{r}_i'}{\partial \gamma} \right) D_i^{-1} \left(\frac{\partial \hat{r}_i'}{\partial \gamma} \right)' \right\} \right]^{-1} \left\{ \sum_{i=1}^{n} \left(\frac{\partial \hat{r}_i'}{\partial \gamma} \right) D_i^{-1} r_i \right\}. \tag{9}$$

Finally, given β and γ the values for the innovation variance parameters λ can be updated using

$$\hat{\lambda} = \left\{ \sum_{i=1}^{n} \left(\frac{\partial \sigma_i^{2'}}{\partial \lambda} \right) W_i^{-1} \left(\frac{\partial \sigma_i^{2'}}{\partial \lambda} \right)' \right\}^{-1} \left\{ \sum_{i=1}^{n} \left(\frac{\partial \sigma_i^{2'}}{\partial \lambda} \right) W_i^{-1} \tilde{\varepsilon}_i^2 \right\}, \tag{10}$$

where $\tilde{\varepsilon}_i^2 = (\varepsilon_i^2 - \sigma_i^2) + D_i \log \sigma_i^2$ and $\log \sigma_i^2$ stands for the $(m_i \times 1)$ vector with jth component $\log \sigma_{ij}^2$ $(i = 1, \ldots, n;\ j = 1, \ldots, m_i)$.

Equations (8)–(10) indicate that, iteratively, the set of all parameters can be estimated using weighted generalised least squares. Moreover, approximate variance-covariance matrices of the estimators $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\lambda}$ can be calculated through the inverses of the quasi-Fisher information matrices $\mathscr{I}_\beta$, $\mathscr{I}_\gamma$ and $\mathscr{I}_\lambda$, respectively, evaluated at $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\lambda}$.

In summary, the algorithm below is used to calculate the parameter estimates in the mean-covariance models in (2).

Algorithm

Step 1. *Given a starting value* $\theta^{(0)} = (\beta^{(0)'}, \gamma^{(0)'}, \lambda^{(0)'})'$, *use the generalised linear models* (2) *to form the lower triangular matrices* $T_i^{(0)}$ *and diagonal matrices* $D_i^{(0)}$. *Then* $\Sigma_i^{(0)}$, *the starting values of* Σ_i, *are obtained.*

Step 2. *Use the weighted generalised least squares estimators* (8)–(10) *to calculate the estimators* $\beta^{(1)}$, $\gamma^{(1)}$ *and* $\lambda^{(1)}$ *of the parameters* β, γ *and* λ, *respectively.*

Step 3. *Replace* $\beta^{(0)}$, $\gamma^{(0)}$ *and* $\lambda^{(0)}$ *with the estimators* $\beta^{(1)}$, $\gamma^{(1)}$ *and* $\lambda^{(1)}$. *Repeat Steps* 1–2 *until convergence of the parameter estimators.*

A convenient starting value for $(\gamma', \lambda')'$ is $\gamma^{(0)} = 0$ and $\lambda^{(0)} = 0$. In other words, the $(m_i \times m_i)$ identity matrix may be chosen as the starting value for the covariance matrix Σ_i.

3·3. *Asymptotic properties*

The consistency and asymptotic Normality of the generalised estimating equation estimators $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\lambda}$ are presented in Theorems 1–2 below. Their proofs are deferred to the Appendix.

Theorem 1. *Suppose there is only one root* $\hat{\theta}_n = (\hat{\beta}_n', \hat{\gamma}_n', \hat{\lambda}_n')'$ *of the generalised estimating equations* (4). *Under some mild regularity conditions stated in the Appendix the generalised estimating equation estimator* $\hat{\theta}_n = (\hat{\beta}_n', \hat{\gamma}_n', \hat{\lambda}_n')'$ *is strongly consistent for the true value* $\theta_0 = (\beta_0', \gamma_0', \lambda_0')'$; *that is,* $\hat{\theta}_n = (\hat{\beta}_n', \hat{\gamma}_n', \hat{\lambda}_n')' \to \theta_0 = (\beta_0', \gamma_0', \lambda_0')'$ *almost surely as* $n \to \infty$.

The proof of asymptotic Normality involves computation of the covariance matrix of the function $S(\theta)/\sqrt{n}=(S_1'(\beta), S_2'(\gamma), S_3'(\lambda))'/\sqrt{n}$, denoted by $V_n=(v_n^{kl})_{k,l=1,2,3}$, where $v_n^{kl}=n^{-1}\,\mathrm{cov}(S_k, S_l)$ for $k\neq l$ and $v_n^{kk}=n^{-1}\,\mathrm{var}(S_k)$ $(k, l=1, 2, 3)$. When evaluated at the true value θ_0 the covariance matrix V_n is assumed to be positive definite. Furthermore, at θ_0 we assume that

$$V_n=\begin{pmatrix} v_n^{11} & v_n^{12} & v_n^{13} \\ v_n^{21} & v_n^{22} & v_n^{23} \\ v_n^{31} & v_n^{32} & v_n^{33} \end{pmatrix} \to V=\begin{pmatrix} v^{11} & v^{12} & v^{13} \\ v^{21} & v^{22} & v^{23} \\ v^{31} & v^{32} & v^{33} \end{pmatrix} \qquad (11)$$

as $n\to\infty$. The constant matrix V in (11) is also assumed to be positive definite.

THEOREM 2. *Suppose that* (11) *above is true. Under some necessary regularity conditions stated in the Appendix the generalised estimating equation estimator* $\hat{\theta}_n=(\hat{\beta}_n', \hat{\gamma}_n', \hat{\lambda}_n')'$ *is asymptotically Normally distributed, with*

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_n-\beta_0 \\ \hat{\gamma}_n-\gamma_0 \\ \hat{\lambda}_n-\lambda_0 \end{pmatrix} \to N\left\{0, \begin{pmatrix} v^{11} & 0 & 0 \\ 0 & v^{22} & 0 \\ 0 & 0 & v^{33} \end{pmatrix}^{-1} \begin{pmatrix} v^{11} & v^{12} & v^{13} \\ v^{21} & v^{22} & v^{23} \\ v^{31} & v^{32} & v^{33} \end{pmatrix} \begin{pmatrix} v^{11} & 0 & 0 \\ 0 & v^{22} & 0 \\ 0 & 0 & v^{33} \end{pmatrix}^{-1}\right\}$$

in distribution as $n\to\infty$, *where the matrices* v^{kl} *are evaluated at the true value* $\theta=\theta_0$ $(k, l=1, 2, 3)$.

Note that when the responses y_i are Normally distributed we have $v^{kl}=0$ $(k\neq l)$ and then the asymptotic covariance matrix above reduces to $\{\mathrm{diag}(v^{11}, v^{22}, v^{33})\}^{-1}$. In other words, the asymptotic covariance matrix is equal to n times the inverse of the quasi-Fisher information matrix, $n\mathscr{I}_\theta^{-1}$. By analysing balanced longitudinal data, Pourahmadi (2000) obtained the Fisher information matrix $\mathscr{I}_\theta$ from the Normal likelihood. He concluded that the off-diagonal blocks $\mathscr{I}_{\gamma\beta}$ and $\mathscr{I}_{\lambda\beta}$ are equal to zero whereas $\mathscr{I}_{\lambda\gamma}=-nZ'D^{-1}B$, where Z is the design matrix for the log-innovation variances, $B=(b_1, \ldots, b_m)'$ is a matrix with $b_t=\sum_{k=1}^{t-1} a_{kt}z_{tk}$ and a_{kt} is the (k, t)th entry of the matrix $A=\Sigma T'$ $(t=1, \ldots, m)$; see the Appendix of Pourahmadi (2000). Since $A=\Sigma T'=T^{-1}D$, a product of the lower triangular matrix T^{-1} and the diagonal matrix D, the matrix A is actually a lower triangular matrix. Accordingly, when $k<t$ we have $a_{kt}=0$ so that $b_t=0$ for $t=1, \ldots, m$. In other words, we obtain $\mathscr{I}_{\lambda\gamma}=0$ because $B=0$. Therefore, Pourahmadi's (2000) asymptotic covariance matrix is actually block diagonal. We conclude that under the Normal distribution assumption our proposed approach reduces precisely to Pourahmadi's (2000) method.

4. TESTS OF HYPOTHESES

The classical score test is based on the first and second derivatives of the loglikelihood evaluated at the parameter estimators under the null hypothesis (Cox & Hinkley, 1974). Within the framework of the generalised estimating equations (4), the quasi-score test based on the derivative of the generalised estimating equations can be constructed and used to undertake tests of hypotheses.

For example, suppose that we are interested in testing $H_0:\beta=\beta_0$, where β_0 is a fixed p-dimensional vector. Let $\alpha=(\gamma', \lambda')'$ and $S_{23}(\alpha)=(S_2(\gamma)', S_3(\lambda)')'$. The covariance matrix

V_n of $S(\theta)/\sqrt{n} = (S_1'(\beta), S_{23}'(\alpha))'/\sqrt{n}$ is partitioned into

$$V_n = \begin{pmatrix} v_n^{11} & v_n^{1\alpha} \\ v_n^{\alpha 1} & v_n^{\alpha\alpha} \end{pmatrix},$$

where

$$v_n^{\alpha\alpha} = \begin{pmatrix} v_n^{22} & v_n^{33} \\ v_n^{32} & v_n^{33} \end{pmatrix}, \quad v_n^{1\alpha} = (v_n^{12}, v_n^{13}) = (v_n^{\alpha 1})'.$$

For this H_0, the parameters $\alpha = (\gamma', \lambda')'$ are nuisance parameters. Hence the quasi-score test statistic for $H_0 : \beta = \beta_0$ is of the form

$$J_{\beta_0} = n^{-1} S_1(\beta_0, \hat{\alpha}_0)' V_n^{\beta_0\beta_0}(\hat{\alpha}_0) S_1(\beta_0, \hat{\alpha}_0), \tag{12}$$

where $\hat{\alpha}_0$ is the generalised estimating equation estimator of α under H_0, $S_1(\beta_0, \hat{\alpha}_0)$ is the score function $S_1(\beta_0)$ but with α replaced by $\hat{\alpha}_0$ and

$$V_n^{\beta_0\beta_0}(\hat{\alpha}_0) = \{v_n^{11} - v_n^{1\alpha}(v_n^{\alpha\alpha})^{-1} v_n^{\alpha 1}\}^{-1}|_{\alpha = \hat{\alpha}_0, \beta = \beta_0}.$$

It can be shown that, under H_0, $J_{\beta_0} \sim \chi_p^2$ asymptotically.

On the other hand, we may wish to test $H_0 : \gamma = \gamma_0$ and $\lambda = \lambda_0$, where γ_0 and λ_0 are fixed q- and d-dimensional vectors, respectively. The quasi-score test statistic is then of the form

$$J_{\alpha_0} = n^{-1} S_{23}(\hat{\beta}_0, \alpha_0)' V_n^{\alpha_0\alpha_0}(\hat{\beta}_0) S_{23}(\hat{\beta}_0, \alpha_0), \tag{13}$$

where $\alpha_0 = (\gamma_0', \lambda_0')'$, $\hat{\beta}_0$ is the generalised estimating equation estimator of β under the null hypothesis $H_0 : \gamma = \gamma_0$ and $\lambda = \lambda_0$, $S_{23}(\hat{\beta}_0, \alpha_0)$ is the score function $S_{23}(\alpha_0)$ but with β replaced by $\hat{\beta}_0$ and $V_n^{\alpha_0\alpha_0}(\hat{\beta}_0) = \{v_n^{\alpha\alpha} - v_n^{1\alpha}(v_n^{11})^{-1} v_n^{\alpha 1}\}^{-1}|_{\beta = \hat{\beta}_0, \alpha = \alpha_0}$. Under H_0, $J_{\alpha_0} \sim \chi_{q+d}^2$ asymptotically.

Similar procedures exist for testing $H_0 : \gamma = \gamma_0$ or $H_0 : \lambda = \lambda_0$ but the details are omitted here.

5. Analysis of the CD4+ cell data

We reanalyse the CD4+ cell data (Diggle et al., 2002) comprising CD4+ cell counts for 369 HIV-infected men. Altogether there are 2376 values of CD4+ cell counts, with several repeated measurements being made for each individual at different times covering a period of approximately eight and a half years. The number m_i of measurements for each individual varies from 1 to 12 and the times are not equally spaced. The CD4+ cell data are highly unbalanced. For further details about design and medical implications of the study we refer to Diggle et al. (2002).

The objective of our analysis is to model jointly the mean and covariance structures for the CD4+ cell data. Based on the principle of model selection and our experience of nonparametric modelling for this dataset, we propose to use three polynomials in time, one of degree 6 and two cubics, to model the mean parameters, the generalised autoregressive parameters and the innovation variances; that is, the covariates x_{ij}, z_{ijk} and z_{ij} take the same forms as in (3) with $p = 7$, $q = 4$ and $d = 4$. Note that there are 54 individuals, about 15%, for whom m_i is less than q and d, implying that the dimensions of the covariates z_{ijk} and z_{ij} can be greater than the number of repeated measurements of some subjects.

We treat the CD4+ cell data as continuous and take the function $g(.)$ to be the identity link. We consider two kinds of correlation structure, compound symmetry and AR(1), for the matrices $R_i(\rho)$ in $W_i = A_i^{\frac{1}{2}} R_i(\rho) A_i^{\frac{1}{2}}$, the working covariance structures of ε_i^2. In each case the parameter ρ measuring the correlation between $\varepsilon_{ij}^2 = (y_{ij} - \hat{y}_{ij})^2$ and $\varepsilon_{ik}^2 = (y_{ik} - \hat{y}_{ik})^2$ takes different values, $\rho = 0{\cdot}2$, $0{\cdot}5$ and $0{\cdot}8$, so that effects of misspecification of $R_i(\rho)$ on the generalised estimating equation estimators $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\lambda}$ can be studied. We also look at the extreme case $\rho = 0$, corresponding to the model with a Normal distribution. Table 1 presents the parameter estimate and the associated standard errors, in parentheses, based on the compound symmetry structure; the results obtained based on the AR(1) structure were very similar.

Table 1: *The CD4+ cell data. Generalised estimating equation estimates of parameters in the mean, generalised autoregressive parameters and innovation variances based on compound symmetry structure, with standard errors in parentheses*

		Compound symmetry		
	Normal	$\rho = 0{\cdot}2$	$\rho = 0{\cdot}5$	$\rho = 0{\cdot}8$
β_1	875·22 (15·25)	875·93 (15·33)	876·20 (15·35)	876·31 (15·36)
β_2	−207·27 (13·23)	−205·85 (13·11)	−205·48 (13·05)	−205·38 (13·02)
β_3	−22·48 (8·38)	−22·82 (8·69)	−22·87 (8·88)	−22·88 (8·97)
β_4	32·21 (4·77)	31·64 (4·73)	31·46 (4·72)	31·40 (4·72)
β_5	−1·31 (0·91)	−1·20 (0·94)	−1·17 (0·96)	−1·17 (0·98)
β_6	−1·92 (0·48)	−1·87 (0·49)	−1·85 (0·49)	−1·85 (0·50)
β_7	0·25 (0·06)	0·24 (0·06)	0·24 (0·06)	0·24 (0·06)
γ_1	0·62 (0·04)	0·62 (0·04)	0·61 (0·04)	0·61 (0·04)
γ_2	−0·51 (0·06)	−0·50 (0·06)	−0·49 (0·06)	−0·49 (0·06)
γ_3	0·15 (0·03)	0·14 (0·03)	0·14 (0·02)	0·14 (0·02)
γ_4	$-1{\cdot}45 \times 10^{-2}$ $(3{\cdot}06 \times 10^{-3})$	$-1{\cdot}40 \times 10^{-2}$ $(2{\cdot}95 \times 10^{-3})$	$-1{\cdot}38 \times 10^{-2}$ $(2{\cdot}90 \times 10^{-3})$	$-1{\cdot}37 \times 10^{-2}$ $(2{\cdot}88 \times 10^{-3})$
λ_1	11·55 (0·03)	11·55 (0·04)	11·55 (0·04)	11·55 (0·05)
λ_2	−0·34 (0·02)	−0·37 (0·02)	−0·38 (0·02)	−0·39 (0·03)
λ_3	$-4{\cdot}71 \times 10^{-2}$ $(8{\cdot}66 \times 10^{-3})$	$-4{\cdot}39 \times 10^{-2}$ $(8{\cdot}11 \times 10^{-3})$	$-4{\cdot}38 \times 10^{-2}$ $(8{\cdot}25 \times 10^{-3})$	$-4{\cdot}27 \times 10^{-2}$ $(8{\cdot}69 \times 10^{-3})$
λ_4	$1{\cdot}80 \times 10^{-2}$ $(2{\cdot}29 \times 10^{-3})$	$1{\cdot}45 \times 10^{-2}$ $(2{\cdot}13 \times 10^{-3})$	$1{\cdot}27 \times 10^{-2}$ $(1{\cdot}72 \times 10^{-3})$	$1{\cdot}19 \times 10^{-2}$ $(1{\cdot}70 \times 10^{-3})$

In Table 1 it is shown that the parameter ρ has little effect on the estimators of β, γ and λ, implying that the generalised estimating equation estimators of parameters are robust against misspecification of the structure of $R_i(\rho)$. This point is confirmed by our simulation studies in § 6. Figure 1 displays the three fitted curves when $R_i(\rho)$ is specified by AR(1) with $\rho = 0{\cdot}5$. Bands made up of the asymptotic 95% confidence intervals are also provided. The mean curve of CD4+ cells changes slowly before seroconversion, at $t = 0$, and then drops steeply before levelling off after two years. The resulting mean trajectory is close to the Diggle et al. (2002) fitted curve based on a smoothing spline. The structures of the generalised autoregressive parameters and (log)innovation variances both clearly display a cubic polynomial pattern.

Next, we compare the proposed approach with the conventional generalised estimating equations in terms of relative efficiency of the estimators of the mean parameters β. The relative efficiency of β_k $(k = 1, \ldots, p)$ is the ratio of the variance of the conventional

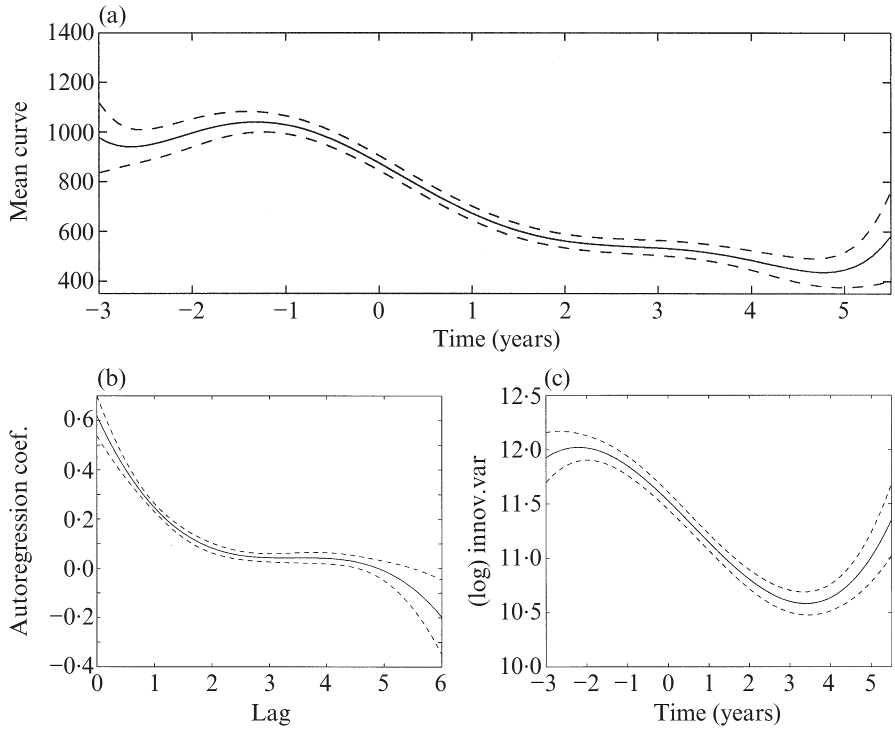

Fig. 1: The CD4+ cell data. The fitted curves of (a) the mean against time, (b) the generalised autoregressive parameters against lag and (c) the (log)innovation variances against time. Dashed curves represent asymptotic 95% confidence intervals.

estimator $\hat{\beta}_k^{\mathrm{C}}$ to the variance of the new estimator $\hat{\beta}_k^{\mathrm{N}}$.

$$e(\hat{\beta}_k) = \mathrm{var}(\hat{\beta}_k^{\mathrm{C}})/\mathrm{var}(\hat{\beta}_k^{\mathrm{N}}). \tag{14}$$

For simplicity, the same working correlation structure is used in both approaches in our analysis. The values for $e(\hat{\beta}_k)$ were almost all in the range 1·07–1·19 for the case of compound symmetry and in the range 1·20–1·67 for the AR(1) case. Obviously, the efficiency of the conventional estimators can be improved by using the new approach.

Finally, we use the quasi-score test statistic (13) to test if the null hypothesis $H_0: \gamma = 0$ and $\lambda = 0$ is true. We find that the value of the test statistic is very large when compared to the χ^2 distribution with $q + d = 8$ degrees of freedom. This also occurs when testing $H_0: \gamma = 0$ or $H_0: \lambda = 0$, which implies that there are strong within-subject correlations or heterogeneous innovation variances for the CD4+ cell data.

6. A simulation study

In the simulation study we use the same design protocol as in the CD4+ cell data. We generate 1000 random samples from Normal and Normal-mixture distributions, respectively. Each sample comprises 369 subjects and the m_i's, the numbers of repeated measurements, are the same as in the real CD4+ data. The Normal distribution studied is $N_{m_i}(\mu_i, \Sigma_i)$ and the Normal-mixture distribution is

$$F_{m_i} = \pi N_{m_i}(\mu_i + \delta_i, \Sigma_i) + (1 - \pi) N_{m_i}(\mu_i, \Sigma_i) \quad (i = 1, 2, \ldots, 369),$$

where μ_i and Σ_i are formed using the parameter estimates in the real data analysis in § 5, π is the mixing weight and δ_i is the mean-shift parameter. In order to see how the proposed approach behaves under different levels of mixtures, we choose $\pi = 0{\cdot}25$, $0{\cdot}5$, $0{\cdot}75$ and $\delta_i = \mu_i/10$, $\mu_i/5$, $\mu_i/3$. Therefore in total we consider nine different combinations of mixtures. The mean is modelled by a 6th-degree polynomial in time and the generalised autoregressive parameters and log-innovation variances by cubic polynomials in lag and time. In each random sample there are about 15% of subjects, 54 out of 369, satisfying $m_i < q$ and $m_i < d$; in other words, the number of repeated measures is less than the dimensions of z_{ijk} and z_{ij}.

In Table 2 we report the averages of the parameter estimate for the simulated data from the Normal distribution, together with the averages of the simulated standard deviations in parentheses. The parameter ρ in $W_i = A_i^{\frac{1}{2}} R_i(\rho) A_i^{\frac{1}{2}}$ takes values 0·2, 0·5 and 0·8 so that the effects of misspecification of W_i on the parameter estimators can be measured, where $R_i(\rho)$ are specified with AR(1) structure. Table 2 shows that the resulting parameter estimators are very close to the true values, implying that the proposed approach works well for unbalanced longitudinal data even when some of the m_i are less than q and d. Moreover, the estimators $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\lambda}$ are robust against misspecification of $R_i(\rho)$.

Table 2: *Simulation study. Average of the parameter estimates for* 1000 *random samples generated from the Normal distribution* (*simulated standard errors are in parentheses*), *where* AR(1) *structures are specified for the covariances of* ε_i^2

	True	$\rho = 0{\cdot}2$	$\rho = 0{\cdot}5$	$\rho = 0{\cdot}8$
β_1	875·22	874·74 (14·04)	874·74 (14·06)	874·76 (14·09)
β_2	−207·27	−207·37 (11·93)	−207·37 (11·94)	−207·38 (11·97)
β_3	−22·48	−22·31 (7·32)	−22·31 (7·32)	−22·31 (7·34)
β_4	32·21	32·15 (4·68)	32·15 (4·69)	32·16 (4·70)
β_5	−1·31	−1·32 (0·79)	−1·32 (0·79)	−1·32 (0·79)
β_6	−1·92	−1·91 (0·47)	−1·91 (0·47)	−1·91 (0·47)
β_7	0·25	0·25 (0·06)	0·25 (0·06)	0·25 (0·06)
γ_1	0·62	0·62 (0·04)	0·62 (0·04)	0·62 (0·04)
γ_2	−0·51	−0·51 (0·07)	−0·51 (0·07)	−0·51 (0·07)
γ_3	0·15	0·15 (0·03)	0·15 (0·03)	0·15 (0·03)
γ_4	−0·02	−0·02 ($3{\cdot}59 \times 10^{-3}$)	−0·02 ($3{\cdot}59 \times 10^{-3}$)	−0·02 ($3{\cdot}60 \times 10^{-3}$)
λ_1	11·55	11·55 (0·04)	11·55 (0·05)	11·55 (0·06)
λ_2	−0·34	−0·34 (0·03)	−0·34 (0·03)	−0·34 (0·04)
λ_3	−0·05	−0·05 (0·01)	−0·05 (0·01)	−0·05 (0·02)
λ_4	0·02	$1{\cdot}98 \times 10^{-2}$ ($3{\cdot}49 \times 10^{-3}$)	$1{\cdot}98 \times 10^{-2}$ ($3{\cdot}81 \times 10^{-3}$)	$1{\cdot}98 \times 10^{-2}$ ($4{\cdot}51 \times 10^{-3}$)

Table 3 gives the simulated efficiencies of the estimators $\hat{\beta}$ relative to the conventional approach. The same working correlation structure, either compound symmetry or AR(1), is specified for y_i and ε_i^2. We also look at efficiency on different degrees of the within-subject correlation by choosing different values of ρ. The relative efficiency is of the same form as (14) but the variance matrix of $\hat{\beta}$ has the more accurate form

$$\operatorname{var}(\hat{\beta}) = \left(\sum_{i=1}^{n} X_i' \hat{\Sigma}_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i' \hat{\Sigma}_i^{-1} \Sigma_i \hat{\Sigma}_i^{-1} X_i\right) \left(\sum_{i=1}^{n} X_i' \hat{\Sigma}_i^{-1} X_i\right)^{-1}, \tag{15}$$

Table 3: *Simulation study. Average of relative efficiency of the mean parameters β_k for* 1000 *random samples from the Normal distribution, with compound symmetry and* AR(1) *structures being specified for the covariances of y_i and ε_i^2*

	Compound symmetry Value of ρ			AR(1) model Value of ρ		
	0·2	0·5	0·8	0·2	0·5	0·8
$e(\hat{\beta}_1)$	1·56	1·93	2·63	1·51	1·54	1·91
$e(\hat{\beta}_2)$	1·82	1·92	2·05	1·77	1·90	2·35
$e(\hat{\beta}_3)$	1·67	1·82	1·93	1·57	1·75	2·32
$e(\hat{\beta}_4)$	1·66	1·76	1·82	1·57	1·73	2·06
$e(\hat{\beta}_5)$	1·96	2·11	2·21	1·76	1·92	2·35
$e(\hat{\beta}_6)$	1·71	1·83	1·90	1·57	1·69	1·97
$e(\hat{\beta}_7)$	2·08	2·21	2·28	1·83	1·90	2·17

where Σ_i are the true covariance matrices, and $\hat{\beta}$ and $\hat{\Sigma}_i$ are the estimators of β and Σ_i, obtained by either the conventional or our proposed approach. Table 3 confirms the conclusion in § 5, that the mean-covariance modelling strategy leads to improvement.

Since, for the Normal mixture distribution $F_{m_i} = \pi N_{m_i}(\mu_i + \delta_i, \Sigma_i) + (1-\pi)N_{m_i}(\mu_i, \Sigma_i)$, the expectation and variance are $\tilde{\mu}_i = \mu_i + \pi\delta_i$ and $\tilde{\Sigma}_i = \Sigma_i + \pi(1-\pi)\delta_i\delta_i'$, it is not appropriate to compare directly the parameter estimators $\hat{\beta}$ in $\tilde{\mu}_i$ and $(\hat{\gamma}, \hat{\lambda})$ in $\tilde{\Sigma}_i$ to the true values of β in μ_i and (γ, λ) in Σ_i unless $\pi = 0$ or $\delta_i = 0$. We therefore compare the estimated mean $\hat{\mu}_i$ and covariance matrix $\hat{\Sigma}_i$ to the true values of $\tilde{\mu}_i$ and $\tilde{\Sigma}_i$. For example, relative errors defined by $\mathrm{err}(\hat{\mu}_i) \equiv \|\hat{\mu}_i - \tilde{\mu}_i\|/\|\tilde{\mu}_i\|$ and $\mathrm{err}(\hat{\Sigma}_i) \equiv \|\hat{\Sigma}_i - \tilde{\Sigma}_i\|/\|\tilde{\Sigma}_i\|$ can be used, where $\|.\|$ denotes the Euclidean norm. Table 4 gives the averages of relative errors for each combination of π and δ_i, showing that the resulting mean estimators are very close to the true values in all cases. The relative error of the covariance matrix estimators, $\mathrm{err}(\hat{\Sigma}_i)$, seems to increase with the mean-shift parameter δ_i but changes little when the mixing weight π varies. In some cases $\mathrm{err}(\hat{\Sigma}_i)$ almost reaches 20% but we regard this as acceptable

Table 4: *Simulation study. Average of relative errors* $\mathrm{err}(\hat{\mu}) = \sum_{i=1}^{n} \mathrm{err}(\hat{\mu}_i)/n$ *and* $\mathrm{err}(\hat{\Sigma}) = \sum_{i=1}^{n} \mathrm{err}(\hat{\Sigma}_i)/n$ *for* 1000 *random samples from the Normal Mixture distribution, with* AR(1) *structure and* $\rho = 0{\cdot}5$ *being specified for the covariances of ε_i^2*

(π, δ_i)	$\mathrm{err}(\hat{\mu})$	$\mathrm{err}(\hat{\Sigma})$
$(0{\cdot}25, \mu_i/10)$	$5{\cdot}90 \times 10^{-3}$	0·02
$(0{\cdot}25, \mu_i/5)$	$1{\cdot}08 \times 10^{-2}$	0·06
$(0{\cdot}25, \mu_i/3)$	$1{\cdot}60 \times 10^{-2}$	0·14
$(0{\cdot}50, \mu_i/10)$	$6{\cdot}11 \times 10^{-3}$	0·02
$(0{\cdot}50, \mu_i/5)$	$1{\cdot}13 \times 10^{-2}$	0·08
$(0{\cdot}50, \mu_i/3)$	$1{\cdot}58 \times 10^{-2}$	0·19
$(0{\cdot}75, \mu_i/10)$	$4{\cdot}40 \times 10^{-3}$	0·02
$(0{\cdot}75, \mu_i/5)$	$6{\cdot}61 \times 10^{-3}$	0·07
$(0{\cdot}75, \mu_i/3)$	$8{\cdot}15 \times 10^{-3}$	0·16

because $\|\hat{\Sigma}_i - \tilde{\Sigma}_i\|$ is based on matrices as large as 12×12. Therefore, the approach behaves well even for unbalanced Normal mixture data.

The relative efficiencies associated with the mean parameters are very similar to those for the Normal case. We therefore conclude that the approach produces efficient estimators of the mean parameters even for nonnormally distributed data. Finally, the quasi-score test statistics in § 4 are used to test the null hypotheses $H_0: \gamma = 0$ and $\lambda = 0$, $H_0: \gamma = 0$ and $H_0: \lambda = 0$. We find that their empirical powers are very close to 1 and their empirical sizes are equal to 0·06, 0·04 and 0·03, respectively, showing that the proposed quasi-score test approach works reasonably well.

7. Discussion

We also studied the balanced case by analysing Kenward's cattle data and conducted the relevant simulation study. The proposed approach also works well in this case; to save space the details are not reported here. On the other hand, when longitudinal data are balanced the degree triple (p, q, d) may be determined by observing the regressograms (Pourahmadi, 1999). In general, it can be determined as the triple that optimises a criterion such as AIC or BIC when the distribution of the responses is known (Pan & MacKenzie, 2003).

Though the methodology works well for a continuous nonnormal distribution under some moment conditions, its applicability to data from discrete distributions requires more attention. Note that the generalised autoregressive parameters and log of innovation variances are generally constrained when Σ_i is structured and this happens for binary and other discrete data. Thus, the use of the last two models in (2) requires a careful choice of z_{ijk} and z_{ij} in such cases.

Acknowledgement

We would like to thank the editor and referees for their very constructive suggestions. The work was supported by a grant from the Engineering and Physical Sciences Research Council and an Overseas Research Scholarship of the U.K.

Appendix

Proofs of theorems

Regularity conditions.

Condition A1. We assume that the dimensions p, q and d of the covariates x_{ij}, z_{ijk} and z_{ij} are fixed, and that the numbers m_i of repeated measurements are fixed. We also assume that the first four moments of the responses exist.

Condition A2. The parameter space Θ is a compact subset of R^{p+q+d}, and the true parameter value θ_0 is in the interior of the parameter space Θ.

Condition A3. The covariates x_{ij}, z_{ijk} and z_{ij}, the vectors

$$\dot{g}^{-1}(X_i\beta) = (\dot{g}^{-1}(x'_{i1}\beta), \ldots, \dot{g}^{-1}(x'_{im_i}\beta))'$$

and the matrices W_i^{-1} are all bounded, meaning that all the elements of the vectors and matrices are bounded.

Proof of Theorem 1. For illustration we only give the proof that $\hat{\beta}_n \to \beta_0$ almost surely. The proofs for $\hat{\gamma}_n$ and $\hat{\lambda}_n$ are similar. According to McCullagh (1983) we have

$$\hat{\beta}_n - \beta_0 = \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial \mu_i'}{\partial \beta} \right) \Sigma_i^{-1} \left(\frac{\partial \mu_i'}{\partial \beta} \right)' \right\}_{\beta_0}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial \mu_i'}{\partial \beta} \right) \Sigma_i^{-1} (y_i - \mu_i) \right\}_{\beta_0} + o_p(n^{-1/2}). \tag{A1}$$

On the other hand, the expectation and variance matrix of $U_i \equiv (\partial \mu_i'/\partial \beta)\Sigma_i^{-1}(y_i - \mu_i)$ at $\beta = \beta_0$ are given by $E_0(U_i) = 0$ and

$$\mathrm{var}_0(U_i) = \left\{ \left(\frac{\partial \mu_i'}{\partial \beta} \right) \Sigma_i^{-1} \left(\frac{\partial \mu_i'}{\partial \beta} \right)' \right\}_{\beta_0} = (G_i^0 X_i)' \Sigma_i^{-1} (G_i^0 X_i), \tag{A2}$$

where $G_i^0 = \mathrm{diag}(\dot{g}^{-1}(x_{i1}'\beta_0), \ldots, \dot{g}^{-1}(x_{im_i}'\beta_i))$ is an $(m_i \times m_i)$ diagonal matrix. Since $\Sigma_i^{-1} = T_i' D_i^{-1} T_i$ the variance matrix in (A2) can be further written as $\mathrm{var}_0(U_i) = (T_i G_i^0 X_i)' D_i^{-1} (T_i G_i^0 X_i)$. Condition A3 above implies that there exists a constant κ_0 such that $\mathrm{var}_0(U_i) \leqslant \kappa_0 1_{p \times p}$ for all i and all $\theta \in \Theta$, where $1_{p \times p}$ is the $(p \times p)$ matrix with all elements being 1's, meaning that all elements of $\mathrm{var}_0(U_i)$ are bounded by κ_0. Thus $\sum_{i=1}^{\infty} \mathrm{var}_0(U_i)/i^2 < \infty$. By Kolmogorov's strong law of large numbers we know that

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial \mu_i'}{\partial \beta} \right) \Sigma_i^{-1} (y_i - \mu_i) \right\}_{\beta_0} \to 0 \tag{A3}$$

almost surely as $n \to \infty$. In the same manner it can be shown that

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial \mu_i'}{\partial \beta} \right) \Sigma_i^{-1} \left(\frac{\partial \mu_i'}{\partial \beta} \right) \right\}_{\beta_0}$$

in (A1) is a bounded matrix. Application of (A3) to (A1) leads to $\hat{\beta}_n - \beta_0 \to 0$ almost surely as $n \to \infty$. The proof is complete. □

Proof of Theorem 2. First, it can be shown that under Conditions A1–A3 the following necessary conditions for asymptotic Normality hold.

Condition A4. The equations in (4) and their derivatives with respect to β, γ and λ exist.

Condition A5. The expectations of the equations in (4) are equal to zero.

Condition A6. The information matrices satisfy

$$E\left[\left\{ \left(\frac{\partial \mu_i'}{\partial \beta} \right) \Sigma_i^{-1} (y_i - \mu_i) \right\} \left\{ \left(\frac{\partial \mu_i'}{\partial \beta} \right) \Sigma_i^{-1} (y_i - \mu_i) \right\}' \right] = -E\left[\frac{\partial}{\partial \beta} \left\{ \left(\frac{\partial \mu_i'}{\partial \beta} \right) \Sigma_i^{-1} (y_i - \mu_i) \right\}' \right],$$

$$E\left[\left\{ \left(\frac{\partial \hat{r}_i'}{\partial \gamma} \right) D_i^{-1} (r_i - \hat{r}_i) \right\} \left\{ \left(\frac{\partial \hat{r}_i'}{\partial \gamma} \right) D_i^{-1} (r_i - \hat{r}_i) \right\}' \right] = -E\left[\frac{\partial}{\partial \gamma} \left\{ \left(\frac{\partial \hat{r}_i'}{\partial \gamma} \right) D_i^{-1} (r_i - \hat{r}_i) \right\}' \right],$$

$$E\left[\left\{ \left(\frac{\partial \sigma_i^{2'}}{\partial \lambda} \right) W_i^{-1} (\varepsilon_i^2 - \sigma_i^2) \right\} \left\{ \left(\frac{\partial \sigma_i^{2'}}{\partial \lambda} \right) W_i^{-1} (\varepsilon_i^2 - \sigma_i^2) \right\}' \right] = -E\left[\frac{\partial}{\partial \lambda} \left\{ \left(\frac{\partial \sigma_i^{2'}}{\partial \lambda} \right) W_i^{-1} (\varepsilon_i^2 - \sigma_i^2) \right\}' \right].$$

Condition A7. Recall that $\theta = (\beta', \gamma', \lambda')'$ and $S(\theta) = (S_1'(\beta), S_2'(\gamma), S_3'(\lambda))'$. As $n \to \infty$ we have that

$$\frac{1}{n} \frac{\partial S'(\theta)}{\partial \theta} - \frac{1}{n} E_0 \left\{ \frac{\partial S'(\theta)}{\partial \theta} \right\} \to 0$$

almost surely, where $E_0\{.\}$ stands for the expectation evaluated at the true value $\theta = \theta_0$.

Condition A8. The following asymptotic result holds at $\theta = \theta_0$:

$$\frac{1}{\sqrt{n}}\begin{Bmatrix} S_1(\beta) \\ S_2(\gamma) \\ S_3(\lambda) \end{Bmatrix} \to N\left\{0, \begin{pmatrix} v^{11} & v^{12} & v^{13} \\ v^{21} & v^{22} & v^{23} \\ v^{31} & v^{32} & v^{33} \end{pmatrix}\right\}$$

in distribution as $n \to 0$, where the asymptotic variance matrix is positive definite.

Conditions A4–A6 are straightforward under Conditions A1–A3. Condition A7 can be obtained in a similar manner to the proof of Theorem 1. Below we give a proof of Condition A8. In fact, Condition A3 implies that

$$E_0\left[\left|\psi'\left\{\left(\frac{\partial\mu_i'}{\partial\beta}\right)\Sigma_i^{-1}(y_i-\mu_i)\right\} + \omega'\left\{\left(\frac{\partial\hat{r}_i'}{\partial\gamma}\right)D_i^{-1}(r_i-\hat{r}_i)\right\} + \varphi'\left\{\left(\frac{\partial\sigma_i^{2'}}{\partial\lambda}\right)W_i^{-1}(\varepsilon_i^2-\sigma_i^2)\right\}\right|^3\right] \leqslant \kappa$$

for any $\psi \in R^p$, $\omega \in R^q$ and $\varphi \in R^d$, where κ is a constant independent of i. Moreover, at $\theta = \theta_0$ we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} V_0\left[\psi'\left\{\left(\frac{\partial\mu_i'}{\partial\beta}\right)\Sigma_i^{-1}(y_i-\mu_i)\right\} + \omega'\left\{\left(\frac{\partial\hat{r}_i'}{\partial\gamma}\right)D_i^{-1}(r_i-\hat{r}_i)\right\} + \varphi'\left\{\left(\frac{\partial\sigma_i^{2'}}{\partial\lambda}\right)W_i^{-1}(\varepsilon_i^2-\sigma_i^2)\right\}\right] \\ = (\psi', \omega', \varphi')\begin{pmatrix} v_n^{11} & v_n^{12} & v_n^{13} \\ v_n^{21} & v_n^{22} & v_n^{23} \\ v_n^{31} & v_n^{32} & v_n^{33} \end{pmatrix}\begin{pmatrix}\psi \\ \omega \\ \varphi\end{pmatrix} \\ \to (\psi', \omega', \varphi')\begin{pmatrix} v^{11} & v^{12} & v^{13} \\ v^{21} & v^{22} & v^{23} \\ v^{31} & v^{32} & v^{33} \end{pmatrix}\begin{pmatrix}\psi \\ \omega \\ \varphi\end{pmatrix} > 0,\end{aligned}$$

because of the positive definiteness of V in (11), where $V_0[.]$ denotes the variance matrix evaluated at the true value $\theta = \theta_0$. Therefore, that Condition A8 holds follows from the Liapounov form of the multivariate central limit theorem.

Next, we give a proof of the asymptotic Normality of the generalised estimating equation estimators $(\hat{\beta}_n', \hat{\gamma}_n', \hat{\lambda}_n')'$. The proof is similar to that in Chiu et al. (1996) and below we only highlight the points of difference. Note that $\hat{\theta}_n = (\hat{\beta}_n', \hat{\gamma}_n', \hat{\lambda}_n')'$ is a consistent sequence of roots to the equations so that $S(\hat{\theta}_n) = 0$. In a neighbourhood of $\theta_0 = (\beta_0', \gamma_0', \lambda_0')'$, the true value of θ, we have that

$$\frac{1}{\sqrt{n}}\begin{Bmatrix} S_1(\beta) \\ S_2(\gamma) \\ S_3(\lambda) \end{Bmatrix}_{(\beta_0,\gamma_0,\lambda_0)} = -\Omega_n\left\{\sqrt{n}\begin{pmatrix}\hat{\beta}_n - \beta_0 \\ \hat{\gamma}_n - \gamma_0 \\ \hat{\lambda}_n - \lambda_0\end{pmatrix}\right\}, \tag{A4}$$

where

$$\Omega_n = \int_0^1 \frac{1}{n}\left\{\frac{\partial S'(\theta)}{\partial\theta}\right\}_{\theta_\zeta} d\zeta \tag{A5}$$

with $\theta_\zeta = (\beta_\zeta', \gamma_\zeta', \lambda_\zeta')' = (\beta_0' + \zeta(\hat{\beta}_n - \beta_0)', \gamma_0' + \zeta(\hat{\gamma}_n - \gamma_0)', \lambda_0' + \zeta(\hat{\lambda}_n - \lambda_0))'$. Using Conditions A5–A7 we obtain

$$-\Omega_n \to \begin{pmatrix} v^{11} & 0 & 0 \\ 0 & v^{22} & 0 \\ 0 & 0 & v^{33} \end{pmatrix}_{(\beta_0,\gamma_0,\lambda_0)}.$$

We apply Condition A8 to (A4) and (A5) and then obtain the conclusion stated in Theorem 2. The proof is complete. □

References

Carey, V. C., Zeger, S. L. & Diggle, P. J. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika* **80**, 517–26.

Chiu, T. Y. M., Leonard, T. & Tsui, K. W. (1996). The matrix-logarithm covariance model. *J. Am. Statist. Assoc.* **91**, 198–210.

Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Daniels, M. & Zhao, Y. (2003). Modelling the random effects covariance matrix in longitudinal data. *Statist. Med.* **22**, 1631–47.

Diggle, P. J. & Verbyla, A. P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* **54**, 403–15.

Diggle, P. J., Heagerty, P., Liang, K. Y. & Zeger, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. Oxford University Press.

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

Liang, K. Y., Zeger, S. L. & Qaqish, B. (1992). Multivariate regression analysis for categorical data (with Discussion). *J. R. Statist. Soc.* B **54**, 3–40.

McCullagh, P. (1983). Quasi-likelihood functions. *Ann. Statist.* **11**, 59–67.

Pan, J. & Mackenzie, G. (2003). Model selection for joint mean-covariance structures in longitudinal studies. *Biometrika* **90**, 239–44.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika* **86**, 677–90.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **87**, 425–35.

Prentice, R. L. & Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47**, 825–39.

Wang, Y. G. & Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. *Biometrika* **90**, 29–41.

Wang, Y. G. & Carey, V. (2004). Unbiased estimating equations from working correlation models for irregularly timed repeated measures. *J. Am. Statist. Assoc.* **99**, 845–53.

[*Received March* 2005. *Revised February* 2006]