



University of Bahrain
**Journal of the Association of Arab Universities for
Basic and Applied Sciences**

www.elsevier.com/locate/jaoubas
www.sciencedirect.com



تقدير جاوس للانحدار والارتباط في البيانات الطولية

محمد الرواش¹، محسن بوراحمدي²

¹ دائرة الرياضيات، جامعة الشارقة، الشارقة، الإمارات العربية المتحدة

² دائرة الإحصاء، جامعة تكساس A&M، تكساس، الولايات المتحدة الأمريكية

المخلص:

يعتبر تقدير العزوم للارتباط باستخدام معادلات التقدير العامة (GEE) غير مضمون النتائج في بعض الأحيان أو غير ممكن أحيانا أخرى. في هذه الدراسة، تم تقديم طريقة تقدير جاوس وبيان أن تقدير الارتباط في حال البيانات الطولية يكون بشكل مقارب غير منحاز وممكن. كما تم اشتقاق خصائص الانحدار والارتباط في حالة العينات الكبيرة، وكما تم توضيح هذه التقديرات باستخدام بيانات واقعية.



University of Bahrain
**Journal of the Association of Arab Universities for
 Basic and Applied Sciences**

www.elsevier.com/locate/jaaubas
 www.sciencedirect.com



ORIGINAL ARTICLE

Gaussian estimation of regression and correlation parameters in longitudinal data

M. Al-Rawwash ^{a,*}, M. Pourahmadi ^b

^a Department of Mathematics, University of Sharjah, Sharjah, United Arab Emirates

^b Department of Statistics, Texas A&M University, College Station, TX, USA

Available online 10 October 2012

KEYWORDS

Covariance matrix;
 Generalized least squares;
 Gaussian estimation;
 Generalized estimating equations;
 Longitudinal data

Abstract The moment estimate of the correlation parameters using generalized estimating equations (GEE) is not guaranteed to exist or to be feasible. We introduce the Gaussian estimation method and show that the estimate of the correlation parameters in longitudinal data setup is asymptotically unbiased and feasible. We derive the large sample properties of the regression and correlation estimates and we illustrate these estimators via real life data example.

© 2012 University of Bahrain. Production and hosting by Elsevier B.V. All rights reserved.

1. Introduction

Longitudinal or repeated measures data arise in several areas of research including social, environmental, medical, biological sciences where many experimental units (human, animals, plants, schools, etc.) are measured repeatedly over time. The broad goal of such studies is to assess changes over time within units and relate them to the treatments, interventions and other relevant covariates using the idea of regression. The difficulty in exact specification of the distribution of nonnormal data (binary, count, etc.) precludes the use of maximum likelihood for parameter estimation. However, the idea of quasi-likelihood (QL) for independent observations (Wedderburn, 1974 and Heyde, 1997) requires only a model for the mean and the relationship between the mean and variance. Liang and Zeger (1986) generalized estimating equations (GEE)

method extended the framework of QL estimation to dependent longitudinal data by specifying models for the marginal means and variances. They relied on a ‘working’ correlation matrix with few parameters (say, α) to write an analog of the generalized least squares normal equations. Some pitfalls and limitations of the GEE estimator of α were pointed out by Crowder (2001) and other researchers. For example, the moment estimator of the correlation parameter may not exist in some simple cases or it may not be feasible. Despite the list of drawbacks of GEE, researchers insisted on the fact that GEE usually works well in practice even in the misspecified structures (Crowder, 2001). Several articles proposed alternative approaches based on minimizing objective functions instead of solving some ad hoc estimating equations to obtain efficient estimators of β and α (Chaganty, 1997; Chaganty and Shults, 1999; Al-Rawwash, 2001; Crowder, 2001; Al-Rawwash and Pourahmadi, 2006).

Chaganty (1997) relied on partial minimization of the generalized least squares criterion to introduce the quasi-least squares (QLS) estimates where the covariance matrix is a function of both regression and correlation parameters (β , α). The estimating equation for β is precisely the same as GEE, but the estimating equation for α is new and different from all previously developed approaches (see Diggle et al., 1994). Shults and Chaganty (1998) used the QLS method for certain

* Corresponding author.

E-mail addresses: malrawwash@sharjah.ac.ae (M. Al-Rawwash), pourahm@stat.tamu.edu (M. Pourahmadi).

^{*} Peer review under responsibility of University of Bahrain.



Production and hosting by Elsevier

correlation models and showed the importance of the QLS compared to the GEE for unbalanced longitudinal data setup. The fact that the QLS estimator of the correlation parameter is asymptotically biased motivated Chaganty and Shults (1999) to propose a method to eliminate this problem. They proposed using a continuous, one-to-one transformation that depends on the working correlation matrix under the QLS scheme. They presented the method for different correlation structures and obtained a bias corrected estimate of the correlation matrix. Pourahmadi (2000) developed the maximum likelihood estimators for the parameters of a generalized linear model for the covariance matrix and discussed the asymptotic results under normality condition. Al-Rawwash (2005) introduced a nonparametric approach to estimate the parameters of interest as an alternative to the classical parametric techniques. Al-Rawwash and Pourahmadi (2006) introduced the generalized Gaussian likelihood function as their objective function and they used data-driven approach to obtain estimates of the parameters assuming minimal distributional requirements. In this paper, we follow the footsteps of Al-Rawwash and Pourahmadi (2006) and use the generalized Gaussian likelihood function as the objective function to pursue the estimation of the regression and the correlation parameters.

The history of Gaussian estimation (GE) goes back to Whittle (1961) in the time series literature where the parameter estimates are the values that maximize the Gaussian likelihood function for certain correlation structures. Several articles discussed the GE in longitudinal data literature where they focused on obtaining the parameter estimates based on minimizing an objective function even when the observations are not normally distributed (Crowder, 1995; Crowder, 2001; Al-Rawwash, 2001; Al-Rawwash and Pourahmadi, 2006). In this article, we assume only the knowledge of the first two moments and not the exact specification of the distribution of the data. The notion of Gaussian estimation is distinct from GEE in that it minimizes an objective function rather than relying on estimating equations that might produce flawed estimators (Chaganty, 1997; Al-Rawwash, 2001; Crowder, 2001). Several articles discussed the idea of modeling longitudinal data sets including but not limited to Vonesh and Chinchilli (1997), Lipsitz et al. (2000), Wang and Carey (2003), Pourahmadi et al. (2007) and Buzkova and Lamley (2008).

The organization of the paper is as follows. In Section 2, we introduce the Gaussian estimation and outline a Newton-Raphson iterative method for computing the estimates. Some closed form solutions for the estimates of the correlation parameter α is given in Section 3 when the correlation structure is AR(1). In Section 4, we establish the consistency and asymptotic normality of the regression and the correlation parameter estimates under mild regularity conditions. In Section 5, we provide a real life data analysis of the GE compared to other well known estimation procedures. Finally, we outline the proof of some results in the appendix.

2. The Gaussian estimation

To set the notation, we consider a random vector $Y = (y_1, y_2, \dots, y_n)'$ comprising the measurements taken on a generic subject at times $t = (t_1, t_2, \dots, t_n)$ and the associated covariates $x_j = (x_{j1}, \dots, x_{jp})'$, $j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, m$.

We plan to make no distributional assumption about Y except for its first two moments. Firstly, the means and variances of the entries of the response are related to the covariates x_j via $g(\mu_j) = x_j' \beta$, $\text{var}(y_j) = \phi v(\mu_j)$, where $\beta = (\beta_1, \dots, \beta_p)'$ is the regression parameters of primary interest, g is an invertible known link function, $v(\cdot)$ is a known variance function and ϕ is a dispersion parameter not depending on β . The importance of the parameter ϕ appears clearly when we handle a nonnormal data set for it expresses the variability beyond the mean. The setup may be carried out for unbalanced longitudinal data, however we focus in this article on the balanced case where $n_i = n$. Secondly, we use the matrix $R = R(\alpha)$ to model correlations among the measurements on the same subject. The matrix R and its parameters α are usually viewed as nuisance, though in some situations they are of primary interest. Now, if we define $\mu = \mu(\beta) = (\mu_1, \dots, \mu_n)'$ and $A(\beta) = \text{diag}(v(\mu_1), \dots, v(\mu_n))$, then the covariance matrix of Y is decomposed into

$$\Sigma = \phi A^{1/2}(\beta) R(\alpha) A(\beta)^{1/2}.$$

In longitudinal data setup, we assume Y_i , $i = 1, 2, \dots, m$, to be the vector of repeated measures on the i th subject with covariance $\Sigma_i = \phi A_i^{1/2} R(\alpha) A_i^{1/2}$. The Gaussian estimates (Al-Rawwash and Pourahmadi, 2006) of (β, α) are the minimizers of the objective function

$$G(\beta, \alpha) = \sum_{i=1}^m [\log |\Sigma_i| + Z_i' R^{-1}(\alpha) Z_i], \quad (1)$$

where $Z_i = A_i^{-1/2}(\beta)(Y_i - \mu_i(\beta))$ is the vector of partially standardized measurement on the i th subject. Even if $G(\cdot, \cdot)$ is differentiable, its differentiation with respect to β is complicated because of the appearance of β in $A_i(\beta)$. This problem is avoided by pretending that $A_i^{-1/2}$ is a function of β^* different from β (sometimes called decoupling). Therefore, if we differentiate (1) with respect to β, α and then substituting β for β^* we arrive at:

$$D'(\beta) A^{-1/2}(\beta) \tilde{R}^{-1}(\alpha) Z = 0, \quad (2)$$

$$\frac{\partial}{\partial \alpha} \log |\tilde{R}(\alpha)| + Z' \frac{\partial \tilde{R}^{-1}(\alpha)}{\partial \alpha} Z = 0. \quad (3)$$

where $D(\beta) = \text{diag}(D_1, D_2, \dots, D_m)$, $D_i = D_i(\beta) = \frac{\partial \mu_i}{\partial \beta}$ for $1 \leq i \leq m$, $\tilde{R} = I_m \otimes R(\alpha)$, $A(\beta) = \text{diag}(A_1(\beta), A_2(\beta), \dots, A_m(\beta))$, $Z = (Z_1', \dots, Z_m')'$ and \otimes is the kronecker product. The estimating Eqs. (2) are precisely the GEEs for β , but (3) is different from the estimating equations developed in the literature to estimate the correlation parameter. Solving these two equations for β and α could give the GE of (β, α) . Since closed form solution is not available in general, we suggest using the Newton-Raphson iterative algorithm as follows. Firstly, we choose an initial value $\hat{\beta}$ for β , then we evaluate the following quantities: $\tilde{A}_i = A_i(\hat{\beta})$, $\tilde{\mu}_i = \mu_i(\hat{\beta})$, $\tilde{Z}_i = \tilde{A}_i^{-1/2} (Y_i - \tilde{\mu}_i)$ and $\tilde{D}_i = D_i(\hat{\beta})$ at $\hat{\beta}$. Accordingly, we solve Eq. (3) for $\tilde{\alpha}$, then we compute $\tilde{R} = R(\tilde{\alpha})$ and $\tilde{\Sigma}_i = \tilde{A}_i^{1/2} \tilde{R} \tilde{A}_i^{1/2}$ for $1 \leq i \leq m$ and construct the covariance matrix $\tilde{\Sigma} = \text{diag}(\tilde{\Sigma}_1, \tilde{\Sigma}_2, \dots, \tilde{\Sigma}_m)$ as well as $D(\tilde{\mu})$ and $\mu(\tilde{\beta})$. Hence, we update the value of $\hat{\beta}$ using $\hat{\beta} = \hat{\beta} + (\tilde{D}' \tilde{\Sigma}^{-1} \tilde{D})^{-1} \tilde{D}' \tilde{\Sigma}^{-1} (Y - \tilde{\mu})$. Finally, we repeat the process until $\hat{\beta} \simeq \tilde{\beta}$, then we take $\hat{\beta}$ as an estimate of β and $\tilde{\alpha}$ as an estimate of α . The least squares estimates $\hat{\beta} = (X'X)^{-1} X'Y$ of β are a good candidate for the initial value of β , while estimates of

α must be chosen so that the positive definiteness of the correlation matrix is guaranteed. The estimate of ϕ depends on the GE results of β and α obtained using Newton-Raphson iterative method which is reduced to $\hat{\phi} = \frac{1}{m} \sum_{i=1}^m \widehat{Z}_i' \widehat{Z}_i$. The unbiasedness of the GE of (β, α) is not guaranteed when the working correlation structure is misspecified. However, we show in Section 4 that the GE of β and α is consistent when the working correlation matrix is correctly specified.

3. Specific correlation structures

In this section, we present closed form solutions for the correlation parameter of AR(1) correlation structure when the correlation matrix is either misspecified or correctly specified. These can be used in Newton-Raphson algorithm and hence avoid an iterative cycle within this algorithm. The impact of the correlation matrix misspecification on the degrees of estimating equations and the form of their solutions are illuminated. Chaganty (1997) and Chaganty and Shults (1999) considered special correlation structures and they derived closed form solutions for the estimator of the correlation parameter using the the QLS and the C-QLS. The asymptotic bias of the QLS estimator of the correlation parameter is corrected by Chaganty and Shults (1999) using a continuous one-to-one function that transformed the biased correlation matrix to an asymptotically unbiased one. We show that GE of α is asymptotically unbiased and feasible when the number of measurements per subject n is greater than 3 or a slightly modified (conditional) form of the Gaussian estimation is used.

In the following example we assume that the true correlation matrix, denoted by \bar{R} , is unstructured, and we intend to give explicit formulas for the estimate of α by solving (3) for the AR(1) 'working' correlation matrices.

Example 1. Let the working correlation structure of the repeated measurements be the AR(1) matrix $R(\rho) = (\rho^{|i-j|})$ with the parameter $\alpha = \rho \in (-1, 1)$. Note that

$$R^{-1}(\rho) = \frac{1}{(1-\rho^2)} [I_n + \rho^2 C_2 - \rho C_1],$$

where C_1 is a tridiagonal matrix with 0 on the diagonal and 1 on the first upper and lower diagonals and $C_2 = \text{diag}(0, 1, \dots, 1, 0)$. Since the entries of $(1-\rho^2)R^{-1}$ are quadratic functions of ρ , the estimating Eq. (3) reduces to the quadratic equation

$$f(\rho) = a_m \rho^2 - 2b_m \rho + a_m = 0,$$

where $a_m = \sum_{i=1}^m Z_i'(\hat{\beta}) C_1 Z_i(\hat{\beta}) - \text{mtr}(C_1 \bar{R})$, $b_m = \sum_{i=1}^m Z_i'(\hat{\beta}) (I_n + C_2) Z_i(\hat{\beta}) - 2m(n-1)$, $\hat{\beta}$ is the current estimate of β and Z_i is defined in (1). Its (unique) feasible solution is given by

$$\tilde{\rho} = \frac{b_m - \sqrt{b_m^2 - a_m^2}}{a_m} = \frac{b_m}{a_m} - \sqrt{\left(\frac{b_m}{a_m}\right)^2 - 1},$$

provided that $|b_m| > |a_m|$. This is in sharp contrast to the QLS estimate of the correlation parameter which is known to be feasible all the time (Chaganty, 1997). The differences between the coefficients of the quadratic equations for GE and QLS estimates of α are in the second terms in a_m and b_m , for which with $\bar{R} = (\rho_{i,j})$ we have $|\text{mtr}(C_1 \bar{R})| = |2m \sum_{i=1}^{n-1} \rho_{i,i+1}| \leq 2m(n-1)$. Nevertheless, there are datasets for which $|b_m| < |a_m|$ and the quadratic equation have complex roots.

It is important to mention that the working and the true correlation matrices may have the same AR(1) structure which consequently makes $\text{corr}(Y_i)$ correctly specified. In this case, the estimating Eq. (3) is cubic in ρ , therefore we shall derive closed formulas for its solutions and study their feasibility.

Example 2. Let both the working and true correlation matrices have AR(1) structures with $R(\rho)$ and $R^{-1}(\rho)$ given in Example 1. Then, (3) reduces to the cubic equation

$$f(\rho) = \rho^3 - a_1 \rho^2 + b_1 \rho - a_1 = 0, \quad (4)$$

where $a_1 = \frac{d_{12}}{m(n-1)}$, $b_1 = \left(\frac{d_{11}+d_{22}}{m(n-1)} - 1\right)$ and $d_{rs} = \sum_{i=1}^m \sum_{j=s}^{n-r+1} z_{ij} z_{i,j+r-s}$ with $Z_i' = (z_{i1}, z_{i2}, \dots, z_{in})$ is defined in (1). To solve the cubic Eq. (4), we follow the standard procedure and set $\Delta = \frac{v^2}{4} + \frac{u^3}{27}$, where

$$u = \frac{1}{3}(3b_1 - a_1^2), \quad v = \frac{1}{27}(2a_1^3 + 9a_1 b_1 - 27a_1),$$

and for later use define

$$\gamma = \arctan \sqrt[3]{\tan \gamma_1}, \quad \gamma_1 = \frac{1}{2} \text{arccotn} \left(\mp \sqrt{\frac{27v^2}{4u^3}} \right),$$

where the upper sign in γ_1 is used if v is positive and the lower sign if v is negative. The nature and number of feasible roots of (4) depend on the signs of Δ and u as discussed below:

Case 1: If $\Delta > 0$ and $u > 0$, then (4) has only one real feasible root, given by

$$\tilde{\rho} = 2\sqrt[3]{\frac{u}{3} \cotn 2\gamma} + \frac{a_1}{3}.$$

Case 2: If $\Delta < 0$ and $u < 0$, then (4) has three real roots given by

$$\tilde{\rho}_k = 2\sqrt[3]{-\frac{u}{3} \cos\left(\frac{\gamma}{3} + \frac{2(k-1)\pi}{3}\right)} + \frac{a_1}{3}, \quad \text{for } k = 1, 2 \text{ and } 3,$$

where $\cos \gamma = \mp \sqrt{-\frac{27v^2}{4u^3}}$, the upper sign in γ is used if v is positive, otherwise, we use the lower sign. Notice that $\tilde{\rho}_1 < \tilde{\rho}_2 < \tilde{\rho}_3$, but the number of feasible roots depends on the data and cannot be decided in advance.

Case 3: If $\Delta = 0$ and $u < 0$, then (4) has the unique feasible closed form solution:

$$\tilde{\rho} = \mp 2\sqrt[3]{-\frac{u}{3} + \frac{a_1}{3}},$$

where the upper sign in $\tilde{\rho}$ is used if v is positive, otherwise we use the lower sign.

3.1. Unique Feasible Solutions

In this section we show that when the number of subjects m is large or when a modified (conditional) Gaussian estimate is used, then a unique feasible correlation parameter estimate satisfying (4) exists.

A number of simulations in Al-Rawwash (2001) corresponding to different values of the triplet (m, n, ρ) revealed that the cubic Eq. (4) has a unique feasible solution for $m \geq 3$. This provided a clue that for m large (4) must have a unique feasible solution in $(-1, 1)$, which we show next. From (4) and

definitions of a_1 and b_1 , it follows that $f(-1) < 0$ and $f(1) > 0$. Thus, if we show that $f(\rho)$ is monotone in $(-1,1)$ when $m \rightarrow \infty$, then existence of a unique feasible root follows. A sufficient condition for $f(\rho)$ to be monotone is for its derivative $f'(\rho) = 3\rho^2 - 2a_1\rho + b_1$ to have at most one real root. But, this happens if and only if its discriminant $\Delta_m = 4a_1^2 - 12b_1$ is nonpositive. Now, applying the law of large numbers (as $m \rightarrow \infty$) to a_1 and b_1 , it follows that

$$\lim_{m \rightarrow \infty} \Delta_m = 4\rho^2 - 12 \left[\frac{2(n-2)}{n-1} - 1 \right],$$

and the latter is nonpositive if and only if $n \geq 4$.

Another situation that gives rise to a unique feasible correlation parameter estimate for the AR(1) 'working' correlation structure is obtained by replacing the objective function (1) by that corresponding to the conditional Gaussian likelihood function.

4. Large Sample Properties

This section presents the large sample properties of the Gaussian estimates $\hat{\beta}$ and $\hat{\alpha}$. We establish their consistency and asymptotic normality as $m \rightarrow \infty$, under certain regularity conditions. Chaganty (1997) showed that the QLS estimate $\hat{\beta}_m$ is consistent, but $\hat{\alpha}_m$ is asymptotically biased and the joint distribution of $(\hat{\beta}_m, \hat{\alpha}_m)$ is asymptotically normal. A corrected version of the QLS to eliminate the bias problem is proposed in Chaganty and Shults (1999) and they considered markov and generalized markov structures to model the unbalanced correlation matrix cases. Thus, compared to the QLS estimate of α , our proposed GE of the correlation parameter is asymptotically unbiased and we intend to compare our results to the QLS as well as the C-QLS estimates. Our proofs are in the spirit of Chaganty (1997); and Sen and Singer (1993, p. 206). Crowder (2001) has presented an alternative proof establishing consistency of the GE, but only asymptotic normality of $\hat{\beta}$.

We start by introducing some preliminary facts and notation about the components of (1). Let $Y_i = (y_{i1}, y_{i2}, \dots, y_{im})'$, $i = 1, 2, \dots, m$ be independent random vectors with $E(Y_i) = \mu_i(\beta)$ and $cov(Y_i) = \Sigma_i(\beta, \alpha)$ where the expectation and the covariance are taken under the true probability distribution. Also, let $R(\alpha)$ be the working correlation matrix and let \bar{R} be the true correlation matrix.

Theorem 1. Let $\theta = (\beta, \alpha)'$, $Z(\beta^*, \beta) = A_i^{-1/2}(\beta^*)(y_i - \mu_i(\beta))$ and set

$$v_i(\beta^*, \theta) = \log|\Sigma_i| + Z_i'(\beta^*, \beta)R^{-1}(\alpha)Z_i(\beta^*, \beta). \quad (5)$$

If $\nabla v_i(\theta)$ and $\nabla^2 v_i(\theta)$ exist everywhere and satisfy $0 < E(\nabla^2 v_i(\theta)) < \infty$, then

- (a) $E(\nabla v_i(\theta)) = 0$,
- (b) $cov(\nabla v_i(\theta)) = \psi_i(\theta)$,
- (c) $E(\nabla^2 v_i(\theta)) = \xi_i(\theta)$,

where

$$\xi_i(\theta) = \begin{pmatrix} \xi_{i11}(\theta) & 0 \\ 0 & \xi_{i22}(\theta) \end{pmatrix}, \psi_i(\theta) = \begin{pmatrix} \psi_{i11}(\theta) & \psi_{i12}(\theta) \\ \psi_{i12}(\theta) & \psi_{i22}(\theta) \end{pmatrix},$$

with the entries of the above matrices given by

$$\begin{aligned} \xi_{i11}(\theta) &= 2D_i(\beta)A_i^{-1/2}(\beta)R^{-1}(\alpha)A_i^{-1/2}(\beta)D_i'(\beta), \\ \psi_{i11}(\theta) &= 4D_i(\beta)A_i^{-1/2}(\beta)R^{-1}(\alpha)\bar{R}R^{-1}(\alpha)A_i^{-1/2}(\beta)D_i'(\beta), \\ \psi_{i12}(\theta) &= 2tr\left\{e_j'D_i(\beta)A_i^{-1/2}(\beta)R^{-1}(\alpha)\otimes\frac{\partial}{\partial\alpha_j}R^{-1}(\alpha)\right\}E(Z_i\otimes Z_iZ_i'). \end{aligned}$$

Assume that if as $m \rightarrow \infty$, we have

$$\frac{1}{m}\sum_{i=1}^m \xi_i(\theta) \rightarrow \xi(\theta), \quad \frac{1}{m}\sum_{i=1}^m \psi_i(\theta) \rightarrow \psi(\theta), \quad (6)$$

where

$$\xi(\theta) = \begin{pmatrix} \xi_{11}(\theta) & 0 \\ 0 & \xi_{22}(\theta) \end{pmatrix}, \quad \psi(\theta) = \begin{pmatrix} \psi_{11}(\theta) & \psi_{12}(\theta) \\ \psi_{12}(\theta) & \psi_{22}(\theta) \end{pmatrix}.$$

Then,

$$\begin{aligned} \xi_{22}(\theta) &= tr\left[R^{-1}\frac{\partial^2 R(\alpha)}{\partial\alpha_j\partial\alpha_k}\right] - tr\left[R^{-1}\frac{\partial R(\alpha)}{\partial\alpha_j}R^{-1}\frac{\partial R(\alpha)}{\partial\alpha_k}\right] \\ &\quad + 2tr\left[R^{-1}\frac{\partial R(\alpha)}{\partial\alpha_j}R^{-1}\frac{\partial R(\alpha)}{\partial\alpha_k}R^{-1}\bar{R}\right] - tr\left[R^{-1}\frac{\partial^2 R(\alpha)}{\partial\alpha_j\partial\alpha_k}R^{-1}\bar{R}\right] \\ \psi_{22}(\theta) &= tr\left[\left\{\frac{\partial}{\partial\alpha_j}R^{-1}(\alpha)\otimes\frac{\partial}{\partial\alpha_k}R^{-1}(\alpha)\right\}\right. \\ &\quad \left.\times E(Z_iZ_i'\otimes Z_iZ_i')\right] - M_j(\alpha)M_k(\alpha), \end{aligned}$$

where $M_j(\alpha) = tr\left(\frac{\partial}{\partial\alpha_j}R^{-1}(\alpha)\bar{R}\right)$, e_j is a $(p \times 1)$ column vector with one at the j^{th} row and zero elsewhere, \bar{R} is the true correlation structure and \otimes is the Kronecker product. It is assumed throughout that components of $E(Z_i \otimes Z_i Z_i')$ and $E(Z_i Z_i' \otimes Z_i Z_i')$ are finite.

Next, we present our results on the large sample properties of the GE of (β, α) . The regularity conditions (6) needed to establish the consistency and asymptotic normality are similar to those used in the multivariate central limit theorem for independent but not identically distributed random vectors.

Theorem 2. Let $\theta = (\beta, \alpha)'$ and assume that (6) holds and as $\delta \rightarrow 0$ we have

$$E\left\{\sup_{\|h\|<\delta} |\nabla^2 v_i(\theta+h) - \nabla^2 v_i(\theta)|\right\} \rightarrow 0.$$

Then, the Gaussian estimate $(\hat{\beta}_m, \hat{\alpha}_m)$ is consistent for (β, α) and we conclude that

$$(\hat{\theta}_m - \theta) \text{ is } AN\left(0, \frac{1}{m}\xi^{-1}(\theta)\psi(\theta)\xi^{-1}(\theta)\right).$$

Theorem 2 establishes the consistency and asymptotic normality of the GE of (β, α) and we outline the proof in the Appendix.

Theorem 3. The marginal distributions of the Gaussian estimates $\hat{\beta}_m$ and $\hat{\alpha}_m$ are asymptotically correlated and

$$\begin{aligned} \hat{\beta}_m \text{ is } AN\left(\beta, \frac{1}{m}\xi_{11}^{-1}(\theta)\psi_{11}(\theta)\xi_{11}^{-1}(\theta)\right), \\ \hat{\alpha}_m \text{ is } AN\left(\alpha, \frac{1}{m}\xi_{22}^{-1}(\theta)\psi_{22}(\theta)\xi_{22}^{-1}(\theta)\right). \end{aligned}$$

In view of the previous theorems, we notice that if $R(\alpha)$ is correctly specified and is equal to \bar{R} , then we have $\psi_{i11}(\theta) = 2\xi_{i11}(\theta)$, $\psi_{11}(\theta) = 2\xi_{11}(\theta)$, and the asymptotic covariance of $\hat{\beta}_m$ reduces to $\frac{2\xi_{11}^{-1}(\theta)}{m}$. Also, the asymptotic distribution of $\hat{\beta}_m$

depends on the parameters β, α, ϕ and \bar{R} while the asymptotic distribution of $\hat{\alpha}_m$ depends on α and \bar{R} but not on β . It also depends on the third and fourth moments of Y_i , which may cause a loss of efficiency if these moments are misspecified. The sandwich-type covariance estimator protects against the impact of misspecification of $R(\alpha)$, however, we may lose some efficiency when using this estimator.

5. Data analysis

Consider the dataset from [Potthoff and Roy \(1964\)](#) that consists of 27 subjects in a dental study (16 boys and 11 girls). The response y_{ij} is the distance, in mm, from the center of each subjects' pituitary to pteryomaxillary fissure recorded at the age of 8, 10, 12 and 14. We treat the data set as normally distributed with a mean vector of $\mu = X\beta$ and covariance matrix Σ . Also, we consider the models used by [Chaganty and Shults \(1999\)](#) and analyze the data using the GE and compare the results with those in [Chaganty and Shults \(1999\)](#). First, we fit the model:

$$\mu_{ij} = \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma_1 x_{i1} * x_{i3} + \gamma_2 x_{i2} * x_{i3}, 1 \leq j \leq 4, 1 \leq i \leq 27$$

where, x_{i1}, x_{i2} are indicator variables for the two genders: boys and girls, respectively. The covariate x_{i3} is the subject's age at the j^{th} measurement time. The estimates of (β, α, ϕ) are computed using the Newton-Raphson iterative method assuming working correlation structures to be:

1. AR(1) structure,
2. An unstructured correlation matrix.

[Table 1](#) contains the estimates and the standard errors for the parameters using the GEE, C-QLS and the GE with AR(1) working correlation structure. Similarly, [Table 2](#) con-

tains the estimates when the working correlation matrix is unstructured. The standard error in both Tables were computed using the model-robust, sandwich-type estimator. From [Tables 1 and 2](#), it can be seen that the estimates of the regression parameters are in a reasonable agreement, but the standard errors are slightly different. Overall, the GE and C-QLS have smaller standard errors than the GEE also the GE has a smaller standard errors for all estimates except for the estimate of γ_2 in [Table 2](#). Finally, [Tables 3](#) show the estimates of the working correlation matrices using the GE for the AR(1) and the unstructured correlation matrix.

Appendix A. Consider the objective function

$$v_i(\beta^*, \theta) = \log|\Sigma_i| + Z_i'(\beta^*, \beta)R^{-1}(\alpha)Z_i(\beta^*, \beta)$$

where, $|\Sigma_i| = \det(\Sigma_i)$, $\theta = (\beta, \alpha)'$, $R(\alpha)$ is the working correlation matrix, $Z(\beta^*, \beta) = A_i^{-1/2}(\beta^*)(y_i - \mu_i(\beta))$ and β is a $(p \times 1)$ vector of unknown regression parameters while α is a $(q \times 1)$ vector of unknown correlation parameters. In this proof, we use arguments similar to those found in [Sen and Singer \(1993, p. 206\)](#).

Define $\epsilon_m(t, \theta)$ by

$$\epsilon_m(t, \theta) = \sum_{i=1}^m \left\{ v_i \left(\beta^*, \theta + \frac{t}{\sqrt{m}} \right) - v_i(\beta^*, \theta) \right\}.$$

Under certain regularity conditions and using the Taylor series expansion for $v_i \left(\beta^*, \theta + \frac{t}{\sqrt{m}} \right)$ around θ and for $\|t\| \leq K$, $0 < K < \infty$, we can rewrite $\epsilon_m(t, \theta)$ as:

$$\epsilon_m(t, \theta) = \frac{1}{\sqrt{m}} \sum_{i=1}^m t' \nabla v_i(\theta) + \frac{1}{2m} \sum_{i=1}^m t' \nabla^2 v_i(\theta^*) t,$$

where θ^* is a point on the line joining θ and $\theta + \frac{t}{\sqrt{m}}$.

Now, adding and subtracting the term $\frac{1}{2m} \sum_{i=1}^m t' \nabla^2 v_i(\theta) t$, we get:

Table 1 Parameter estimates of the dental data using AR(1) structure.

Parameter	GEE		C-QLS		GE	
	Est.	Std.	Est.	Std.	Est.	Std.
β_1	16.5946	1.2788	16.5931	1.2781	16.5925	1.2778
β_2	17.3213	0.7780	17.3215	0.7776	17.3216	0.7774
γ_1	0.7965	0.1050	0.7695	0.1049	0.7696	0.1049
γ_2	0.4838	0.0629	0.4837	0.0629	0.4837	0.0629
ϕ	4.9107		4.9106		4.9105	

Table 2 Parameter estimates of the dental data using unstructured correlation.

Parameter	GEE		C-QLS		GE	
	Est.	Std.	Est.	Std.	Est.	Std.
β_1	16.3236	1.1701	16.0523	1.1288	15.8841	1.1232
β_2	17.3973	0.7244	17.4018	0.6972	17.4103	0.6944
γ_1	0.7881	0.0983	0.8122	0.0939	0.8276	0.0929
γ_2	0.4781	0.0639	0.4770	0.0632	0.4760	0.0642
ϕ	4.9058		4.9076		4.9101	

Table 3 Estimated correlation matrix using GE.

AR(1)	Unstructured						
–	0.6082	0.3700	0.2250	–	0.5201	0.6135	0.4873
0.6082	–	0.6082	0.3700	0.5201	–	0.5218	0.7433
0.3700	0.6082	–	0.6082	0.6135	0.5218	–	0.6921
0.2250	0.3700	0.6082	–	0.4873	0.7433	0.6921	–

$$\epsilon_m(t, \theta) = \frac{1}{\sqrt{m}} \sum_{i=1}^m t' \nabla v_i(\theta) + \frac{1}{2m} \sum_{i=1}^m t' \nabla^2 v_i(\theta) t + \frac{1}{2} t' \eta_m(t, \theta) t \quad (\text{A.1})$$

where,

$$\eta_m(t, \theta) = \frac{1}{m} \sum_{i=1}^m \{\nabla^2 v_i(\theta^*) - \nabla^2 v_i(\theta)\}.$$

Given $\delta > 0$, there exists an integer $m_0 = m_0(\delta)$ such that

$$\frac{1}{\sqrt{m}} \|t\| \leq \frac{1}{\sqrt{m}} K < \delta,$$

for all $m \geq m_0$. Therefore, for sufficiently large m , we have

$$\begin{aligned} |\eta_m(t)| &\leq \frac{1}{m} \sum_{i=1}^m \sup_{\{h: \|h\| \leq \frac{\|t\|}{\sqrt{m}}\}} |\nabla^2 v_i(\theta + h) - \nabla^2 v_i(\theta)| \\ &\leq \frac{1}{m} \sum_{i=1}^m \sup_{\{h: \|h\| \leq \delta\}} |\nabla^2 v_i(\theta + h) - \nabla^2 v_i(\theta)| \end{aligned}$$

Hence, using Khintchine strong law of large numbers (Sen and Singer, 1993, p 69) and assuming that $E(\nabla^2 v_i(\theta))$ is finite, we have

$$\frac{1}{m} \sum_{i=1}^m \sup_{\{h: \|h\| \leq \frac{\|t\|}{\sqrt{m}}\}} |\nabla^2 v_i(\theta + h) - \nabla^2 v_i(\theta)| \rightarrow 0,$$

almost surely as $\delta \rightarrow 0$, therefore we have $\sup \{|\eta_m(t)|, \|t\| \leq K\} \rightarrow 0$ almost surely as $m \rightarrow \infty$.

This will be true since the value of θ^* lies on the line joining θ and $\theta + \frac{t}{\sqrt{m}}$. Now

$$\begin{aligned} \epsilon_m(t, \theta) &= \frac{1}{\sqrt{m}} \sum_{i=1}^m t' \nabla v_i(\theta) + \frac{1}{2m} \sum_{i=1}^m t' \zeta_i(\theta) t + \frac{1}{2m} \sum_{i=1}^m t' \{\nabla^2 v_i(\theta) \\ &\quad - \zeta_i(\theta)\} t + \frac{1}{2} t' \eta_m(t) t. \end{aligned}$$

Hence, using the weak law of large numbers, we get

$$\frac{1}{2m} \sum_{i=1}^m t' \{\nabla^2 v_i(\theta) - \zeta_i(\theta)\} t = o_p(1).$$

Therefore,

$$\epsilon_m(t, \theta) = \frac{1}{\sqrt{m}} \sum_{i=1}^m t' \nabla v_i(\theta) + \frac{1}{2m} \sum_{i=1}^m t' \zeta_i(\theta) t + o_p(1).$$

In order to find the value of t that minimizes the term $\epsilon_m(t, \theta)$, we will find the first derivative of $\epsilon_m(t, \theta)$ with respect to t and disregarding the term $o_p(1)$, hence this will give us the following

$$\hat{t}_m = - \left(\frac{1}{m} \sum_{i=1}^m \zeta_i(\theta) \right)^{-1} \left(\frac{1}{\sqrt{m}} \sum_{i=1}^m \nabla v_i(\theta) \right) + o_p(1). \quad (\text{A.2})$$

Notice that using the definition of $\epsilon_m(t, \theta)$, we conclude that \hat{t}_m will also correspond closely to the Gaussian estimate of θ , which is attained at $\hat{\theta}_m$, therefore we conclude that $\hat{\theta}_m$ can be written as follows

$$\hat{\theta}_m = \theta + \frac{\hat{t}_m}{\sqrt{m}} + o_p\left(\frac{1}{\sqrt{m}}\right). \quad (\text{A.3})$$

But since $\frac{1}{m} \sum_{i=1}^m \nabla v_i(\theta) \rightarrow E \nabla v_i(\theta) = 0$ and using Theorem 1, (A.2), (A.3) and using the weak law of large numbers, we conclude that:

$$\hat{\theta}_m - \theta \rightarrow 0 \quad (\text{A.4})$$

in probability as $m \rightarrow \infty$.

The consistency of Gaussian estimates could be easily concluded from (A.4). Finally, under the assumptions of A.2, A.3, A.4 and (6), the results in Theorem 1, the multivariate central limit theorem and Slutsky's theorem, we conclude that

$$\begin{aligned} \hat{\theta}_m - \theta &= \frac{\hat{t}_m}{\sqrt{m}} \\ &\quad + o_p\left(\frac{1}{\sqrt{m}}\right) \text{ is } AN\left(0, \frac{1}{m} \xi^{-1}(\theta) \psi(\theta) \xi^{-1}(\theta)\right), \end{aligned} \quad (\text{A.5})$$

which is the end of the proof.

References

- Al-Rawwash, M., 2001. Gaussian estimation and modeling covariance in longitudinal data analysis. Ph.D. thesis, Northern Illinois University, USA.
- Al-Rawwash, M., 2005. Covariance matrix estimation using repeated measurements when data are incomplete. *Applied Mathematics and Computation* 167, 299–315.
- Al-Rawwash, M., Pourahmadi, M., 2006. Gaussian estimation and joint modeling of dispersions and correlations in longitudinal data. *Computer Methods and Programs in Biomedicine* 82, 106–113.
- Buzkova, P., Lamley, T., 2008. Semiparametric log-linear regression for longitudinal measurements subject to outcome-dependent follow-up. *Journal of Statistical Planning and Inference* 138, 2450–2461.
- Chaganty, N.R., 1997. An alternative approach to the analysis of longitudinal data via generalized estimating equations. *Journal of Statistics Planning and Inference* 63, 39–54.
- Chaganty, N.R., Shults, J., 1999. On eliminating the asymptotic bias in the quasi-least squares estimate of the correlation parameter. *Journal of Statistics Planning and Inference* 76, 145–161.
- Crowder, M., 1995. On the use of a working correlation matrix in using generalized linear models for repeated measurements. *Biometrika* 82, 407–410.
- Crowder, M., 2001. On repeated measures analysis with misspecified covariance structure. *Journal of Royal Statistical Society B* 63, 55–62.
- Diggle, P., Liang, K., Zeger, S., 1994. *Analysis of Longitudinal Data*. Oxford University Press, Oxford.

- Heyde, C.C., 1997. *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. Springer-Verlag, New York.
- Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, 73 13–22.
- Lipsitz, S.R., Molenberghs, G., Fitzmaurice, G.M., Ibrahim, J., 2000. GEE with Gaussian estimation of the correlations when data are incomplete. *Biometrics* 56, 528–536.
- Potthoff, R.F., Roy, S.N., 1964. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51, 665–680.
- Pourahmadi, M., 2000. Maximum likelihood estimation of generalized linear multivariate normal covariance matrix. *Biometrika* 87, 425–435.
- Pourahmadi, M., Daniels, M., Park, T., 2007. Simultaneous modelling of the cholesky decomposition of several covariance matrices. *Journal of Multivariate Analysis* 98, 568–587.
- Sen, P., Singer, J., 1993. *Large sample methods in statistics. An Introduction with Applications*. Chapman and Hall, New York.
- Shults, J., Chaganty, N.R., 1998. Analysis of serially correlated data using quasi-least squares. *Biometrics* 54, 1622–1630.
- Vonesh, E.F., Chinchilli, V.M., 1997. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York.
- Wang, Y., Carey, V., 2003. Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. *Biometrika* 90, 29–41.
- Wedderburn, R.W.M., 1974. Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* 61, 439–447.
- Whittle, P., 1961. Gaussian estimation in stationary time series. *Bulltten Institute of International Statistics* 39, 105–112, livraison 2.