

An improved quadratic inference function for parameter estimation in the analysis of correlated data

Philip M. Westgate^{a,*†} and Thomas M. Braun^b

Generalized estimating equations (GEE) are commonly employed for the analysis of correlated data. However, the quadratic inference function (QIF) method is increasing in popularity because of its multiple theoretical advantages over GEE. We base our focus on the fact that the QIF method is more efficient than GEE when the working covariance structure for the data is misspecified. It has been shown that because of the use of an empirical weighting covariance matrix inside its estimating equations, the QIF method's realized estimation performance can potentially be inferior to GEE's when the number of independent clusters is not large. We therefore propose an alternative weighting matrix for the QIF, which asymptotically is an optimally weighted combination of the empirical covariance matrix and its model-based version, which is derived by minimizing its expected quadratic loss. Use of the proposed weighting matrix maintains the large-sample advantages the QIF approach has over GEE and, as shown via simulation, improves small-sample parameter estimation. We also illustrated the proposed method in the analysis of a longitudinal study. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: correlated data; efficiency; estimating equations; expected quadratic loss; marginal model

1. Introduction

The use of generalized estimating equations (GEE) [1] for the marginal analysis of correlated data is very common in practice. Cluster randomized trials (CRTs) and longitudinal studies are two examples in which correlated data arise. In CRTs, independent groups of subjects are randomized to different trial arms, and subjects within the same group, or cluster, are correlated because of unknown clustering effects. In longitudinal studies, independent subjects contribute multiple observations over time, and outcomes from the same subject are correlated. An example is a study we later look at in which 59 epileptic subjects, who were randomized to either a treatment drug or control, were followed over time, with interest in the number of seizures each person had during four successional 2-week intervals [2, 3].

The quadratic inference function (QIF) method, proposed by Qu *et al.* [4], is an attractive alternative to GEE when fitting a marginal model. The QIF method rewrites the inverse of the working correlation matrix as the weighted sum of m basis matrices inside the GEE, and then the corresponding m distinct GEE components are optimally weighted using the generalized method of moments (GMM) [5]. Theoretically, the use of these optimal estimating equations makes the QIF method more efficient than GEE when both methods utilize the same incorrect covariance structure, whereas both methods are equally efficient when incorporating the true structure. Furthermore, the QIF method is more robust to outliers relative to GEE [6], and the QIF can itself be used as a statistic in goodness-of-fit and likelihood ratio score-type tests [4, 7].

^aDepartment of Biostatistics, College of Public Health, University of Kentucky, Lexington, KY 40536, U.S.A.

^bDepartment of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, U.S.A.

*Correspondence to: Philip M. Westgate, Department of Biostatistics, College of Public Health, University of Kentucky, Lexington, KY 40536, U.S.A.

†E-mail: philip.westgate@uky.edu

Westgate and Braun [8] and Westgate [9] showed that for finite-sample sizes, the variances of parameter estimates from the QIF method can be larger than theoretically expected, implying that the corresponding estimates from GEE can potentially be less variable in some settings. Therefore, which of these two methods should realistically be used will be unknown in practice, although theoretical efficiency results imply that the QIF approach should always be utilized. They explain that the empirical covariance weighting matrix implemented inside the QIF is the source of both its efficiency advantage and potential drawback in small to moderately sized samples. Specifically, the smaller the number of clusters of subjects in CRTs or the number of subjects with repeated measures, the more variable this weighting matrix will be. In practice, estimated empirical covariances are used in this weighting matrix, and because of this matrix's finite-sample variability, the use of these estimates will increase the variances of the final parameter estimates [8–11].

This paper develops an improvement to the weighting covariance matrix employed with the QIF approach that is meant to eliminate potential small-sample estimation deficiencies relative to GEE, while maintaining the QIF's large-sample advantages. Specifically, we propose utilizing a weighted combination of the empirical covariance matrix and its respective model-based version, in which the corresponding weights minimize the expected quadratic loss of the resulting asymptotically optimal matrix. Section 2 discusses GEE and the QIF method in more detail. In Section 3, we develop the proposed weighting matrix for the QIF method, discuss the corresponding estimation procedure, and derive the estimated covariance matrix of resulting parameter estimates. Via simulation study in general finite-sample settings, Section 4 illustrates the realized estimation performance from use of the improved weighting matrix with the QIF as compared with the use of GEE, the typical QIF, and a QIF proposed by Han and Song [12]. The performances of these methods are then contrasted in application to the seizure study in Section 5. Section 6 then gives the concluding remarks, whereas the Appendix presents proofs justifying the use of the proposed weighting matrix.

2. Generalized estimating equations and the quadratic inference function method

We let $Y_i = [Y_{i1}, \dots, Y_{in_i}]^T$ denote the vector of outcomes with marginal mean $\mu_i = E(Y_i)$ for the i th independent cluster of correlated data, $i = 1, \dots, N$. Furthermore, for the j th observation, $j = 1, \dots, n_i$, in the i th cluster, we have $f(\mu_{ij}) = \mathbf{x}_{ij}\boldsymbol{\beta}$, where f is a known link function, $\mathbf{x}_{ij} = [1, x_{ij1}, \dots, x_{ij(p-1)}]$ is a vector of covariate values, and $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_{p-1}]^T$ is a $p \times 1$ vector of regression parameters.

We can acquire estimates for $\boldsymbol{\beta}$ by setting the GEE equal to 0,

$$\sum_{i=1}^N \mathbf{D}_i^T \mathbf{V}_i^{-1} (Y_i - \mu_i) = \mathbf{0},$$

where $\mathbf{D}_i = \partial \mu_i / \partial \boldsymbol{\beta}^T$ and $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}_i \mathbf{A}_i^{1/2}$ is the working model-based covariance structure for Y_i . Here, \mathbf{A}_i is a diagonal matrix of working marginal variances for the n_i outcomes, and \mathbf{R}_i is their working correlation structure. As GEE can be inefficient when the working covariance structure is incorrect, Qu *et al.* [4] proposed the QIF approach, which assumes $\mathbf{R}_i^{-1} \approx \sum_{r=1}^m \gamma_{ri} \mathbf{M}_{ri}$. Here, \mathbf{M}_{ri} , $r = 1, \dots, m$, are known basis matrices, and γ_{ri} , $r = 1, \dots, m$, are functions of a correlation parameter(s) and possibly cluster size [4]. Typically, $m = 2$ is used to at least approximate the inverse of independence, unstructured, exchangeable, and autoregressive (AR)-1 correlation structures [3, 7]. We can then write the GEE as

$$\sum_{r=1}^m \sum_{i=1}^N \gamma_{ri} \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_{ri} \mathbf{A}_i^{-1/2} (Y_i - \mu_i) = \sum_{r=1}^m \sum_{i=1}^N \gamma_{ri} g_{ri}(\boldsymbol{\beta}) = \mathbf{0}.$$

So that we can optimally weight these m distinct GEE components, the QIF uses extended score equations (ESEs), given as

$$\bar{g}_N(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^N g_i(\boldsymbol{\beta}) = \begin{bmatrix} \frac{1}{N} \sum_{i=1}^N g_{1i}(\boldsymbol{\beta}) \\ \vdots \\ \frac{1}{N} \sum_{i=1}^N g_{mi}(\boldsymbol{\beta}) \end{bmatrix}.$$

These are the m components from GEE, but without γ_{ri} , $r = 1, \dots, m$; $i = 1, 2, \dots, N$. The ESEs are then used in Hansen's [5] GMMs to create the QIF, $Q_N(\beta) = N \bar{g}_N^T(\beta) C_N^{-1}(\beta) \bar{g}_N(\beta)$, in which $C_N(\beta) = (1/N) \sum_{i=1}^N g_i(\beta) g_i^T(\beta)$ is the empirical covariance matrix for the ESEs and is used to estimate the optimal weighting matrix, $E[C_N(\beta)] = \Sigma_N = (1/N) \sum_{i=1}^N \text{Cov}[g_i(\beta)]$. $\arg \min_{\beta} Q_N(\beta)$ gives the estimate for β , which is asymptotically equivalent to solving

$$N \dot{g}_N^T(\beta) C_N^{-1}(\beta) \bar{g}_N(\beta) = \mathbf{0} \tag{1}$$

for β , in which $\dot{g}_N(\beta) = E[\partial \bar{g}_N(\beta) / \partial \beta^T]$. As $C_N(\beta) - \Sigma_N \xrightarrow{p} 0$ [4, 12], the estimating equations in Equation (1) are asymptotically equivalent to $N \dot{g}_N^T(\beta) \Sigma_N^{-1} \bar{g}_N(\beta) = \mathbf{0}$, which are estimating equations that have optimally weighted the ESEs and thus the m distinct GEE components [4, 13, 14]. Therefore, the QIF method is more efficient than GEE when the working and true covariance structures differ, and the two methods are equally efficient when using the correct structure [4]. We note, however, as detailed in Westgate and Braun [8], that if cluster sizes vary and a working exchangeable correlation structure is utilized, this efficiency result may not hold unless the ESEs replace g_{ri} with $\alpha_{ri} g_{ri}$, $r = 1, \dots, m$; $i = 1, \dots, N$.

Westgate and Braun [8] and Westgate [9] demonstrated that the variable empirical nature of $C_N(\beta)$ and the use of a working consistent estimate, $\hat{\beta}$, for β inside C_N result in the QIF method's potentially inferior, rather than superior, estimation performance relative to GEE. Specifically, $N \dot{g}_N^T(\hat{\beta}) C_N^{-1}(\hat{\beta}) \bar{g}_N(\hat{\beta}) = \mathbf{0}$ gives the estimating equations employed in practice with the QIF approach, with final estimates denoted by $\hat{\beta}$. As the variable empirical nature of C_N is a cause for the QIF approach to not realistically attain its theoretical efficiency from the use of Σ_N as the weighting matrix, we propose an alternative weighting matrix to use in place of C_N that is more stable in terms of expected quadratic loss in finite samples and is still asymptotically optimal.

3. An improved quadratic inference function approach

3.1. An improved weighting matrix

In some settings consisting of small to moderately sized samples, GEE can work better than the QIF method for parameter estimation owing to employing model-based covariance structures to weight outcomes. Therefore, when C_N is quite variable, a potentially better estimate for Σ_N would utilize the corresponding model-based version, $M_N = (1/N) \sum_{i=1}^N \widehat{\text{Cov}}[g_i(\beta)]$, which incorporates the working covariance structures. Specifically, $\widehat{\text{Cov}}[g_i(\beta)] = \hat{E}[g_i(\beta) g_i^T(\beta)] = \hat{E}[B_i e_i e_i^T B_i^T] = B_i A_i^{-1/2} V_i A_i^{-1/2} B_i^T = B_i R_i B_i^T$, where

$$B_i = \begin{bmatrix} D_i^T A_i^{-1/2} M_{1i} \\ \vdots \\ D_i^T A_i^{-1/2} M_{mi} \end{bmatrix},$$

$e_i = A_i^{-1/2}(Y_i - \mu_i)$ are defined by Han and Song [12], and \hat{E} denotes the estimated expected value incorporating the model-based covariance structure.

Sole use of M_N , rather than C_N , as the weighting covariance matrix in Equation (1) can also be disadvantageous, however. Specifically, the QIF approach will no longer have an efficiency advantage over GEE. Furthermore, even for small N , if the true covariance structure is misspecified, in some settings, the corresponding bias in M_N can be more detrimental than the variability in C_N with respect to parameter estimation. We therefore propose replacing $C_N(\beta)$ in Equation (1) with an improved weighting matrix, $C_N^* = \rho_N M_N + (1 - \rho_N) C_N(\beta)$, that optimally takes into account both the variability in C_N and the bias in M_N , while maintaining the theoretical advantages the QIF approach has over GEE. In practice, we note that $\tilde{\beta}$ will replace β inside C_N . Here, $\rho_N = \tau_N^2 / (\alpha_N^2 + \tau_N^2) = \tau_N^2 / \delta_N^2$, $\alpha_N^2 = \|\mathbf{M}_N - \Sigma_N\|^2$, $\tau_N^2 = E[\|C_N(\beta) - \Sigma_N\|^2]$, and $\delta_N^2 = E[\|C_N(\beta) - M_N\|^2]$, where $\|\mathbf{K}\| = \sqrt{\text{tr}(\mathbf{K} \mathbf{K}^T)} / p$ for some arbitrary $p \times p$ matrix \mathbf{K} [15]. This value for ρ_N minimizes expected quadratic loss, given by $E[\|C_N^* - \Sigma_N\|^2]$. Furthermore, τ_N^2 and α_N^2 take into account the variability in

C_N and bias in M_N , respectively. Additionally, for several conditions that are typically met in practice, $E[\|C_N^* - \Sigma_N\|^2] \rightarrow 0$ as $N \rightarrow \infty$, implying that $C_N^* - \Sigma_N \xrightarrow{p} 0$ [12, 15] (proof in Appendix).

The proposed weighting matrix is related to the works of Ledoit and Wolf [15] and Han and Song [12]. Ledoit and Wolf [15] proposed a well-conditioned estimated covariance matrix that is a weighted combination of the identity and sample covariance matrices. Han and Song [12] extended this idea for the QIF, proposing $S_N = \rho_N \mu_N I + (1 - \rho_N)C_N(\beta)$, in which I is the identity matrix, μ_N is the average value for the diagonal elements of Σ_N , and ρ_N minimizes $E[\|S_N - \Sigma_N\|^2]$. They derived this alternative weighting matrix, referred to as the linear shrinkage estimator, as C_N may not be invertible in some study designs. However, although the use of S_N can lead to more stable results because this particular problem is fixed, it is not designed to improve the QIF method's estimation performance in general settings. Particularly, $\mu_N I$ is not meant to model Σ_N , whereas this is the sole purpose of M_N .

Currently assuming β is known, M_N and ρ_N need to be estimated in practice. Similar to Ledoit and Wolf [15] and Han and Song [12], we propose the following:

- The estimator for M_N is \hat{M}_N , in which covariance parameters are estimated.
- The estimator for δ_N^2 is $d_N^2 = \|C_N(\beta) - \hat{M}_N\|^2$.
- The estimator for τ_N^2 is $t_N^2 = \min[\bar{t}_N^2, d_N^2]$, $\bar{t}_N^2 = (1/N^2) \sum_{i=1}^N \|g_i(\beta)g_i(\beta)^T - C_N(\beta)\|^2$.
- The estimator for α_N^2 is $a_N^2 = d_N^2 - t_N^2$.
- The estimator for C_N^* is $\hat{C}_N^* = (t_N^2/d_N^2)\hat{M}_N + (a_N^2/d_N^2)C_N(\beta) = \hat{\rho}_N \hat{M}_N + (1 - \hat{\rho}_N)C_N(\beta)$.

The use of \bar{t}_N^2 is appropriate in the settings of Han and Song [12], as they deal with balanced covariate designs. However, in many general applications, the covariances of the N extended score components will likely vary, inducing bias in \bar{t}_N^2 . Particularly, $\text{Bias}(\bar{t}_N^2) \approx (1/N^2) \sum_{i=1}^N \|\text{Cov}[g_i(\beta)]\|^2 - (1/N^3) \|\sum_{i=1}^N \text{Cov}[g_i(\beta)]\|^2$ (Appendix). We can estimate bias using model-based covariances, yielding $\hat{t}_N^2 = \max(0, \min[\bar{t}_N^2 - \widehat{\text{Bias}}(\bar{t}_N^2), d_N^2])$ as an alternative estimate for τ_N^2 . Results using \hat{t}_N^2 and t_N^2 are asymptotically equivalent.

The Appendix gives the justifications for estimators. Specifically, $a_N^2 - \alpha_N^2$, $t_N^2 - \tau_N^2$, and $d_N^2 - \delta_N^2$ all converge in quadratic mean to 0 as $N \rightarrow \infty$, under the assumption that $E[\|\hat{M}_N - M_N\|^4] \rightarrow 0$ as $N \rightarrow \infty$. Furthermore, $E[\|\hat{C}_N^* - \Sigma_N\|^2] \rightarrow 0$ as $N \rightarrow \infty$, implying that \hat{C}_N^* is asymptotically optimal because $\hat{C}_N^* - \Sigma_N \xrightarrow{p} 0$ [12, 15] (proof in Appendix). Finally, $\tilde{\beta}$ must replace β inside \hat{C}_N^* . However, because $\tilde{\beta} - \beta \xrightarrow{p} 0$, we still have $\hat{C}_N^* - \Sigma_N \xrightarrow{p} 0$.

3.2. The estimation procedure

As discussed by Westgate (submitted), the estimation procedure that iteratively updates the parameter estimates inside C_N has been suggested in the QIF literature. Specifically, the procedure is given by

$$\tilde{\beta}^{(k+1)} \approx \tilde{\beta}^{(k)} - \left[\dot{g}_N^T(\tilde{\beta}^{(k)}) C_N^{-1}(\tilde{\beta}^{(k)}) \dot{g}_N(\tilde{\beta}^{(k)}) \right]^{-1} \dot{g}_N^T(\tilde{\beta}^{(k)}) C_N^{-1}(\tilde{\beta}^{(k)}) \bar{g}_N(\tilde{\beta}^{(k)});$$

$$k = 1, 2, \dots,$$

and iterates until convergence, yielding $\hat{\beta}$. Furthermore, if the value for the QIF increases at any given step of the estimation procedure, Loader and Pilla [16] proposed multiplying the quantity that is subtracted from $\tilde{\beta}^{(k)}$ by a scalar that reduces in value until the QIF decreases in value. Westgate (submitted) showed via simulation that choice of the initial estimates, $\tilde{\beta}^{(1)}$, whether from GEE using independence or GEE using the working covariance structure that is implemented with the QIF, has only small influence on the variances of final estimates. However, holding the initial estimates fixed inside C_N throughout the estimation procedure was shown to decrease the mean squared errors (MSEs) of final estimates, especially in small-sample settings when the working AR-1 correlation structure was implemented. Thus, we propose the use of the following estimation procedure that employs our proposed weighting matrix:

$$\tilde{\beta}^{(k+1)} \approx \tilde{\beta}^{(k)} - \left[\dot{g}_N^T(\tilde{\beta}^{(1)}) \hat{C}_N^{*-1}(\tilde{\beta}^{(1)}) \dot{g}_N(\tilde{\beta}^{(1)}) \right]^{-1} \dot{g}_N^T(\tilde{\beta}^{(1)}) \hat{C}_N^{*-1}(\tilde{\beta}^{(1)}) \bar{g}_N(\tilde{\beta}^{(k)});$$

$$k = 1, 2, \dots$$

3.3. A bias-corrected covariance matrix for parameter estimates

$\widehat{\text{cov}}(\hat{\beta}) = (1/N)\mathbf{J}_N^{-1}$ gives the asymptotic empirical sandwich estimator for the covariance matrix of parameter estimates from the QIF method, where $\mathbf{J}_N = \dot{g}_N^T C_N^{-1}(\beta) \dot{g}_N$ and \dot{g}_N denotes $\dot{g}_N(\beta)$. As discussed by Westgate [9], estimates are used in place of β inside C_N , inducing two types of finite-sample biases inside this covariance estimator. First, as already discussed, the use of estimates inside C_N during the estimation procedure increases the variances in the final estimates, which \mathbf{J}_N does not take into account. Second, $C_N(\hat{\beta})$ is biased for Σ_N when N is small as estimated residuals, $\hat{r}_i = Y_i - \hat{\mu}_i$, $i = 1, \dots, N$, will tend to be too small on average. Westgate [9] therefore proposed the following consistent bias-corrected estimate for the covariance of $\hat{\beta}$: $\widehat{\text{cov}}(\hat{\beta}) = (1/N)(I_p + G)\mathbf{J}_N^{-1} \dot{g}_N^T C_N^{-1} \tilde{C}_N C_N^{-1} \dot{g}_N \mathbf{J}_N^{-1} (I_p + G)^T$. Here, I_p is a $p \times p$ identity matrix and $G = -\partial [\mathbf{J}_N^{-1} \dot{g}_N^T C_N^{-1}(\beta^*) \tilde{g}_N(\beta)] / \partial \beta^{*T} |_{\beta^* = \beta}$, in which the k th column is given by

$$\mathbf{J}_N^{-1} \dot{g}_N^T C_N^{-1}(\beta) \frac{\partial C_N(\beta)}{\partial \beta_{k-1}} C_N^{-1}(\beta) \tilde{g}_N(\beta),$$

where

$$\frac{\partial C_N(\beta)}{\partial \beta_{k-1}} = \frac{1}{N} \sum_{i=1}^N \left(\frac{\partial g_i(\beta)}{\partial \beta_{k-1}} g_i^T(\beta) + g_i(\beta) \frac{\partial g_i^T(\beta)}{\partial \beta_{k-1}} \right).$$

Also, $\tilde{C}_N(\hat{\beta})$ replaces $\hat{r}_i \hat{r}_i^T$, $i = 1, \dots, N$, which comprise $C_N(\hat{\beta})$, with bias-corrected versions given by either $(\mathbf{I}_i + \mathbf{O}_i)^{-1} \hat{r}_i \hat{r}_i^T (\mathbf{I}_i + \mathbf{O}_i^T)^{-1}$ or $(\mathbf{I}_i + \mathbf{O}_i)^{-1} \hat{r}_i \hat{r}_i^T$, $i = 1, \dots, N$. Here,

$$\mathbf{O}_i = \frac{1}{N} \mathbf{D}_i (I_p + G) \mathbf{J}_N^{-1} \dot{g}_N^T C_N^{-1} \begin{bmatrix} \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_{1i} \mathbf{A}_i^{-1/2} \\ \vdots \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_{mi} \mathbf{A}_i^{-1/2} \end{bmatrix}.$$

The use of this G matrix, derived originally for GMM estimators by Windmeijer [10, 11], accounts for the first type of bias, whereas the use of \mathbf{O}_i accounts for the second source of bias. In the simulation study presented by Westgate [9], the use of $(\mathbf{I}_i + \mathbf{O}_i)^{-1} \hat{r}_i \hat{r}_i^T (\mathbf{I}_i + \mathbf{O}_i^T)^{-1}$, similar to the correction Mancl and DeRouen [17] derived for the sandwich estimator with GEE, performed better than the use of $(\mathbf{I}_i + \mathbf{O}_i)^{-1} \hat{r}_i \hat{r}_i^T$, similar to the alternative correction derived by Kauermann and Carroll [18].

As our proposed weighting covariance matrix is a weighted combination of C_N and \hat{M}_N , $(1/N)\mathbf{J}_N^{*-1} \dot{g}_N^T \hat{C}_N^{*-1} C_N \hat{C}_N^{*-1} \dot{g}_N \mathbf{J}_N^{*-1}$ gives the sandwich estimator for the covariance matrix of the resulting parameter estimates, $\hat{\beta}$, when ignoring both types of bias. Here, \mathbf{J}_N^{*-1} replaces C_N with \hat{C}_N^* inside \mathbf{J}_N . Modifying the derivations in Westgate [9], we propose the use of the following consistent bias-corrected estimator, with unknown parameters replaced by their corresponding estimates:

$$\widehat{\text{cov}}(\hat{\beta}) = (1/N)[I_p + (1 - \rho_N)\tilde{G}]\mathbf{J}_N^{*-1} \dot{g}_N^T \hat{C}_N^{*-1} \tilde{C}_N \hat{C}_N^{*-1} \dot{g}_N \mathbf{J}_N^{*-1} [I_p + (1 - \rho_N)\tilde{G}]^T.$$

Here, \tilde{G} replaces C_N and \mathbf{J}_N with \hat{C}_N^* and \mathbf{J}_N^* , respectively, inside the formula for G . Furthermore, we use these same replacements, along with $(1 - \rho_N)\tilde{G}$ in place of G , within \mathbf{O}_i inside \tilde{C}_N .

4. Simulation study

4.1. Study description

To assess the estimation performances of GEE using a common correlation parameter, the typical QIF, the QIF proposed by Han and Song [12], and our proposed QIF with the improved weighting matrix, we use empirical MSE quantities that are the sum of the empirical MSEs from all nonintercept parameters. Tables report MSE ratios, which take the empirical MSE quantity from GEE and divides it by the MSE value from the specified QIF approach that incorporates the same working correlation structure, and the empirical mean of the estimated weights given to \hat{M}_N inside our proposed weighting matrix.

Table I. Ratios of empirical MSEs from generalized estimating equations divided by empirical MSEs from the corresponding QIF version, with the use of a working AR-1 correlation structure.

Setting	N	QIF		QIF HS	Proposed QIF with t_N^2		Proposed QIF with \hat{t}_N^2	
		MSE ratio ¹	MSE ratio ²	MSE ratio	$\hat{E}(\hat{\rho}_N)$	MSE ratio	$\hat{E}(\hat{\rho}_N)$	MSE ratio
(1.1)	25	0.88	0.91	0.70	0.94	0.99	0.92	0.99
(1.2)	200	0.98	0.98	0.82	0.98	1.00	0.97	1.00
(1.3)	25	1.21	1.25	0.73	0.56	1.15	0.54	1.16
(1.4)	200	1.37	1.37	1.24	0.08	1.36	0.08	1.36
(2.1)	40	0.80	0.82	0.78	1.00	0.96	0.97	0.96
(2.2)	200	0.94	0.94	0.91	1.00	0.97	0.97	0.97
(2.3)	40	0.84	0.87	0.91	0.98	0.97	0.93	0.97
(2.4)	200	0.99	0.99	0.99	0.46	1.00	0.26	1.00
(3.1)	59	0.80	0.91	0.39	0.86	1.00	0.78	0.99
(3.2)	200	0.96	0.98	0.49	0.90	1.01	0.86	1.01
(3.3)	59	0.84	0.95	0.39	0.94	1.00	0.88	1.00
(3.4)	200	0.98	1.00	0.51	0.89	1.00	0.85	1.00

Empirical means of estimated weights, $\hat{E}(\hat{\rho}_N)$, given to \hat{M}_N in both versions of our proposed QIF approach are also given. In each scenario, the true correlation structure was AR-1 (exchangeable) for the first (last) two settings. QIF denotes results from the typical quadratic inference function approach. QIF HS denotes results from the QIF approach using the weighting matrix proposed by Han and Song [12]. t_N^2 and \hat{t}_N^2 are the estimators used to obtain $\hat{\rho}_N$ for our proposed QIF approach.

¹ N , number of independent subjects; MSE, mean squared error.

²Iteratively updating working parameter estimates throughout the estimation procedure inside C_N .

³Holding initial working parameter estimates fixed throughout the estimation procedure inside C_N .

We present results from three general longitudinal study designs and two general CRTs, utilizing four settings in each scenario. Tables I and II present results from the use of working AR-1 and exchangeable correlation structures, respectively. Each setting was examined via 1000 simulations. We obtained correlation and dispersion parameter estimates used in model-based covariances for our proposed weighting matrix from GEE to reduce simulation time, although estimates could be found iteratively as is performed with GEE. In scenarios 1–3, the true correlation structure was AR-1 (exchangeable) for the first (last) two settings. However, scenarios 4 and 5 are representative of CRTs in which the true and working correlation structures are exchangeable, as AR-1 is not feasible, and corresponding results are thus presented only in Table II.

In scenario 1, the marginal model is given by

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \epsilon_{ij}; \quad \epsilon_{ij} \sim N(0, 1); \quad j = 1, \dots, 10,$$

in which $x_{1ij} = j/10$ and $x_{2ij} \sim N(j/10, 1)$, $i = 1, \dots, N$, similar to a design presented by Qu *et al.* [4]. The number of subjects was 25 (200) for the first (second) and third (fourth) settings. The correlation parameter was 0.7, whereas $\beta = [0, 0.5, 1]^T$.

In scenario 2, the marginal model is given by

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij}; \quad j = 1, 2, 3, 4,$$

where π_{ij} is the marginal probability for the j th response from subject i , and $\text{logit}(\pi_{ij}) = \log[\pi_{ij}/(1 - \pi_{ij})]$, x_{1i} was given a value of 0 or 1, depending upon the arm of the trial to which the i th subject belonged, and $x_{2ij} \sim \text{Uniform}(0, 1)$ was generated independently from all observations within and across subjects. The number of subjects in each of two trial arms was 20 (100) for the first (second) and third (fourth) settings. The correlation parameter was 0.7, whereas $\beta = [0, 0, 0.1]^T$. These settings have similarities to those used by Song *et al.* [7]. We generated outcomes using the method presented by Qaqish [19].

Scenario 3 is a representation of our applied example in Section 5, using the marginal model given by

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4ij}; \quad j = 1, 2, 3, 4,$$

Table II. Ratios of MSEs from generalized estimating equations divided by empirical MSEs from the corresponding QIF version, with the use of a working exchangeable correlation structure.

Setting	N	QIF		QIF HS	Proposed QIF with t_N^2		Proposed QIF with \hat{t}_N^2	
		MSE ratio ¹	MSE ratio ²	MSE ratio	$\hat{E}(\hat{\rho}_N)$	MSE ratio	$\hat{E}(\hat{\rho}_N)$	MSE ratio
(1.1)	25	0.96	0.97	0.32	1.00	1.00	1.00	1.00
(1.2)	200	0.99	0.99	0.89	1.00	1.00	1.00	1.00
(1.3)	25	0.95	0.95	0.04	1.00	1.00	1.00	1.00
(1.4)	200	1.00	1.00	0.50	1.00	1.00	1.00	1.00
(2.1)	40	0.88	0.89	0.91	1.00	1.00	0.99	1.00
(2.2)	200	1.00	0.99	1.00	1.00	1.01	0.99	1.01
(2.3)	40	0.89	0.91	0.90	1.00	1.00	0.99	1.00
(2.4)	200	0.97	0.97	0.96	1.00	0.98	0.99	0.98
(3.1)	59	0.91	0.96	1.01	0.63	0.99	0.50	0.99
(3.2)	200	0.98	0.98	1.00	0.56	1.00	0.47	1.00
(3.3)	59	0.93	0.99	1.01	0.87	1.00	0.79	1.00
(3.4)	200	0.97	0.98	1.00	0.91	1.00	0.88	1.00
(4.1)	20	0.68	0.78	0.79	0.92	0.96	0.74	0.92
(4.2)	200	0.92	0.93	0.77	0.97	0.99	0.94	0.99
(4.3)	40	0.81	0.84	0.71	0.93	0.97	0.75	0.93
(4.4)	400	0.95	0.95	0.68	0.77	0.97	0.60	0.97
(5.1)	20	1.09	1.25	1.45	0.62	1.16	0.28	1.22
(5.2)	100	1.41	1.42	1.50	0.36	1.30	0.11	1.38
(5.3)	20	1.02	1.09	1.20	0.78	1.05	0.42	1.07
(5.4)	100	1.24	1.25	1.31	0.64	1.12	0.37	1.17

Empirical means of estimated weights, $\hat{E}(\hat{\rho}_N)$, given to \hat{M}_N in both versions of our proposed QIF approach are also given. In scenarios 1–3, the true correlation structure was AR-1 (exchangeable) for the first (last) two settings, whereas the true structure was exchangeable in scenarios 4 and 5. QIF denotes results from the typical quadratic inference function approach. QIF HS denotes results from the QIF approach using the weighting matrix proposed by Han and Song [12]. t_N^2 and \hat{t}_N^2 are the estimators used to obtain $\hat{\rho}_N$ for our proposed QIF approach.

N, number of independent subjects in scenarios 1–3 or the number of independent clusters in scenarios 4 and 5; MSE, mean squared error.

¹ Iteratively updating working parameter estimates throughout the estimation procedure inside C_N .

² Holding initial working parameter estimates fixed throughout the estimation procedure inside C_N .

where μ_{ij} , $j = 1, 2, 3, 4$; $i = 1, \dots, N$, denotes mean counts, x_{1i} is a subject-level indicator equal to 1 for 31 (105) subjects when $N = 59$ ($N = 200$), x_{2i} and x_{3i} are generated independently across subjects from $N(1.75, 0.5)$ and $\text{Uniform}(2.9, 3.7)$, respectively, and $x_{4ij} = j$. The correlation parameter was 0.3, whereas the marginal variance for the i th subject's j th outcome is $4\mu_{ij}$ and $\beta = [-2, 0, 1, 0.65, -0.05]^T$. We generated outcomes using a procedure given by Madsen and Dalthorp [20].

In scenarios 4 and 5, the marginal model is given by

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 x_{1i}; \quad j = 1, \dots, n_i.$$

We randomized an equal number of clusters to the intervention and control arms of the trial, and x_{1i} was an indicator for intervention assignment. We allowed the number of clusters to vary from 20 to 400 across settings. We independently generated cluster sizes from $\text{Uniform}(10, 50)$ and $\text{Uniform}(25, 150)$ in scenarios 4 and 5, respectively. In the first two settings of scenario 4, marginal probabilities and correlations were 0.1 and 0.05, respectively, across all clusters. In the last two settings, marginal probabilities (correlations) were 0.5 (0.3) and 0.3 (0.2) for control and intervention clusters, respectively. In scenario 5, marginal probabilities were 0.5. The exchangeable correlation value for the i th cluster was $\exp(\omega_1 + \omega_2 * n_i) / (1 + \exp(\omega_1 + \omega_2 * n_i))$, in which ω_1 and ω_2 were -0.05 (-1.5) and -0.025 (-0.02), respectively, in the first (last) two settings, allowing correlations to range from 0.02 to 0.34 (0.01 to 0.12). Westgate and Braun [8] utilized similar settings. We generated outcomes using the beta-binomial distribution.

4.2. Study results

Results show that the typical QIF approach can produce estimates with greater variability than the corresponding estimates from GEE that incorporate the same working covariance structure in settings consisting of a small to moderately sized sample. This result was also evident in some settings in which the QIF had a theoretical efficiency advantage due to the use of a misspecified working covariance structure. As also shown by Westgate (submitted), in multiple settings, the estimation procedure that does not iteratively update the parameter estimates inside C_N did improve upon the commonly used estimation procedure in which the estimates are updated throughout the procedure, although this did not guarantee a better performance relative to GEE. Alternatively, the use of our proposed weighting matrix with the QIF was in general shown to be superior. Specifically, when the typical QIF approach led to larger MSEs than GEE, corresponding MSEs from our proposed QIF were only negligibly different relative to the MSEs from GEE. Furthermore, when MSEs from the typical QIF were notably smaller than MSEs from GEE, our proposed QIF also produced MSEs that were superior to MSEs from GEE, although the typical QIF produced smaller MSEs in some of these settings.

Results show that the weighting matrix proposed by Han and Song [12] generally did not work as well as our proposed matrix, and the QIF approach that incorporates their weighting matrix was also detrimental relative to GEE and the typical QIF approach in many settings. Specifically, although the QIF approach that incorporates their weighting matrix performed best in scenario 5, it was typically detrimental relative to all other methods under comparison in scenario 3 when using a working AR-1 correlation structure and in scenarios 1 and 4. In other settings, although not always performing as well as our proposed matrix, the use of their weighting matrix typically led to empirical MSEs that were relatively close to the corresponding empirical MSEs from one or more of the methods under comparison.

The utility of our proposed weighting matrix with the QIF is most apparent in settings with smaller values for N in each scenario. However, for larger N , MSEs from the typical QIF were never too much larger than the corresponding MSEs from GEE and were notably smaller in some settings. Additionally, for the larger N , the use of our proposed weighting matrix with the QIF performed similarly to the typical QIF with either estimation procedure. Although this seems to imply that the choice between C_N and \hat{C}_N^* with the QIF does not matter for large N , we still suggest the use of \hat{C}_N^* . How large N needs to be to obtain similar MSE results with respect to these two weighting matrices will be unknown in practice, implying that the proposed matrix should be utilized to ensure that estimation precision is not lost relative to GEE, while still maintaining the possibility of gaining precision.

Both versions of our proposed weighting matrix for the QIF led to parameter estimates with almost equivalent MSEs in the majority of longitudinal study settings, although some small differences were evident in the CRT scenarios. This result occurred as estimated mean weights, $\hat{E}(\hat{\rho}_N)$, from these two

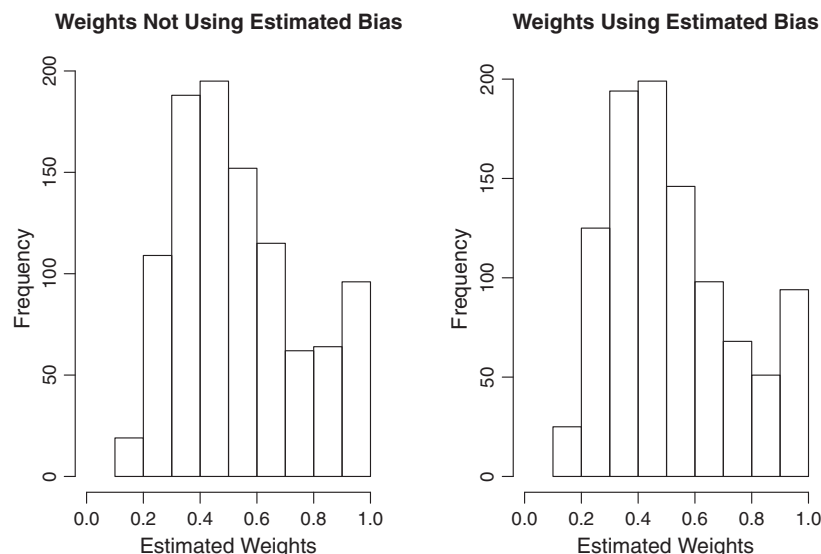


Figure 1. Values of $\hat{\rho}_N$, either using $(\hat{\tau}_N^2)$ or not using $(\hat{\tau}_N^2)$ the estimated bias of $\hat{\tau}_N^2$, from utilizing our proposed method with a working autoregressive-1 structure to analyze the 1000 simulated datasets from setting (1.3).

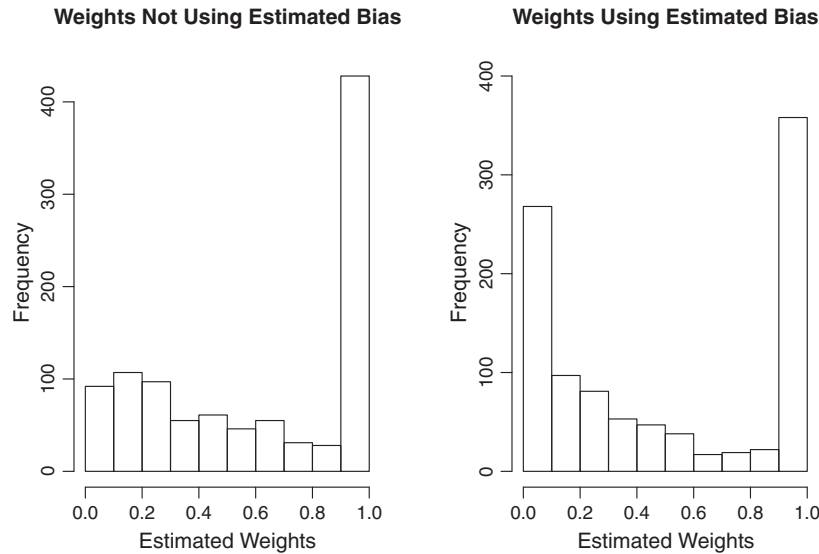


Figure 2. Values of $\hat{\rho}_N$, either using (\hat{t}_N^2) or not using (t_N^2) the estimated bias of \hat{t}_N^2 , from utilizing our proposed method with a working exchangeable structure to analyze the 1000 simulated datasets from setting (3.1).

methods either were very similar or had little influence in the longitudinal settings, whereas large variations in cluster sizes in the CRT settings increased the bias in \hat{t}_N^2 , and thus, notable differences in mean empirical weights were observed. As examples of what individual values look like, Figures 1 and 2 present histograms of $\hat{\rho}_N$ from settings (1.3) and (3.1), respectively. In scenarios 1 and 2, $\hat{E}(\hat{\rho}_N)$ was always close to 1 when the true and working correlation structures were AR-1. However, when the working structure was AR-1 but the true correlation was exchangeable, the amount of weight given to C_N in our proposed matrix increased with N because of the bias in the model-based covariance matrix and the decrease in variability in C_N . Furthermore, when a working exchangeable structure is used in scenarios 1 and 2, the ordinary QIF did not perform quite as well as GEE, and our proposed weighting matrix with the QIF performed as well as GEE because $\hat{\rho}_N = 1$ was typically observed. In scenario 3, $\hat{E}(\hat{\rho}_N)$ was large when the working covariance structure was correct, and $\hat{E}(\hat{\rho}_N)$ was slightly smaller for the larger N in settings of a misspecified structure. Similarly, in the third and fourth settings of scenario 4 and all settings of scenario 5, $\hat{E}(\hat{\rho}_N)$ decreased as N increased owing to C_N being less variable while utilizing an incorrect covariance structure.

5. Application

To illustrate differences between GEE and the QIF approaches, we analyze the dataset from the epileptic seizures study, adding to the results presented by Westgate [9]. We can find details about the study, in addition to the dataset, in the works of Thall and Vail [2] and Song [3]. In addition to subjects receiving standard chemotherapy, 28 served as controls whereas 31 subjects were randomized to receive a treatment drug. The number of seizures a subject had within each of four successional 2-week intervals served as the outcome of interest. Song [3] and Song *et al.* [7] fit the following marginal model to show that the QIF, relative to GEE, is robust to outliers:

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4ij}; \quad j = 1, 2, 3, 4.$$

Here, μ_{ij} denotes the mean number of seizures for the i th subject during the j th 2-week interval, x_{1i} is equal to 1 (0) if subject i received the drug (placebo), x_{2i} is the natural log of one quarter of the number of seizures the i th subject experienced throughout the 8 weeks prior to randomization, x_{3i} is the natural log of the i th subject's age, and $x_{4ij} = j$ denotes the j th 2-week interval.

Table III presents results from analyses, including and excluding a patient with outlying seizure counts [3, 7], using working AR-1 and exchangeable correlation structures. As discussed in Section 3.3,

Table III. Parameter, standard error (in parentheses), and weight estimates from analyses of the epileptic seizures dataset.

	Including the outlying subject					Excluding the outlying subject				
	GEE	QIF ¹	QIF ²	QIF ³	Proposed QIF ⁴	GEE	QIF ¹	QIF ²	QIF ³	Proposed QIF ⁴
Working AR-1 correlation structure										
$\hat{\beta}_1$	-0.02 (0.28)	-0.05 (0.17)	-0.04 (0.16)	-0.08 (0.31)	-0.03 (0.27)	-0.25 (0.16)	-0.28 (0.18)	-0.27 (0.16)	-0.27 (0.24)	-0.25 (0.16)
$\hat{\beta}_2$	1.25 (0.28)	1.19 (0.14)	1.20 (0.13)	1.15 (0.24)	1.24 (0.26)	0.99 (0.09)	0.96 (0.08)	0.97 (0.08)	0.93 (0.10)	0.99 (0.09)
$\hat{\beta}_3$	0.65 (0.34)	0.58 (0.36)	0.60 (0.37)	0.11 (0.74)	0.65 (0.33)	0.78 (0.28)	0.68 (0.37)	0.68 (0.34)	0.25 (0.76)	0.78 (0.27)
$\hat{\beta}_4$	-0.06 (0.04)	-0.05 (0.03)	-0.05 (0.03)	-0.07 (0.03)	-0.06 (0.03)	-0.05 (0.04)	-0.05 (0.04)	-0.05 (0.03)	-0.07 (0.03)	-0.05 (0.04)
$\hat{\rho}_N$				0.24	0.91				0.29	1.00
Working exchangeable correlation structure										
$\hat{\beta}_1$	-0.01 (0.27)	-0.02 (0.18)	-0.02 (0.18)	-0.02 (0.27)	-0.01 (0.27)	-0.23 (0.17)	-0.17 (0.17)	-0.17 (0.16)	-0.23 (0.17)	-0.23 (0.17)
$\hat{\beta}_2$	1.23 (0.26)	1.17 (0.17)	1.17 (0.16)	1.22 (0.25)	1.23 (0.26)	0.99 (0.09)	0.97 (0.09)	0.97 (0.08)	0.99 (0.09)	0.99 (0.09)
$\hat{\beta}_3$	0.60 (0.34)	0.51 (0.35)	0.52 (0.36)	0.58 (0.33)	0.60 (0.34)	0.72 (0.28)	0.62 (0.34)	0.65 (0.32)	0.70 (0.28)	0.72 (0.28)
$\hat{\beta}_4$	-0.06 (0.04)	-0.06 (0.03)	-0.06 (0.03)	-0.06 (0.04)	-0.06 (0.04)	-0.04 (0.04)	-0.03 (0.03)	-0.03 (0.03)	-0.04 (0.04)	-0.04 (0.04)
$\hat{\rho}_N$				0.26	1.00				0.20	1.00

$\hat{\rho}_N$ is either the estimated weight given to the model-based covariance matrix within the proposed QIF's weighting matrix or the estimate for $\mu_N I$ within the weighting matrix proposed by Han and Song [12] for the QIF.

AR, autoregressive; GEE, generalized estimating equations; QIF, quadratic inference function.

¹The typical QIF approach that uses the estimation procedure in which working parameter estimates are iteratively updated inside C_N .

²The typical QIF approach that uses the estimation procedure in which initial parameter estimates are fixed inside C_N .

³The QIF approach using the weighting matrix proposed by Han and Song [12].

⁴Choice of t_N^2 or \hat{t}_N^2 with the QIF's proposed weighting matrix led to equivalent results, except for the analysis in which $\hat{\rho}_N = 0.91$. Here, this is the estimated weight using t_N^2 . When \hat{t}_N^2 is used, $\hat{\rho}_N = 0.89$.

standard error (SE) estimates utilize a correction based on the work of Mancl and DeRouen [17]. Results are similar to results seen in the simulation study of scenario 3. Specifically, only small, but notable, differences existed in parameter and SE estimates from GEE and both the typical and proposed QIF approaches. Furthermore, relative to the other methods and when a working AR-1 correlation structure is incorporated, use of the weighting matrix proposed by Han and Song [12] gave notably larger SE estimates for $\hat{\beta}_1$ and $\hat{\beta}_3$ and distinctly smaller values for $\hat{\beta}_2$. Additionally, large values of $\hat{\rho}_N$ were utilized with our proposed weighting matrix for the QIF approach, resulting in estimates that were similar to or the same as parameter and SE estimates from GEE. The use of four covariates in the model is partially the reason for these large values of $\hat{\rho}_N$, as the dimension of, and thus variability within, C_N increases with the number of covariates [8], which leads to an increase in the allocated weight to the model-based matrix.

It is interesting to note that when the outlying seizure counts from the one patient are included, use of the typical QIF resulted in smaller SE estimates for $\hat{\beta}_1$ and $\hat{\beta}_2$ as compared with GEE and use of the proposed QIF, likely because of the original QIF's robustness to this one subject's outcomes. This result suggests that the proposed QIF will not be as robust to outliers as the original QIF, particularly when $\hat{\rho}_N > 0$. However, after excluding this subjects' observations from the analyses, SE estimates for $\hat{\beta}_3$ were smaller for GEE and the proposed QIF approach, whereas only negligible differences were observed for other SE estimates. We do caution, however, that we cannot make decisive conclusions with respect to estimation precision for the different analyses of this dataset because results are based on estimated, rather than true, SEs.

6. Concluding remarks

Although the QIF method has a theoretical efficiency advantage over GEE, its estimation performance may actually be inferior in some finite-sample settings. To improve the QIF method in this regard, we proposed a new weighting covariance matrix, which is related to the weighting matrix proposed by Han and Song [12], that is a weighted combination of corresponding model-based and empirical covariance matrices, optimally taking their respective bias and variability into account. Specifically, the utilized weight minimizes the expected quadratic loss of the proposed matrix and allows for the implementation of an asymptotically optimal weighting matrix, thus maintaining the QIF's theoretical efficiency.

In practice, the optimal weight, ρ_N , needs to be estimated, and a simulation study covering general longitudinal study and CRT scenarios showed, overall, that the proposed QIF approach was superior to the typical QIF, the QIF of Han and Song [12], and GEE. Specifically, MSEs from the proposed QIF were always similar to or smaller than the corresponding MSEs from GEE. Furthermore, although in some settings the MSEs from the typical QIF were slightly smaller than MSEs from the proposed QIF, the typical QIF's estimation performance was notably inferior to that of GEE's and our proposed QIF's in many other settings in which N was not large. Additionally, even for large N , we caution against the general usage of the weighting matrix proposed by Han and Song [12], as this QIF approach was detrimental relative to the other methods in multiple settings, while only performing better than our proposed weighting matrix in one scenario, as their weighting matrix gives weight to a matrix that does not model the true covariance matrix. In practice, we do not know how large N must be such that the typical QIF and our proposed version produce estimates with similar MSEs, and we therefore suggest the use of our proposed weighting matrix for any N . Additionally, as the sample estimate for the variability in C_N contains bias, two different methods for obtaining $\hat{\rho}_N$ with our proposed weighting matrix were suggested. However, no notable differences were seen in terms of estimation performance, except in some CRT settings, implying that either approach can be utilized.

How robust the proposed QIF method will be to outliers will depend on how much estimated weight is given to the model-based covariance matrix. Specifically, this approach will be more (less) robust to outliers relative to GEE (the typical QIF) when this weight is less than 1 (greater than 0). Furthermore, as both versions of the proposed weighting matrix are asymptotically optimal, the QIF's large-sample properties are maintained, and the inference function values can still be used as test statistics. Future research is needed to determine the validity of the proposed QIF as a test statistic, however, as it partially employs a covariance matrix that could be biased.

An R function that implements the proposed QIF approach can be obtained by contacting the author at philip.westgate@uky.edu.

Appendix

General forms for given proofs come from the work of Han and Song [12] and Ledoit and Wolf [15]. Following Han and Song [12], we let c be a finite, generic constant that can change in value. Additionally, define $\langle \mathbf{K}_1, \mathbf{K}_2 \rangle = \text{tr}(\mathbf{K}_1 \mathbf{K}_2^T) / p$ for some arbitrary $p \times p$ matrices \mathbf{K}_1 and \mathbf{K}_2 [15].

All proofs assume that $\|\hat{\mathbf{M}}_N - \Sigma_N\|^2 > 0$ and $E[\|\hat{\mathbf{M}}_N - C_N(\boldsymbol{\beta})\|^2]$ does not converge to 0. Specifically, we assume that the working covariance structure is misspecified in some manner. In reality, we have $N < \infty$, and even if the covariance structure is correctly specified for all data, our method will still work. Additionally, if $E[\|\hat{\mathbf{M}}_N - C_N(\boldsymbol{\beta})\|^2] \rightarrow 0$, for large N , it would not make a difference how much weight is given to $\hat{\mathbf{M}}_N$ and $C_N(\boldsymbol{\beta})$.

The proof that $\rho_N = \tau_N^2 / (\alpha_N^2 + \tau_N^2)$ minimizes $E[\|C_N^* - \Sigma_N\|^2]$ closely resembles the corresponding proof given by Ledoit and Wolf [15].

The following proofs are based on the lemma given by Han and Song [12] and its corresponding conditions. The first two parts of their lemma are equivalent in our scenario, whereas $\delta_N^2 = \alpha_N^2 + \tau_N^2$ is easily verified. Additionally, using arguments similar to Han and Song [12], we can show that $\|\hat{\mathbf{M}}_N\|$, δ_N^2 , and α_N^2 remain bounded, whereas Han and Song [12] show that τ_N^2 is bounded and $\tau_N^2 \rightarrow 0$.

We now prove that $d_N^2 - \delta_N^2$ converges in quadratic mean to 0 as $N \rightarrow \infty$, under the assumption that $E[\|\hat{\mathbf{M}}_N - \mathbf{M}_N\|^4] \rightarrow 0$ as $N \rightarrow \infty$. Han and Song [12] proved that $E[(\bar{\tau}_N^2 - \tau_N^2)^2] \rightarrow 0$ as $N \rightarrow \infty$, and the proof that $E[(\bar{t}_N^2 - \tau_N^2)^2] \rightarrow 0$ as $N \rightarrow \infty$ then follows from Ledoit and Wolf [15]. The derivation for the bias in \bar{t}_N^2 and the proof that $E[(\hat{t}_N^2 - \tau_N^2)^2] \rightarrow 0$ as $N \rightarrow \infty$ are given at the end of this Appendix. The fact that $E[(\hat{a}_N^2 - \alpha_N^2)^2] \rightarrow 0$ then follows.

Proof

$d_N^2 - \delta_N^2 = (\|C_N(\boldsymbol{\beta}) - \mathbf{M}_N\|^2 - E[\|C_N(\boldsymbol{\beta}) - \mathbf{M}_N\|^2]) + \|\hat{\mathbf{M}}_N - \mathbf{M}_N\|^2 - 2 \langle C_N(\boldsymbol{\beta}) - \mathbf{M}_N, \hat{\mathbf{M}}_N - \mathbf{M}_N \rangle$. Following Han and Song [12], we show that the expected value of the square of each of these three terms goes to 0, thus proving $E[(d_N^2 - \delta_N^2)^2] \rightarrow 0$. With respect to the first term, $\|C_N(\boldsymbol{\beta}) - \mathbf{M}_N\|^2 - E[\|C_N(\boldsymbol{\beta}) - \mathbf{M}_N\|^2] = [\|C_N(\boldsymbol{\beta})\|^2 - E[\|C_N(\boldsymbol{\beta})\|^2]] - 2 \langle C_N(\boldsymbol{\beta}) - \boldsymbol{\Sigma}_N, \mathbf{M}_N \rangle$. Han and Song [12] proved that $E[(\|C_N(\boldsymbol{\beta})\|^2 - E[\|C_N(\boldsymbol{\beta})\|^2])^2] \rightarrow 0$. When the Cauchy–Schwarz inequality is used, $E[(2 \langle C_N(\boldsymbol{\beta}) - \boldsymbol{\Sigma}_N, \mathbf{M}_N \rangle)^2] \rightarrow 0$ as $N \rightarrow \infty$. The expected value of the square of the second term converges to 0 by assumption. Finally, when the Cauchy–Schwarz inequality is used, $E[(\langle C_N(\boldsymbol{\beta}) - \mathbf{M}_N, \hat{\mathbf{M}}_N - \mathbf{M}_N \rangle)^2] \rightarrow 0$. \square

We now prove two theorems similar to those given by Han and Song [12]. We note that five conditions given by Han and Song [12] must be met for the previous proofs and both theorems to be valid.

Theorem 1

For $\boldsymbol{\beta} \in \mathbf{B}$, $E[\|C_N^* - \boldsymbol{\Sigma}_N\|^2] \rightarrow 0$ as $N \rightarrow \infty$, implying that $C_N^* - \boldsymbol{\Sigma}_N \rightarrow 0$.

Proof of theorem 1

$0 \leq \|C_N^* - \boldsymbol{\Sigma}_N\|^2 \leq (\tau_N^2/\delta_N^2)^2 \|\mathbf{M}_N - \boldsymbol{\Sigma}_N\|^2 + \|C_N(\boldsymbol{\beta}) - \boldsymbol{\Sigma}_N\|^2 + 2(\tau_N^2/\delta_N^2)(\alpha_N^2/\delta_N^2 \langle \mathbf{M}_N - \boldsymbol{\Sigma}_N, C_N(\boldsymbol{\beta}) - \boldsymbol{\Sigma}_N \rangle)$, which has an expectation that converges to 0 because $E[(\tau_N^2/\delta_N^2)^2 \|\mathbf{M}_N - \boldsymbol{\Sigma}_N\|^2] = \tau_N^2 (\tau_N^2/\delta_N^2) (\alpha_N^2/\delta_N^2) \leq \tau_N^2 \rightarrow 0$ and $E[2(\tau_N^2/\delta_N^2)(\alpha_N^2/\delta_N^2) \langle \mathbf{M}_N - \boldsymbol{\Sigma}_N, C_N(\boldsymbol{\beta}) - \boldsymbol{\Sigma}_N \rangle] = 0$. \square

Theorem 2

For $\boldsymbol{\beta} \in \mathbf{B}$, $E[\|\hat{C}_N^* - \boldsymbol{\Sigma}_N\|^2] \rightarrow 0$ as $N \rightarrow \infty$, implying that $\hat{C}_N^* - \boldsymbol{\Sigma}_N \rightarrow 0$.

Proof of theorem 2

Following Ledoit and Wolf [15], $0 \leq \|\hat{C}_N^* - C_N^*\|^2 \leq \|\hat{\mathbf{M}}_N - \mathbf{M}_N\|^2 + ((a_N^2/d_N^2) - (\alpha_N^2/\delta_N^2))^2 d_N^2 + 2(\tau_N^2/\delta_N^2)((a_N^2/d_N^2) - (\alpha_N^2/\delta_N^2)) \langle \hat{\mathbf{M}}_N - \mathbf{M}_N, C_N(\boldsymbol{\beta}) - \hat{\mathbf{M}}_N \rangle$. We now show that $E[\|\hat{C}_N^* - C_N^*\|^2] \rightarrow 0$: by assumption, $E[\|\hat{\mathbf{M}}_N - \mathbf{M}_N\|^2] \rightarrow 0$. We have

$$\left(\frac{a_N^2}{d_N^2} - \frac{\alpha_N^2}{\delta_N^2}\right)^2 d_N^2 = (a_N^2 \delta_N^2 - \alpha_N^2 d_N^2)^2 / (d_N^4 \delta_N^4),$$

and Ledoit and Wolf [15] proved that $E[(a_N^2 \delta_N^2 - \alpha_N^2 d_N^2)^2 / (d_N^4 \delta_N^4)] \rightarrow 0$ as $N \rightarrow \infty$. Finally, $E[2(\tau_N^2/\delta_N^2)(a_N^2/d_N^2 - \alpha_N^2/\delta_N^2) \langle \hat{\mathbf{M}}_N - \mathbf{M}_N, C_N(\boldsymbol{\beta}) - \hat{\mathbf{M}}_N \rangle] \rightarrow 0$ by the Cauchy–Schwarz inequality and the following: $0 \leq E[\langle \hat{\mathbf{M}}_N - \mathbf{M}_N, C_N(\boldsymbol{\beta}) - \hat{\mathbf{M}}_N \rangle^2] \leq E[\|\hat{\mathbf{M}}_N - \mathbf{M}_N\|^2 \|C_N(\boldsymbol{\beta}) - \hat{\mathbf{M}}_N\|^2] \leq \sqrt{E[\|\hat{\mathbf{M}}_N - \mathbf{M}_N\|^4]} \sqrt{E[\|C_N(\boldsymbol{\beta}) - \hat{\mathbf{M}}_N\|^4]} \leq c \sqrt{E[\|\hat{\mathbf{M}}_N - \mathbf{M}_N\|^4]} \rightarrow 0$ by assumption, and $0 \leq 4(\tau_N^2/\delta_N^2)^2 E[(a_N^2/d_N^2 - \alpha_N^2/\delta_N^2)^2] \rightarrow 0$.

We have now shown that $E[\|\hat{C}_N^* - C_N^*\|^2] \rightarrow 0$ as $N \rightarrow \infty$. Using this in conjunction with Theorem 1 and the proof given by Han and Song [12], we have $E[\|\hat{C}_N^* - \boldsymbol{\Sigma}_N\|^2] \rightarrow 0$. \square

We now derive the bias in $\bar{\tau}_N^2$ and then prove that $E[(\bar{\tau}_N^2 - \tau_N^2)^2] \rightarrow 0$ as $N \rightarrow \infty$.

Let $\bar{x}_N = (1/N) \sum_{i=1}^N x_i$ represent any one of the $(mp)^2$ elements comprising $C_N(\boldsymbol{\beta})$, the sample average over the N ESEs, and $\mu_N = E[\bar{x}_N] = (1/N) \sum_{i=1}^N \mu_i$. We have $\text{var}(\bar{x}_N) = (1/N^2) \sum_{i=1}^N \text{var}(x_i) = (1/N^2) \sum_{i=1}^N \sigma_i^2$, $\bar{\tau}_N^2$ estimates $\text{var}(\bar{x}_N)$ with $(1/N^2) \sum_{i=1}^N (x_i - \bar{x}_N)^2$, and

$$(1/N^2) \sum_{i=1}^N E[(x_i - \bar{x}_N)^2] = [(N-1)/N] \text{Var}(\bar{x}_N) + (1/N^2) \sum_{i=1}^N \mu_i^2 - (1/N^3) \left(\sum_{i=1}^N \mu_i\right)^2.$$

As $(N - 1)/N \approx 1$, we ignore the first term, and therefore we can approximate the bias of the corresponding arbitrary element within $\bar{\tau}_N^2$ by using $(1/N^2) \sum_{i=1}^N \mu_i^2 - (1/N^3) \left(\sum_{i=1}^N \mu_i \right)^2$. Because we need an estimate for the sum of the variances for each of the $(2p)^2$ elements comprising $C_N(\boldsymbol{\beta})$, we need to sum the biases from each of the variance estimates, leading to $\text{Bias}(\bar{\tau}_N^2) \approx (1/N^2) \sum_{i=1}^N \|\text{Cov}[g_i(\boldsymbol{\beta})]\|^2 - (1/N^3) \|\sum_{i=1}^N \text{Cov}[g_i(\boldsymbol{\beta})]\|^2$. We can verify that the bias terms go to 0 as $N \rightarrow \infty$ when using model-based covariances.

We now prove $E \left[(\hat{\tau}_N^2 - \tau_N^2)^2 \right] \rightarrow 0$ as $N \rightarrow \infty$.

Proof

We have already established that $E \left[\bar{\tau}_N^2 - \widehat{\text{Bias}}(\bar{\tau}_N^2) \right] = E \left[\bar{\tau}_N^2 \right] - \widehat{\text{Bias}}(\bar{\tau}_N^2) \rightarrow 0$.

$$\begin{aligned} \left[\bar{\tau}_N^2 - \widehat{\text{Bias}}(\bar{\tau}_N^2) \right]^2 &= (\bar{\tau}_N^2)^2 + \left[(1/N^2) \sum_{i=1}^N \|\mathbf{B}_i \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{B}_i^T\|^2 \right]^2 + (1/N^2) \|\mathbf{M}_N\|^4 \\ &\quad - 2\bar{\tau}_N^2 \left[(1/N^2) \sum_{i=1}^N \|\mathbf{B}_i \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{B}_i^T\|^2 \right] + 2\bar{\tau}_N^2 (1/N) \|\mathbf{M}_N\|^2 \\ &\quad - 2 \left[(1/N^2) \sum_{i=1}^N \|\mathbf{B}_i \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{B}_i^T\|^2 \right] [(1/N) \|\mathbf{M}_N\|^2] \end{aligned}$$

To prove that $E \left(\left[\bar{\tau}_N^2 - \widehat{\text{Bias}}(\bar{\tau}_N^2) \right]^2 \right) \rightarrow 0$, we need to show that the expected value of each of these six terms goes to 0 as $N \rightarrow \infty$. The second, third, and sixth terms are each comprised of two terms that do not contain random variables and have been shown to go to 0, therefore implying that these three terms go to 0. Han and Song [12] showed that $E \left[(\bar{\tau}_N^2)^2 \right] \rightarrow 0$. The fourth and fifth terms are comprised of a respective bias term and $\bar{\tau}_N^2$. As we have already shown that the bias terms are bounded and both the bias terms and $\bar{\tau}_N^2$ go to 0, we therefore have the result that the fourth and fifth terms also go to 0.

Following Han and Song [12], $E \left[\left(\bar{\tau}_N^2 - \widehat{\text{Bias}}(\bar{\tau}_N^2) - \tau_N^2 \right)^2 \right] = E \left(\left[\bar{\tau}_N^2 - \widehat{\text{Bias}}(\bar{\tau}_N^2) \right]^2 \right) - 2\tau_N^2 E \left[\bar{\tau}_N^2 - \widehat{\text{Bias}}(\bar{\tau}_N^2) \right] + (\tau_N^2)^2 \rightarrow 0$ as $N \rightarrow \infty$, and the work by Ledoit and Wolf [15] proves that $E \left[(\hat{\tau}_N^2 - \tau_N^2)^2 \right] \rightarrow 0$ as $N \rightarrow \infty$. \square

Acknowledgements

We would like to thank Dr. Peter Song and Peisong Han for their very helpful advice with respect to our work.

References

1. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
2. Thall PF, Vail SC. Some covariance models for longitudinal count data with overdispersion. *Biometrics* 1990; **46**:657–671.
3. Song PX-K. *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer: New York, 2007.
4. Qu A, Lindsay BG, Li B. Improving generalised estimating equations using quadratic inference functions. *Biometrika* 2000; **87**:823–836.
5. Hansen LP. Large sample properties of generalized method of moments estimators. *Econometrica* 1982; **50**:1029–1054.
6. Qu A, Song PX-K. Assessing robustness of generalised estimating equations and quadratic inference functions. *Biometrika* 2004; **91**:447–459.
7. Song PX-K, Jiang Z, Park E, Qu A. Quadratic inference functions in marginal models for longitudinal data. *Statistics in Medicine* 2009; **28**:3683–3696.
8. Westgate PM, Braun TM. The effect of cluster size imbalance and covariates on the estimation performance of quadratic inference functions. *Statistics in Medicine* 2012; **31**:2209–2222.
9. Westgate PM. A bias-corrected covariance estimate for improved inference with quadratic inference functions. *Statistics in Medicine* 2012; **31**:4003–4022. DOI: 10.1002/sim.5479.
10. Windmeijer F. A finite sample correction for the variance of linear two-step GMM estimators. *Working Paper Series No. W00/19*, Institute for Fiscal Studies, London, 2000.
11. Windmeijer F. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 2005; **126**:25–51.

12. Han P, Song PX-K. A note on improving quadratic inference functions using a linear shrinkage approach. *Statistics and Probability Letters* 2011; **81**:438–445.
13. Lindsay BG. Conditional score functions: some optimality results. *Biometrika* 1982; **69**:503–512.
14. Small CG, McLeish DL. *Hilbert Space Methods in Probability and Statistical Inference*. Wiley: New York, 1994.
15. Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 2004; **88**:365–411.
16. Loader C, Pilla RS. Iteratively reweighted generalized least squares for estimation and testing with correlated data: an inference function framework. *Journal of Computational and Graphical Statistics* 2007; **16**:925–945.
17. Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics* 2001; **57**:126–134.
18. Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 2001; **96**:1387–1396.
19. Qaqish BF. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* 2003; **90**:455–463.
20. Madsen L, Dalthorp D. Simulating correlated count data. *Environmental and Ecological Statistics* 2007; **14**:129–148.