

Quadratic inference functions in marginal models for longitudinal data

Peter X.-K. Song^{1,*}, Zhichang Jiang², Eunjoo Park³ and Annie Qu⁴

¹*Department of Biostatistics, UM School of Public Health, University of Michigan, 1420 Washington Heights,
Ann Arbor, MI 48109-2029, U.S.A.*

²*Alberta Cancer Board, Edmonton, AB, Canada T6G 1Z2*

³*St. Paul's Hospital, Vancouver, BC, Canada V6Z 1T6*

⁴*Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, U.S.A.*

SUMMARY

The quadratic inference function (QIF) is a new statistical methodology developed for the estimation and inference in longitudinal data analysis using marginal models. This method is an alternative to the popular generalized estimating equations approach, and it has several useful properties such as robustness, a goodness-of-fit test and model selection. This paper presents an introductory review of the QIF, with a strong emphasis on its applications. In particular, a recently developed SAS MACRO QIF is illustrated in this paper to obtain numerical results. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: efficiency; GEE; goodness-of-fit; model selection; robustness; SAS macro

1. INTRODUCTION

Longitudinal data are pervasive in medical and health sciences. Sequentially observed over time, longitudinal data are collected either from an observational study or a designed experiment in which response variables pertain to a sequence of events with outcomes recorded at certain time-points during a study period. One primary objective of a longitudinal study is to investigate how the variability of the responses may be addressed in time with covariates. For instance, a longitudinal clinical trial is often used to examine time-varying drug efficacy in treating a disease, which cannot be studied by a cross-sectional study. Using longitudinal data, one can separate the so-called cohort

*Correspondence to: Peter X.-K. Song, Department of Biostatistics, UM School of Public Health, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, U.S.A.

†E-mail: pxsong@umich.edu

Contract/grant sponsor: NSF

effects and age (or time) effects, by characterizing changes within each individual in reference to the subject's baseline status.

There are three important methodologies widely used to analyze longitudinal data, including marginal models, mixed-effects models and transition models. Each modeling approach serves specific analytic purposes to answer relevant scientific questions. This paper focuses on the methodology of the marginal model, which primarily aims to estimate the population-average effects of covariates on the response of interest. The celebrated method of statistical inference among marginal models is the generalized estimating equation (GEE) [1] approach proposed by Liang and Zeger in 1986. This method has gained considerable popularity in the analysis of longitudinal data in the past two decades or so. However, the application of the GEE method is restricted by some critical underlying assumptions, outlined as follows.

First, data to fit the marginal model are assumed to be sampled from a relatively homogeneous population, in the sense that the variation in the response is mostly due to differences in covariates as well as sampling noise, but not due to a subject-specific variation (or population heterogeneity). Second, the first moment mean model is assumed to be correctly specified, which is a critical condition to ensure the consistency of the GEE estimation. Third, the nuisance correlation parameter is assumed to be properly estimated; root- n consistency is required in order for the GEE estimator to have valid asymptotic normality. The current version of SAS PROC GENMOD is built on a procedure in which the correlation parameters are estimated separately from the GEE estimation of the regression coefficients. From a theoretical point of view, several papers such as [2, 3] have pointed out that the GEE might fail to produce consistent estimators if the correlation parameters are not consistently estimated. In practice, however, this inconsistency rarely happens. Fourth, missing values in the data are assumed to be missing completely at random (MCAR).

In addition, the GEE method is very sensitive to outliers or contaminated data (e.g. [4, 5]), which is worrisome in practice. Another practical shortcoming of the GEE method is that model selection is complicated by the lack of an objective function in its estimation procedure. Pan's [6] Akaike Information Criterion-(AIC) type model selection criterion for the GEE method is derived under the quasi-likelihood function ignoring the serial correlation.

To overcome some of the difficulties in the use of the GEE method, many improvements have been proposed in the literature. Among these improvements, the quadratic inference function (QIF) approach proposed by Qu *et al.* [7] has received some attention in the recent years, as it improved on the GEE method in several aspects simultaneously, summarized as follows:

- (i) The QIF provides statistical inference under the same model assumptions as the GEE requires. Like the GEE method, a misspecified working correlation does not affect the consistency of the regression parameter estimation, and the QIF provides a robust sandwich estimator for the variance of the regression parameter estimator. When the working correlation structure is correctly specified, both the QIF and GEE are equally efficient. However, when the working correlation structure is misspecified, the QIF is more efficient than the GEE. According to Fitzmaurice *et al.* [8], misspecification of the working correlation affects the GEE's efficiency of regression parameter estimation, especially when time-varying covariates are included in the analysis.
- (ii) The QIF provides a goodness-of-fit test for checking the validity of the assumption for the first moment mean model. This is a crucial condition for the estimation consistency that the GEE requires but is unable to validate easily. Hardin and Hilbe [9] introduced several summary statistics similar to those considered in the GLMs under the independence

- working correlation. Horton *et al.* [10] proposed an *ad hoc* goodness-of-fit chi-square test for binomial models via a deciles-based partition. The difficulty of developing a general goodness-of-fit test lies in the fact that the GEE does not have an objective function in its estimation procedure. According to [11], multiple roots could be a potential issue for the GEE. For example, if data are sampled from a mixture of multiple normals, the multimodality can lead to a multiple-root issue. Although this does not happen very often, the QIF would avoid this potential problem as the QIF estimation is carried out through global minimization.
- (iii) The QIF is robust against outliers or contaminated data since it has been proved to have a bounded influence function [4]. In contrast, the GEE has an unbounded influence function, and consequently it is highly sensitive to abnormal observations. This kind of robustness property is appealing, as outliers occur frequently in practice. Making a decision on whether outliers should be included or excluded from a model fit, in the case of the GEE, is not trivial and can be sometimes confusing, especially when the two options result in different statistical conclusions. Sometimes, a procedure that is sensitive to deletions may be useful to detect influential cases and learn some interesting features about the data.
 - (iv) The QIF is built upon an objective function, which is analogous to twice the negative log-likelihood. See also Hall [12]. This naturally leads to the formulation of model selection criteria such as the one similar to the AIC or the Bayes Information Criterion (BIC) in estimating equation settings. Moreover, using the proposed criterion, one can proceed in a stepwise regression-type analysis to trim away unimportant covariates and reach a model with a balance of parsimony and goodness-of-fit. This sequential model selection procedure is widely used in practice, but is unavailable in the GEE method. Currently, this procedure can be applied with the QIF method manually, and an automatic procedure will be made available in the SAS QIF macro in the near future.

This paper aims to provide an introductory review of the QIF, with extensive numerical illustrations on the application of this method to longitudinal data analysis. Various simulation results are included to numerically demonstrate the properties of the QIF discussed above, where its comparison with the GEE method is quantified. In particular, in a data analysis example, we demonstrate the features of a SAS MACRO developed recently by Song and Jiang at the University of Waterloo. We hope that the availability of the QIF software will attract more practitioners to apply this new powerful method.

This paper is organized as follows. Section 2 presents a brief introduction to the QIF method, including simulation results. Section 3 focuses on the applications of the QIF, including the SAS MACRO QIF, and a data example to illustrate the use of the macro. Section 4 contains concluding remarks.

2. QUADRATIC INFERENCE FUNCTIONS

The main motivation behind the QIF is to improve on the efficiency of the GEE and to provide an inference function for hypothesis testing. Like the GEE method, the QIF is useful when the inference of the population-average effect is the main interest, and it only requires specifying the first two moments of the likelihood function.

2.1. Estimation

Let y_{it} be an outcome variable and x_{it} be a $q \times 1$ vector of covariates, corresponding to observations recorded at times $t = 1, \dots, n_i$ for subjects $i = 1, \dots, N$. Assume that the observations from different subjects are independent, but those within the same subject are dependent. A marginal approach such as the GEE assumes that the marginal mean μ_{ij} is a function of the covariates through a link function g with $g(\mu_{ij}) = x'_{ij}\beta$, and the variance of y_{ij} is a function of the mean $\text{var}(y_{ij}) = \phi V(\mu_{ij})$, where ϕ is the dispersion parameter. The GEE solves the equation

$$\sum_{i=1}^N \dot{\mu}_i' V_i^{-1} (y_i - \mu_i) = 0 \quad (1)$$

where $\dot{\mu}_i = \partial \mu_i / \partial \beta$ is an $n_i \times q$ matrix, and $V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2}$ with A_i being the diagonal matrix of the marginal variances, $\text{var}(y_{ij})$, and $R_i(\alpha)$ being the working correlation matrix.

The QIF [7] is derived by observing that the inverse of the working correlation matrix can be approximated by a linear combination of several basis matrices:

$$R^{-1} \approx \sum_{l=0}^k a_l M_l \quad (2)$$

where M_0 is the identity matrix, M_1, \dots, M_k are known basis matrices with 0 or 1 as components and a_0, \dots, a_k are unknown coefficients. This expression (2) holds exactly for some common working correlation structures. For example, if the working correlation is exchangeable, then $R^{-1} = a_0 I + a_1 M_1$, where M_1 is 0 on the diagonal and 1 elsewhere. If it is an AR-1 working correlation, $R^{-1} = a_0^* I + a_1^* M_1^* + a_2^* M_2^*$, where M_1^* has 1 on the sub-diagonal and 0 elsewhere and M_2^* has 1 on the two corner components of the diagonal. In cases when it is difficult to determine a suitable working correlation structure, one can also choose a hybrid working correlation that combines basis matrices from several working correlations [13]. The basis matrices corresponding to the unstructured correlation structure is given by Qu and Lindsay [14].

Plugging expression (2) into the GEE in (1) leads to a linear combination of the elements of the following extended score vector:

$$\bar{g}_N(\beta) = \frac{1}{N} \sum_{i=1}^N g_i(\beta) \approx \frac{1}{N} \begin{pmatrix} \sum_{i=1}^N \dot{\mu}_i' A_i^{-1} (y_i - \mu_i) \\ \vdots \\ \sum_{i=1}^N \dot{\mu}_i' A_i^{-1/2} M_k A_i^{-1/2} (y_i - \mu_i) \end{pmatrix}$$

That is, the GEE expression (1) takes the form of $N a' \bar{g}_N(\beta)$ with $a = (a_0, a_1, \dots, a_k)'$ being the vector of the coefficients in the expansion (2). As there are more equations than unknown parameters, the generalized method of moments [13] can be applied by minimizing the QIF:

$$Q_N(\beta) = N \bar{g}_N(\beta) C_N^{-1}(\beta) \bar{g}_N(\beta) \quad (3)$$

where $C_N(\beta) = N^{-1} \sum_{i=1}^N g_i(\beta) g_i'(\beta)$ is the sample covariance matrix. It is worth pointing out that the objective function defined in (3) contains only the regression parameter β , and only the basis

matrices from the working correlation structure are used to formulate this function. In other words, the QIF estimator, $\hat{\beta} = \arg \min_{\beta} Q_N(\beta)$, is obtained with no need to estimate the nuisance correlation parameter. Hence, the QIF method does not rely on whether an appropriate estimation of the correlation parameter is available or not. In addition, since the estimate is obtained by minimizing a quadratic function, there exists a unique global solution, which avoids the multiple-root problem.

The QIF estimator $\hat{\beta}$ has the usual large sample properties, namely, it is \sqrt{N} -consistent, with the asymptotic variance proportional to the inverse of the Godambe information matrix (or the sandwich estimator), $(D'\Sigma^{-1}D)^{-1}$, where the sensitivity matrix is $D = E\{\partial g_i(\beta)/\partial \beta\}$, the variability matrix is $\Sigma = E\{g_i(\beta)g_i'(\beta)\}$ and g_i is the extended score with a single observation. If the dimension of the extended score is the same as the dimension of the parameters, then minimizing the QIF is equivalent to solving $\bar{g}_N(\beta) = 0$, and consequently, the asymptotic covariance matrix of the resulting estimator is $(D^{-1}\Sigma D^{-1})/N$, namely, the robust sandwich estimator.

The asymptotic variance of the QIF estimator attains the minimum in the sense of the Löwner ordering [15]. This property ensures that the QIF improves the efficiency of the GEE when the working correlation is misspecified and remains as efficient as the GEE when the working correlation is correct. Qu *et al.* [7] conducted a simple simulation study based on a 10-dimensional multivariate normal model, in which the relative efficiency between the GEE and QIF was examined. This simulation study found that under certain misspecified working correlations, the improvement of the QIF in efficiency may be more than twice the GEE approach. Without making any extra model assumptions, the QIF can substantially improve the estimation efficiency over the GEE. It is noted that when the correlation structure is correctly specified in the setting of the multivariate normal model, the resulting GEE estimation is identical to the maximum likelihood estimation; hence, the GEE will achieve the best efficiency. The efficiency gain in cases of misspecified correlation structures results simply from the best linear combination of moment conditions, which avoids the estimation of the nuisance correlation parameter. In practice, the true correlation is never known; therefore, the QIF appears to be more appealing as far as estimation efficiency is concerned. In addition, the intracluster correlation can be estimated as part of the data analysis using, for example, the same estimators in the GEE, if needed.

2.2. Testing hypothesis

The QIF can also be regarded as an inference function. Because it is an analog to twice the negative log likelihood, it has properties similar to the likelihood ratio test. For example, to test $H_0: \beta = \beta_0$, one can use the test statistic $Q(\beta_0) - Q(\hat{\beta})$ since it follows a chi-square distribution asymptotically with $q = \dim(\beta)$ degrees of freedom. If there are partitions of interest parameters ψ and nuisance parameters λ , say $\beta = (\psi, \lambda)$, then under $H_0: \psi = \psi_0$, the test statistic $Q(\psi_0, \tilde{\lambda}) - Q(\hat{\psi}, \hat{\lambda})$ follows an asymptotically chi-square distribution with p degrees of freedom, where p is the dimension of parameter ψ . To test whether the first moment condition or the mean model assumption $H_0: E\{g(\beta_0)\} = 0$ holds or not, Hansen's [15] over-identifying moment condition test can be applied. If the first moment condition holds, then $Q(\hat{\beta})$ follows a chi-square distribution asymptotically with $r - q$ degrees of freedom, where r is the dimension of the extended score vector \bar{g}_N .

According to Qu *et al.* [7], when the sample size is moderate, all these test statistics follow a chi-square distribution asymptotically whether the working correlation is valid or not. This result contrasts to Rotnitzky and Jewell's [16] generalized score test, which is sensitive to the correct specification of the working correlation in the sense of type I error control. In addition, the power

Table I. Type I errors and Empirical relative power (ERP) between the GEE's Wald test and the QIF's score-type test (Power(QIF)/Power(GEE)), for the significance of treatment based on 1000 rounds of simulation and true correlation structure AR-1.

Model	Exchangeable						AR-1					
	$\alpha=0.3$			$\alpha=0.7$			$\alpha=0.3$			$\alpha=0.7$		
	Type I error		Power	Type I error		Power	Type I error		Power	Type I error		Power
	QIF	GEE	ERP	QIF	GEE	ERP	QIF	GEE	ERP	QIF	GEE	ERP
Normal	0.055	0.056	1.03	0.046	0.043	1.00	0.052	0.052	1.01	0.052	0.056	1.02
Binomial	0.046	0.046	1.09	0.052	0.046	1.05	0.052	0.053	1.07	0.054	0.053	1.08

of the test under the local alternative is optimal within the linear class of estimating functions. However, when the unstructured correlation is used, both GEE and QIF methods should perform similarly in the aspect of controlling for the type I error and achieving optimal test power.

A simulation study was conducted to compare the relative power of the GEE's Wald test and the QIF's score-type test. We considered a hypothetical situation for data generation: each of N subjects had 4 (n) repeated measurements of a trait, either continuous or dichotomous, and at each time two covariates, exposure (E) and treatment type (T), were recorded. Here, exposure E was a time-dependent continuous variable-generated independently from a uniform distribution on interval $(0, 1)$, while treatment status T was a binary variable generated from a Bernoulli distribution with probability 0.5 for each subject. Once the covariates were generated and the regression coefficient values were fixed, the response variable could be derived according to its distribution assumption, either normal or binomial. For normal responses, let $y_{it} = \beta_0 + \beta_1 E_{it} + \beta_2 T_i + \varepsilon_{it}$, $i = 1, \dots, 200$, $t = 1, \dots, 4$, with $(\varepsilon_{i1}, \dots, \varepsilon_{i4})' \sim \text{MVN}(0, \sigma^2 R(\alpha))$, where the true variance $\sigma^2 = 1$ and the correlation matrix $R(\alpha)$ will vary in the simulation. As for dichotomous responses, an algorithm suggested by Preisser *et al.* [17] was used with the same mean model subject to a logit link and sample size $N = 500$.

In the simulation of the normal data model, the hypothesis of interest to be tested was the association between the trait and the treatment, namely $H_0: \beta_2 = 0$ versus $H_1: \beta_2 \neq 0$. The true values were set as $\beta_0 = \beta_1 = 1.0$ and $\beta_2 = 0.2$ (under the normal data model). In the simulation of the binary data model, we considered the null $H_0: \beta_2 = 0.5$ (under a standard treatment) versus the alternative $H_1: \beta_2 = 0.6$ (under a new treatment), and the other true values were given by $\beta_0 = \beta_1 = 0.1$. Among several scenarios we considered, we here report one case with the true correlation structure being AR-1 and the misspecified correlation structure being exchangeable. Table I lists the summary results of both type I errors and empirical relative power (ERP) between the GEE's Wald test and QIF's score-type test, defined as Power(QIF)/Power(GEE), based on 1000 rounds of simulation. When ERP is bigger (or smaller) than 1, the QIF would correspondingly be more (or less) powerful than the GEE. It is easy to see from Table I that, in general, the QIF is more powerful than the GEE, even if the working correlation structure is correctly specified. This is in agreement with the known fact that the likelihood ratio score test usually has a better small-sample performance than the Wald test. This property gives rise to an important practical impact on clinical trial design regarding sample size determination. That is, to achieve the same power in a hypothesis test for treatment effects in a longitudinal study, the QIF requires a smaller sample size than the GEE.

2.3. Robustness

An important advantage of the QIF approach is that the QIF estimator is robust to outlying observations. This is because the influence function of the QIF is bounded while the influence function of the GEE is not bounded [4]. Specifically, the QIF possesses a redescending property; that is, the contribution of a single anomalous observation to the estimating function goes to zero as the residual of the outlying observation goes to infinity. This property results from the use of an empirical variance estimator, or the weighting matrix C_N , of the extended scores. Note that the inverse of C_N assigns smaller weights for observations with larger variances. Thus, the QIF automatically downweights outlying observations. This downweighting, however, is not in effect for the independence working structure, since in that case the QIF is equivalent to the GEE estimator.

Through a simulation study, Qu and Song [4] showed substantial differences in robustness between the QIF and the GEE methods. Their results indicate that the GEE is highly sensitive to abnormal observations, the consistency of the GEE estimator no longer holds, and large biases result from the data contamination. In contrast, the QIF estimator is affected only marginally, and when the proportion of contaminated clustered is not too high, say 30 per cent or lower, the QIF estimator works well.

2.4. Model selection

When the marginal model is used to fit the data, two primary tasks concerning model selection include the selection of covariates and the selection of a correlation structure. Traditional variable selection criteria, such as AIC and BIC, widely used to select covariates, are difficult to derive in the marginal modeling approach because the likelihood function is not fully specified. Pan [6] proposed a generalization of the AIC criterion, called the quasi-likelihood information criterion (QIC) under the independence model. The QIC uses the quasi-likelihood of the independence model as an objective function and hence ignores the serial correlation in the data; see also [9]. In contrast, it seems rather natural to extend the AIC and BIC to the QIF approach that takes account the correlation of the data into the model selection measure, where twice the negative log-likelihood function is replaced by the QIF objective function due to their similar asymptotic properties. For a given subset of indices, \mathcal{M} , let \mathcal{M}_c be its complement. Let $\hat{\beta}_{\mathcal{M}}$ be the minimizer of $Q_N(\beta_{\mathcal{M}})$. The QIF for the full model, $\mathcal{M} \cup \mathcal{M}_c$, can be simply viewed as a function of $\beta_{\mathcal{M}}$ by setting $\beta_{\mathcal{M}_c} = 0$. A penalized QIF is defined as follows:

$$Q_N(\hat{\beta}_{\mathcal{M}}) + \lambda \times |\mathcal{M}|$$

where $|\mathcal{M}|$ is the cardinality of \mathcal{M} . The AIC and BIC correspond to $\lambda = 2$ and $\ln(N)$, respectively. Wang and Qu [18] showed that the penalized QIF with the BIC penalty has a well-known model selection consistency property. That is, with the probability approaching one, the QIF with the BIC penalty selects the true model as the sample size N goes to infinity. Simulation studies given by Wang and Qu [18] indicated that the AIC and BIC worked well for the selection of covariates, and better than Pan's QIC-type criterion.

However, in a separate simulation study we found that neither the AIC nor the BIC performed well for the selection of the correlation structure, and that there is a tendency that both criteria over-select the exchangeable correlation over the AR-1 correlation. To overcome this problem, we

propose a new selection criterion based on the trace of the Godambe Information (TGI) (or the sandwich covariance) matrices (previously denoted by $J = D'\Sigma^{-1}D$), defined as follows:

$$\text{TGI} = \text{trace}(J_2 - J_1)$$

where their inverse matrices, J_1^{-1} and J_2^{-1} , are the asymptotic covariance matrices obtained from the two analyses under the respective correlation structures. A heuristic justification of this criterion is given as follows. In a quasi-likelihood setting (e.g. [9, p. 140]), the AIC criterion is $-2\text{QL}(\beta) + 2\text{trace}(D^{-1}\Sigma)$, where the first term depends on the regression coefficients, but is independent of the correlation structure as it is formed under the independence model. In other words, the first QL term will be irrelevant to, and only the penalty term in this AIC will be involved in, the selection of correlation structure. It is unfortunate that in the QIF matrix D is not invertible, because it is generally not a square matrix; hence, we propose to replace the penalty by $\text{trace}\{(D'\Sigma^{-1}D)^{-1}\}$, namely $\text{trace}(J^{-1})$. This replacement is intuitively sensible as the diagonal elements of matrix J^{-1} determine the standard errors of the estimates. The smaller $\text{trace}(J^{-1})$ reflects a better model fit. For computational ease, an equivalent measure would be $\text{trace}(J)$, with a larger value favoring a better model fit. Thus, when the TGI is positive, we prefer the correlation structure J_2 over J_1 , given that in both cases the mean model is fixed and the numbers of the correlation parameters are the same.

To examine the performance of the proposed TGI, we conducted a simulation study based on the following setting: each of 50 (N) subjects had 4 (n) repeated measurements, and at each time two covariates, exposure (E) and treatment type (T), were recorded. Here, exposure E is a time-dependent continuous variable generated independently from a uniform distribution on interval (0, 1), while treatment status T is a binary variable generated from a Bernoulli distribution with probability 0.5 for each subject. Once the covariates are generated and the regression coefficient values are fixed, the response variable can be derived according to its distribution assumption, either normal or binomial. For normal responses, let $y_{it} = \beta_0 + \beta_1 E_{it} + \beta_2 T_i + \varepsilon_{it}$, $i = 1, \dots, 50$, $t = 1, \dots, 4$, with $(\varepsilon_{i1}, \dots, \varepsilon_{i4})' \sim \text{MVN}(0, \sigma^2 R(\alpha))$, where the true $\sigma^2 = 1$. As for binary response, an algorithm suggested by Preisser *et al.* [17] was used to generate correlated binary responses with a given correlation structure.

Table II. The per cent of the true correlation structure being selected among 1000 simulations by the TGI and BIC criteria based on the QIF method. Longitudinal data of 50 subjects and 4 repeated measurements for each subject are generated from a normal model or a binomial model with two covariates, respectively.

Model	True Corr	α	Selection per cent	
			TGI	BIC
Normal	Exch	0.3	58.9	82.8
		0.7	84.9	85.1
	AR-1	0.3	84.9	10.0
		0.7	82.5	11.2
Binomial	Exch	0.3	82.7	37.3
		0.7	92.4	39.5
	AR-1	0.3	79.7	33.9
		0.7	91.3	32.2

Among 1000 simulations, each was fitted by two correlation structures, exchangeable and AR-1. The per cent of the true structure being selected by the proposed TGI and the BIC criteria is listed in Table II. It is evident that the TGI performed much better than the BIC in selecting the true correlation structure. The performance of the AIC was almost identical to that of the BIC, hence, it is also not recommended for the selection of correlation structure. Once again, we would like to point out that the QIF method produces consistent estimators of the regression coefficients even if the correlation structure is misspecified. Thus, the selection of correlation structure is perhaps secondary in comparison with the selection of covariates.

3. APPLICATIONS

This section gives an introduction to a newly developed SAS MACRO QIF, and presents a data example to illustrate the use of this macro in practice.

3.1. SAS macro QIF

The SAS MACRO QIF was developed by Song and Jiang at the University of Waterloo in 2006 to provide a tool that fits marginal models. It uses PROC GENMOD to generate initial values for Newton–Raphson iterations, PROC IML to carry out the main calculations concerning the search of the minimizer, and an Output Delivery System to deliver results. It was developed under SAS/STAT Version 9.1.3. The initial coding was made in MATLAB by Qu and her research group at the Department of Statistics, Oregon State University. Part of the SAS macro for fitting the normal marginal model was translated from the original MATLAB coding. The other part of the macro was independently developed by Song and Jiang.

The current version of the macro implements several types of models with the error distributions and canonical link functions listed in Table III. Four types of working correlation structures are implemented in the macro, including Independence, Exchangeable (or Compound Symmetry), AR-1 and Unstructured. In the AR-1 case, only two major basis matrices (i.e. I and M_1^*) were used. For the case of the independence correlation structure, this macro produces exactly the same results as those given by PROC GENMOD.

The syntax for the macro was designed to be similar to that of PROC GENMOD, such that it is easy to use. There are some macro-specific options, and interested readers may refer to the user's manual written by Song and Jiang available at URL address: <http://www-personal.umich.edu/~pxsong/QIFmanual.pdf>.

The output from this macro is optional. The default is `print=Y`, but the print can be muted by choosing `print=N`. All the output tables can be extracted by calling their names given in the macro.

Table III. Distributions and link functions are implemented in SAS MACRO QIF.

Model	Error distribution	Link
Linear regression	Normal	identity
Logistic regression	Binomial	logit
Log-Linear regression	Poisson	log
Gamma regression	Gamma	reciprocal

The source code of the macro can be downloaded at the following URL address: http://www.stats.uwaterloo.ca/stats_navigation/downloadRequest.shtml.

3.2. Data example

Now we illustrate an application of the macro via one data analysis example. We use epileptic seizures data reported by Thall and Vail [19] from a clinical trial of 59 epileptics. For each patient, the number of epileptic seizures was recorded during a baseline period of 8 weeks. Patients were then randomized to two treatment arms, one with the anti-epileptic drug progabide, and the other with a placebo, in addition to standard chemotherapy. The number of seizures was recorded in four consecutive two-week periods after the randomization. The scientific question was whether the drug progabide helps to reduce the rate of epileptic seizures. Figure 1 shows a longitudinal plot of the data across the two treatment arms, where the square-root of the observed seizure counts are plotted against the times of visits. This transformation on the responses was invoked for better visualization of patients' trajectories. It is easy to see that in the progabide group (the top panel), one patient (ID 207) clearly exhibits an outlying trajectory from the other patients. Several analyses (e.g. [20]) have noted the existence of this abnormal patient ID 207 who recorded an excessively large number of seizures over the study period in comparison with the rest of the 58 patients. For example, we invoked SAS/IML software developed in [21] for GEE regression diagnostics to obtain a plot of Cook's distance versus leverages at the cluster level, as shown in Figure 2, under the bias-corrected variance estimator. This figure confirms clearly that patient 207 is an influential outlier. It is of interest to observe how the QIF and the GEE perform in the presence of an outlying patient in the data.

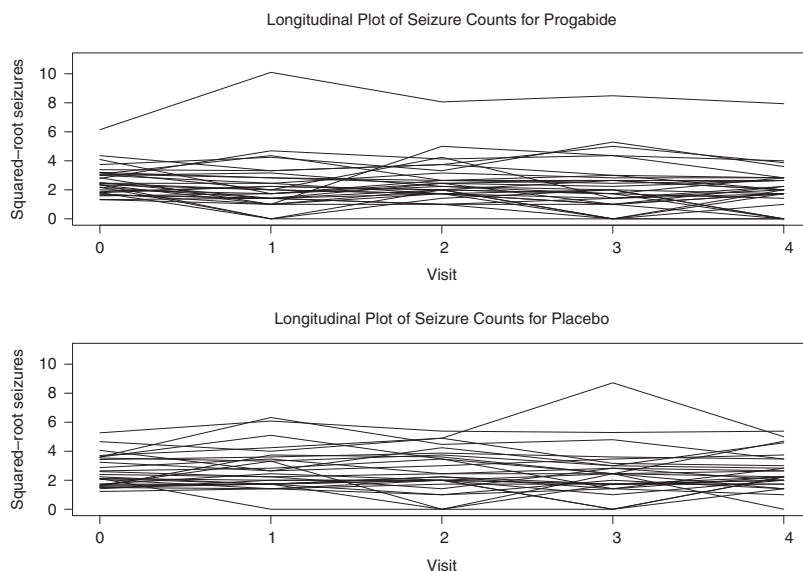


Figure 1. Longitudinal plots of observed seizures of epileptic patients across two treatment arms.

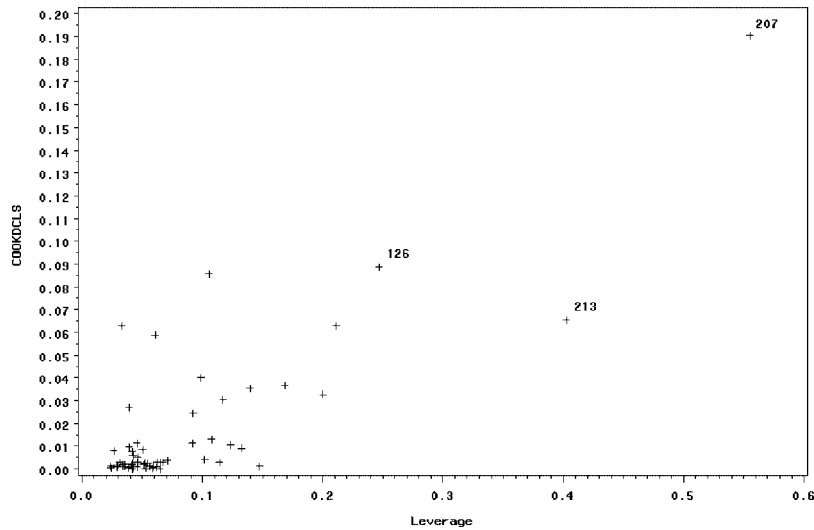


Figure 2. Diagnostic plot of Cook's distance (y axis labeled by 'COOKDCLS') versus leverages (x axis labeled by 'Leverage') at cluster level under the bias-corrected variance estimator.

We compare the robustness of the two methods by DFBETAS, namely the difference of estimated regression coefficients obtained with and without the outlier ID207. Precisely, for each parameter estimation, we look at a ratio of the relative change given as follows [22]:

$$RC(\theta_j) = \frac{|\theta_{j,gee}^{with} - \theta_{j,gee}^{without}|}{s.e.(\theta_{j,gee}^{without})} \bigg/ \frac{|\theta_{j,qif}^{with} - \theta_{j,qif}^{without}|}{s.e.(\theta_{j,qif}^{without})}$$

If $RC > 1$, then the outlier would affect the GEE more severely than the QIF. In other words, the larger the RC, the more robust is the QIF compared with the GEE.

Following Diggle *et al.* [20], suppose the number of seizures in a two-week period Y_{ij} follows a marginal Poisson(μ_{ij}) distribution, where $\mu_{ij} = E(Y_{ij} | \mathbf{x}_{ij}, t_{ij})$ is the expected number of seizures during the j th two-week period for subject i . The first moment is specified by a log-linear model

$$\log(\mu_{ij}) = \beta_0 + \beta_1 \text{bsln}_i + \beta_2 \text{trt}_i + \beta_3 \text{logage}_i + \beta_4 \text{vst}_j \tag{4}$$

where *bsln* is log(baseline seizure counts/4), *trt* is a binary covariate equal to 1 for progabide and 0 for placebo, *logage* is log(age) and *vst* is time of visit equal to 1, 2, 3, 4. Using SAS PROC GENMOD and the SAS MACRO QIF, we fit the data to the above marginal log-linear model and obtained the estimates and standard errors given in Table IV, where AR-1 was chosen as the working correlation. The SAS MACRO code that we executed is given as follows:

```
%qif(data=epilepsy, yvar=y, xvar=baseline trt logage visit,
      id=id, dist=poisson, corr=ar-1,
      print=y, outpar=par1, outqif=qif1, outcov=cov1, outres=resid);
run;
```

The p -value of the goodness-of-fit Q statistic, according to χ^2_5 , is 0.5934 for the complete data, and 0.3161 for the data without ID207. The 5 degrees freedom is obtained by $r - q$ in Section 2.2,

Table IV. A summary of SAS PROC GENMOD and SAS MACRO QIF outputs in the analysis of the epileptic seizures data with the AR-1 working correlation.

Par	Complete data				Without patient #207			
	Estimate		Std err		Estimate		Std err	
	GEE	QIF	GEE	QIF	GEE	QIF	GEE	QIF
intcpt	-2.522	-2.233	1.034	1.006	-2.380	-2.017	0.863	0.892
bsln	1.247	1.193	0.163	0.099	0.987	0.960	0.080	0.066
trt	-0.020	-0.046	0.190	0.141	-0.255	-0.281	0.152	0.146
logage	0.653	0.581	0.287	0.270	0.783	0.680	0.247	0.261
vst	-0.064	-0.052	0.034	0.026	-0.045	-0.047	0.035	0.031
<i>Q</i> -stat	—	3.7	—	—	—	5.9	—	—
TGI	—	40.26	—	—	—	21.76	—	—

where $r = 10$ corresponds to the number of components in the extended score vector $\bar{g}_N(\beta)$ and $q = 5$ is the number of regression coefficients in this log-linear model. In both cases, the QIF goodness-of-fit test indicate that the log-linear model specification (4) is appropriate, in the sense that the mean zero assumption is satisfied. Obviously, the treatment is not statistically significant, where the p -value changed from 0.093 obtained by the GEE to 0.054 obtained by the QIF in the case of excluding patient 207. However, both baseline seizure and age are statistically significant.

The final step of the analysis was to determine which correlation structure, AR-1 or exchangeable, would be chosen to make the final conclusion of the analysis. We did the calculation of the $TGI(J_2 - J_1)$, where J_2 is the AR-1 and J_1 is the exchangeable. The TGI values in Table IV are positive regardless of patient ID 207 being included or excluded in the analysis. This means that the data favor the AR-1 correlation.

As part of the sensitivity analysis, we also fit the data with both independence and exchangeable correlation structures. As expected, the point estimates and standard errors are similar over the three correlation structures, and we found no conflicting results regarding statistical significance in these three correlation scenarios. The BIC values are 24.1, 22.3 and 20.3 for AR-1, exchangeable and independence correlation structures when patient 207 is included in the analysis. When patient 207 is excluded from the analysis, the BIC values become 26.2, 23.1 and 20.3, respectively. This suggests that the data of patient 207 did not seem to affect the order of the BIC values (or the Q statistics) among the three correlation structures. Once again, it is worth noting that according to the simulation study in Section 2.4, the TGI appears to be a better method for the selection of correlation structure than the BIC. However, under given a correlation structure Wang and Qu [18] show that the BIC is appropriate to select covariates.

For comparison of robustness against the outlier ID 207, the RC between the QIF and the GEE is given in Table V for each regression coefficient. We also applied Hammill and Preisser's SAS/IML GEE diagnostics package [21] to examine the DFBETA measures for the influence of patient 207 on individual estimates of covariate effects (the figures are not shown), and found that this influential outlier in fact substantially affects two covariates: *trt* and *vst*. In Table V, the most striking is the estimation for the effect of covariate *vst*. The QIF appears to be more robust than the GEE on this estimation. In addition, since the effect of treatment is found to be insignificant in all the analyses, assessing the influence of patient 207 on the estimation of treatment

Table V. Relative changes (RC) of point estimates with respect to ID 207 between GEE and QIF under AR-1 correlation.

	Covariate				
	intercept	baseline	treatment	log-age	visit
RC	0.68	0.92	0.96	1.39	3.37

effect does not seem to be important in this data analysis. To close this analysis, we would like to comment that a better assessment for the robustness performance of the two methods could be made with a larger sample size. In this analysis, we identified the influence of patient 207 on the two covariates via Hammill and Preisser's software, and then focused on our evaluation on these covariates exclusively. There are other aspects of robustness worth further exploration.

4. CONCLUDING REMARKS

In this paper, we have shown that the QIF is an important and powerful alternative to the popular GEE. In particular, the QIF is robust to outliers or contaminated data and provides both a goodness-of-fit test to validate the first moment marginal mean assumption and a model selection criterion to perform a stepwise regression analysis. These properties are either unavailable or difficult to establish in the GEE method. The availability of the SAS MACRO QIF facilitates the application of the QIF in the analysis of longitudinal data.

However, the QIF has some limitations. First, the QIF depends on the availability of the basis matrices for a given correlation structure. Currently, the QIF is established only for four types of working correlation structures: Independence, Exchangeability, AR-1 and Unstructured. Although these four structures cover most important cases, it would be of interest to develop the QIF that accommodates flexible correlation structures. Second, the current version of a QIF cannot handle the unequally spaced repeated measurements. It is worth mentioning that there is one competitive alternative to the GEE for the ease of incorporating new correlation structures, developed by Shults and her colleagues in a series of papers including [23, 24]. Third, similar to the GEE, when missing data are present, the QIF works only under MCAR, and a certain modification along the lines of Robins' *et al.* [25, 26] inverse probability weighting is worth exploring. Finally, although the proposed TGI criterion works reasonably well, it tended to over-select the unstructured correlation if no penalty on the number of parameters in the correlation structure was imposed. The issue of how to add a proper penalty on the TGI needs further investigation. In addition, the idea of the TGI is, in general, applicable to the GEE method as well. However, this criterion can be difficult to apply when both potential inconsistency of the asymptotic covariance matrix when the working correlation is misspecified, and bias caused by outliers.

The authors have developed and tested the SAS macro QIF through simulation experiments and data analyses. The availability of a reliable software package for the QIF is helpful for general practitioners'. One future upgrade that we would like to work out soon is an automatic stepwise procedure in the macro. We certainly welcome any comments and suggested improvements from readers on this software.

ACKNOWLEDGEMENTS

The authors are very grateful to the anonymous reviewers for their constructive suggestions that led to significant improvement in the exposition of the paper. This research was supported by NSF grants to both the first author (Song) and the last author (Qu).

REFERENCES

1. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
2. Crowder M. On consistency and inconsistency of estimating equations. *Econometric Theory* 1986; **3**:305–330.
3. Crowder M. On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika* 1995; **82**:407–410.
4. Qu A, Song PX-K. Assessing robustness of generalised estimating equations and quadratic inference functions. *Biometrika* 2004; **91**:447–459.
5. Preisser JS, Qaqish BF. Deletion diagnostics for generalized estimating equations. *Biometrika* 1996; **83**:551–562.
6. Pan W. Model selection in estimating equations. *Biometrics* 2001; **57**:529–534.
7. Qu A, Lindsay BG, Li B. Improving generalized estimating equations using quadratic inference functions. *Biometrika* 2000; **87**:823–836.
8. Fitzmaurice GM, Laird NM, Rotnitzky A. Regression models for discrete longitudinal responses. *Statistical Science* 1993; **8**:284–309.
9. Hardin JW, Hilbe JM. *Generalized Estimating Equations*. Chapman & Hall: Boca Raton, 2003.
10. Horton NJ, Bebbchuk JD, Jones CL, Lipsitz SR, Catalano PJ, Zahner GEP, Fitzmaurice GM. Goodness-of-fit for GEE: an example with mental health service utilization. *Statistics in Medicine* 1999; **18**:213–222.
11. Small CG, Wang J, Yang Z. Eliminating multiple root problems in estimation. *Statistical Science* 2000; **15**:313–341.
12. Hall DB. On the application of extended quasi-likelihood to the clustered data case. *Canadian Journal of Statistics* 2001; **29**:77–97.
13. Qu A, Li R. Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics* 2006; **62**:379–391.
14. Qu A, Lindsay BG. Building adaptive estimating equations when inverse of covariance estimation is difficult. *Journal of the Royal Statistical Society, Series B* 2003; **65**:127–142.
15. Hansen L. Large sample properties of generalized methods of moments estimators. *Econometrica* 1982; **50**:1029–1055.
16. Rotnitzky A, Jewell N. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* 1990; **90**:485–497.
17. Preisser JS, Lohman KK, Rathouz PJ. Performance of weighted estimating equations for longitudinal binary data with drop outs missing at random. *Statistics in Medicine* 2002; **21**:3035–3054.
18. Wang L, Qu A. Consistent model selection and data-driven smooth tests for longitudinal data in the estimating equation approach. *Journal of Royal Statistical Society, Series B* 2009; **71**:177–190.
19. Thall PF, Vail SC. Some covariance models for longitudinal count data with overdispersion. *Biometrics* 1990; **46**:657–671.
20. Diggle PJ, Heagerty P, Liang K-Y, Zeger SL. *Analysis of Longitudinal Data* (2nd edn). Oxford University Press: New York, 2002.
21. Hammill BG, Preisser JS. A SAS/IML software program for GEE and regression diagnostics. *Computational Statistics and Data Analysis* 2006; **51**:1197–1212.
22. Song PX-K. *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer: New York, 2007.
23. Shults J, Chaganty NR. Analysis of serially correlated data using quasi-least squares. *Biometrics* 1998; **54**:1622–1630.
24. Shults J, Whitt M, Kumanyika S. Analysis of data with multiple sources of correlation in the framework of generalized estimating equations. *Statistics in Medicine* 2004; **23**:3209–3226.
25. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 1994; **89**:846–866.
26. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes under the presence of missing data. *Journal of the American Statistical Association* 1995; **90**:106–121.