Biometrika (2004), **91**, 2, *pp.* 447–459

Printed in Great Britain

Assessing robustness of generalised estimating equations and quadratic inference functions

By ANNIE QU

Department of Statistics, Oregon State University, Corvallis, Oregon 97331, U.S.A.

qu@stat.orst.edu

and PETER X.-K. SONG

Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada M3J 1P3

song@mathstat.yorku.ca

Summary

In the presence of data contamination or outliers, some empirical studies have indicated that the two methods of generalised estimating equations and quadratic inference functions appear to have rather different robustness behaviour. This paper presents a theoretical investigation from the perspective of the influence function to identify the causes for the difference. We show that quadratic inference functions lead to bounded influence functions and the corresponding M-estimator has a redescending property, but the generalised estimating equation approach does not. We also illustrate that, unlike generalised estimating equations, quadratic inference functions can still provide consistent estimators even if part of the data is contaminated. We conclude that the quadratic inference function is a preferable method to the generalised estimating equation as far as robustness is concerned. This conclusion is supported by simulations and real-data examples.

Some key words: Data contamination; Generalised method of moments; Influence function; Longitudinal data; M-estimator; Outlier; Redescending property.

1. Introduction

In longitudinal or clustered data, observations are repeatedly measured from the same subject or cluster, and therefore the within-cluster correlation has to be accounted for in order to make a proper statistical inference. The method of generalised estimating equations (Liang & Zeger, 1986) has gained popularity in estimation of parameters. The generalised estimating equation approach requires correct specification of the first two moments of a model. However, these moment assumptions can be distorted by contaminated or irregular measurements. As a result, the generalised estimating equation method fails to give consistent estimators, and more seriously this will lead to incorrect conclusions (Preisser & Qaqish, 1996, 1999; Mills et al., 2002).

Downweighting and deleting putatively contaminated clusters are two common ways suggested by Preisser & Qaqish (1996, 1999) of ensuring consistent estimation in the generalised estimating equation method. However, the implementation of these strategies relies on whether, or how, potentially problematic clusters are identified beforehand. Furthermore, it is also difficult to judge whether a large residual is attributed to an irregular measurement on the response or on the covariates. Therefore, it is generally uncertain which downweighting strategy to take between the 'Mallows' class on the covariates and the 'Schweppe' class (Pregibon, 1982; Künsch et al., 1989) on the response. The nonparametric smoothing spline M-estimator is proposed by He et al. (2002) for obtaining a consistent estimator in the presence of outliers, but the correlated nature of measurements is not incorporated in their approach.

As an alternative to the generalised estimating equation, Qu et al. (2000) proposed the quadratic inference function to improve the efficiency of the regression parameter estimator when the working correlation of the generalised estimating equation is misspecified. Notably, the method appears to create a downweighting strategy automatically in the estimation procedure, so that it behaves robustly against irregular measurements arising from either response or covariate variables. In fact, its greater robustness compared to the generalised estimating equation method was observed in several empirical studies, and this motivated us to explore theoretical explanations in the present paper.

A robust estimator characteristically has a bounded influence function (Hampel et al., 1986). In this paper, we show that the influence function of the quadratic inference function estimator is bounded, whereas the influence function of the generalised estimating equation estimator is unbounded. An intuition behind this difference is that the generalised estimating equation method imposes a parametric working covariance matrix that is essentially independent of any residual variations. In contrast, the estimating equations derived from the quadratic inference function involve the variability of residuals, which gives rise to an automatic downweighting for any observations associated with large residual values.

In particular, the minimiser of the quadratic inference function has a redescending property (Holland & Welsch, 1977; Hampel et al., 1986, p. 150); that is, the associated estimating function $g(z)$ tends to zero as the Euclidean norm of z goes to infinity. This implies that the truncation based on certain tuning constants (Huber, 1981; Ronchetti & Trojani, 2001; Mills et al., 2002) in M-estimation emerges automatically in quadratic inference function estimation.

2. GENERALISED ESTIMATING EQUATIONS AND QUADRATIC INFERENCE FUNCTIONS

Let y_{it} be a response variable and let x_{it} be a $q \times 1$ vector of covariates, measured at time $t = 1, \ldots, n_i$, for subjects $i = 1, \ldots, N$. We assume that the model satisfies the first moment model assumption that

$$\mu_{it} = E(y_{it}) = \mu(x_{it}'\beta), \tag{1}$$

where $\mu(.)$ is a known link function, and β is the regression parameter.

To estimate β, Liang & Zeger (1986) proposed generalised estimating equations

$$s(\beta) = \sum (\dot{\mu}_i)' W_i^{-1}(y_i - \mu_i) = 0, \tag{2}$$

where $y_i = (y_{i1}, \ldots, y_{in_i})'$, $\mu_i = (\mu_{i1}, \ldots, \mu_{in_i})'$, $\dot{\mu}_i = \partial\mu_i/\partial\beta$ and $W_i = A_i^{\frac{1}{2}} R A_i^{\frac{1}{2}}$, with A_i being the diagonal matrix of marginal variances for the ith cluster and $R = R(\alpha)$ being the working correlation matrix. If the model assumption (1) is satisfied, then the generalised estimating equation estimator is consistent regardless of whether the working correlation is correctly specified or not. Furthermore, the generalised estimating equation estimator is efficient when the working correlation is correctly specified.

The quadratic inference function method proposed by Qu et al. (2000) does not require more assumptions than does the generalised estimating equation method, but yields a substantial improvement in efficiency for the estimator of β when the working correlation is misspecified, and equal efficiency to the generalised estimating equation when the working correlation is correct. The fomulation is based on the assumption that the inverse of the working correlation R can be approximated by a linear combination of basis matrices; that is,

$$R^{-1} = \sum_{i=1}^{m} \alpha_i M_i, \tag{3}$$

where $M_1, \ldots, M_m$ are known matrices and $\alpha_1, \ldots, \alpha_m$ are unknown constants. Therefore, the generalised estimating equation (2) becomes

$$\sum_{i=1}^{N} (\dot{\mu}_i)' A_i^{-\frac{1}{2}} (\alpha_1 M_1 + \ldots + \alpha_m M_m) A_i^{-\frac{1}{2}} (y_i - \mu_i) = 0,$$

which is a linear combination of elements of the extended score vector

$$g(\beta) = \sum_{i=1}^{N} g_i(\beta) = \begin{pmatrix} \sum_{i=1}^{N} (\dot{\mu}_i)' A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} (y_i - \mu_i) \\ \vdots \\ \sum_{i=1}^{N} (\dot{\mu}_i)' A_i^{-\frac{1}{2}} M_m A_i^{-\frac{1}{2}} (y_i - \mu_i) \end{pmatrix}. \tag{4}$$

Since the dimension of the extended score in (4) is greater than the number of unknown parameters, one cannot directly solve $g(\beta) = 0$ for β. Instead, we follow Hansen's (1982) generalised method of moments and estimate β by minimising the quadratic distance function

$$Q(\beta) = g' C^{-1} g, \tag{5}$$

where $C = \sum_{i=1}^{N} g_i(\beta) g_i'(\beta)$ is the sample variance matrix of g. The empirical estimator C plays an important role in robust estimation for which we will provide more details in the next section. The objective function Q is referred to as the quadratic inference function since it also provides an inference function for testing. Note that, if an independent working correlation, or exchangeable correlation for balanced data, is assumed, the quadratic inference function estimator is identical to the generalised estimating equation estimator, because they have the same estimating functions.

It is easy to see that the Q function in (5) is bounded between 0 and N. The lower bound of 0 is obvious since C is nonnegative definite. To establish the upper bound of N, let $H = (g_1, \ldots, g_N)'$ and let 1_N be an N-element vector with 1 for all components. Denote the projection matrix by $P_H = H(H'H)^{-1}H'$. It follows immediately that, by orthogonal projection and idempotent properties of the projection matrix,

$$Q = 1_N' H (H'H)^{-1} H' 1_N = 1_N' P_H 1_N = \|P_H 1_N\|^2 \leqslant \|1_N\|^2 = N.$$

Here $\|.\|$ denotes the Euclidean norm of a matrix $A = (a_{ij})$, defined by $\|A\| = (\sum_i \sum_j a_{ij}^2)^{\frac{1}{2}}$. Note that $\|A\| = \{\mathrm{tr}(A'A)\}^{\frac{1}{2}}$. The bounded quadratic inference function is important for deriving the redescending M-estimator from the quadratic inference function. Figure 1 illustrates the boundedness property. The next section provides details of how Fig. 1 was drawn.

Qu et al. (2000) showed that the quadratic inference function estimator of β is consistent and asymptotically normal, and that the asymptotic variance matrix $(D'\Sigma^{-1}D)^{-1}$ attains the minimum in the sense of Löwner ordering, where $D = E(\partial g/\partial\beta)$ and $\Sigma = \text{var}\{g(\beta)\}$. Furthermore, the quadratic inference function has desirable inferential properties as a likelihood ratio test. In particular, one may construct a goodness-of-fit test statistic $Q(\hat{\beta})$ for testing the model assumption (1). Hansen (1982) showed that the asymptotic distribution of $Q(\hat{\beta})$ is χ^2 with $\{\dim(g) - \dim(\beta)\}$ degrees of freedom under the model assumption (1). Note that these properties hold whether or not the working correlation is correctly specified.

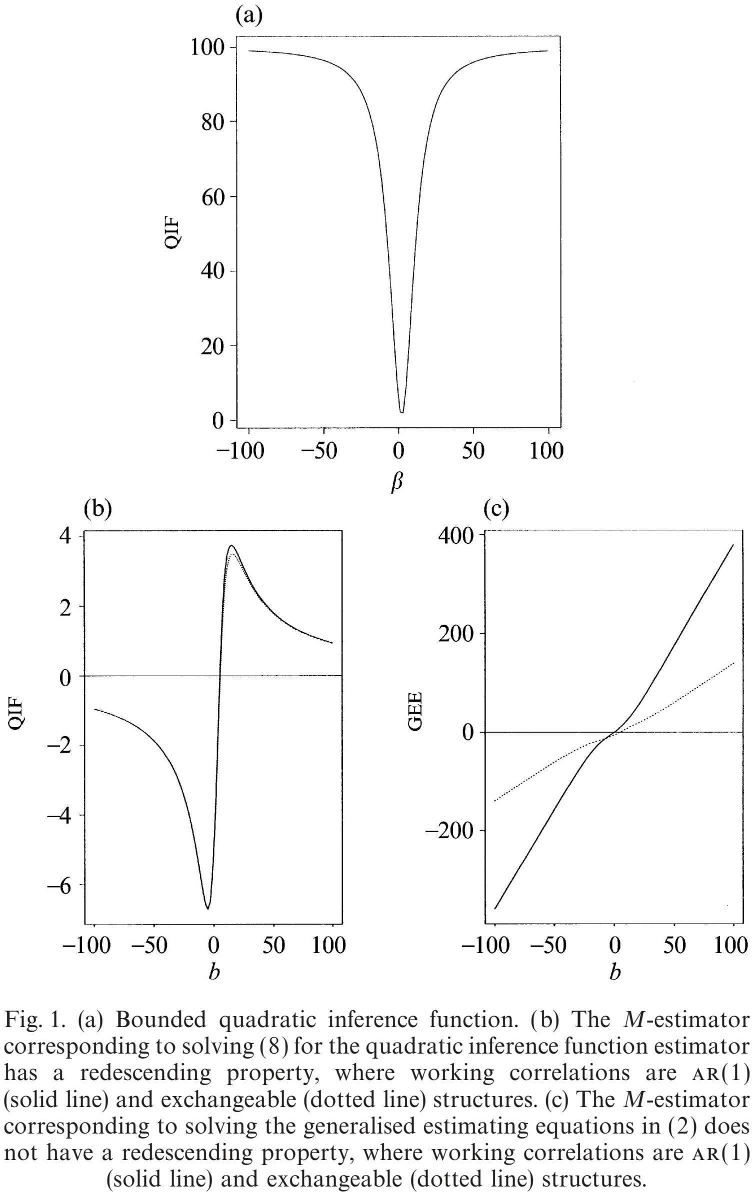

Fig. 1. (a) Bounded quadratic inference function. (b) The M-estimator corresponding to solving (8) for the quadratic inference function estimator has a redescending property, where working correlations are AR(1) (solid line) and exchangeable (dotted line) structures. (c) The M-estimator corresponding to solving the generalised estimating equations in (2) does not have a redescending property, where working correlations are AR(1) (solid line) and exchangeable (dotted line) structures.

3. Robustness property

Hampel (1974) defined the influence function of the β-estimator T as

$$\text{IF}(z, T, P_\beta) = \lim_{\varepsilon \to 0} \frac{T\{(1-\varepsilon)P_\beta + \varepsilon\Delta_z\} - T(P_\beta)}{\varepsilon}, \tag{6}$$

where P_β is the probability measure of the assumed model, and Δ_z is the probability measure with mass 1 at the single contaminated data point z. Heuristically, the influence function measures the asymptotic bias caused by fractional data contamination. It is known that an estimator T whose influence function is unbounded may have an unbounded asymptotic bias under single-point data contamination.

An M-estimator is defined as the solution to the estimating equation

$$\sum_{i=1}^{N} s_i(z_i, \beta) = 0$$

for specified functions s_i. Let z denote the suspect observation under investigation. Hampel et al. (1986) showed that the influence function of the M-estimator is

$$\text{IF}(z, T, P_\beta) = -\{E_\beta(\dot{s})\}^{-1} s(z, T), \tag{7}$$

where s may be s_i if the ith observation happens to be the z, and $\dot{s}$ is the first derivative of s with respect to β.

Since for a given z the influence function of an M-estimator is proportional to the estimating function $s(z, T)$, this implies that the influence function of an M-estimator is bounded if and only if $s(z, T)$ is bounded. Theorem 1 below claims that the quadratic inference function estimator has a bounded influence function and the corresponding M-estimator has a redescending property. In contrast, we show that the generalised estimating equation does not have such robustness properties.

Note that minimising the quadratic inference function in (5) is asymptotically equivalent to solving

$$\dot{g}'C^{-1}g = \sum_i D'C^{-1}g_i = 0, \tag{8}$$

since $\dot{g}$ is nonrandom. The quadratic inference function estimator derived by solving (8) is therefore an M-estimator.

Suppose cluster i, say, is being investigated. To avoid possible confusion in notation, we use g_i to indicate the cluster of interest, but let $z = y_i - \mu_i$ be the residual associated with cluster i under the model assumption (1), in which we omit the subscript for simplicity.

Theorem 1. *The quadratic inference function estimator has a redescending property; that is, the estimating function $D'C^{-1}g_i(z)$ is bounded and approaches zero as $\|z\| \to \infty$.*

Proof. For contaminated cluster i, denote the corresponding residual by z, and let $\|z\| \to \infty$. Since $C = \sum g_j g_j' = G + g_i(z)g_i(z)'$, where $G = \sum_{j \neq i} g_j g_j'$, then, by Rao (1973, p. 33),

$$C^{-1} = \{G + g_i(z)g_i'(z)\}^{-1} = G^{-1} - \frac{1}{1 + g_i'(z)G^{-1}g_i(z)} G^{-1}g_i(z)g_i'(z)G^{-1},$$

provided that both C^{-1} and G^{-1} exist. It follows that

$$\begin{aligned} C^{-1}g_i(z) &= G^{-1}g_i(z) - \frac{g_i'(z)G^{-1}g_i(z)}{1+g_i'(z)G^{-1}g_i(z)}G^{-1}g_i(z) \\ &= \frac{1}{1+g_i'(z)G^{-1}g_i(z)}G^{-1}g_i(z). \end{aligned}$$

Since $g_i(z)$ given in equation (4) is effectively linear in z, it is easy to see that

$$\begin{aligned} \|D'C^{-1}g_i(z)\|^2 &= \frac{g_i'(z)G^{-1}DD'G^{-1}g_i(z)}{1+g_i'(z)G^{-1}g_i(z)} \frac{1}{1+g_i'(z)G^{-1}g_i(z)} \\ &\to 0 \end{aligned}$$

as $\|z\| \to \infty$, where on the right-hand side the first term is bounded and the second term tends to zero when $\|z\| \to \infty$. Therefore, the quadratic inference function estimator has a redescending property with the horizontal axis as the asymptote. □

The implication of Theorem 1 is as follows. If there is an unduly large residual z caused by either an irregular measurement of response y, or an irregular predictor μ caused by misspecification of the model or irregular values of covariates, the estimating equation in (8) will automatically downweight the corresponding cluster through the inverse of the C matrix. This inverse matrix penalises any large residual values in the estimating equations. This property also indicates that the truncation usually applied in robustness estimation is inherent here.

On the other hand, the generalised estimating equation in (2) is not bounded when a residual $\|z\| \to \infty$. Note that the term $\dot{\mu}_i$ is a function of covariate x_i and marginal mean μ_i, and that the working variance matrix for the ith cluster is $W_i = A_i^{\frac{1}{2}} R A_i^{\frac{1}{2}}$, where $A_i = \phi \operatorname{diag}\{V(\mu_{i1}), \ldots, V(\mu_{in_i})\}$ is a parametric function of the marginal mean μ_i and is not associated with the response y_i, and the working correlation $R(\alpha)$ is the same for all clusters. Here $V(.)$ is a known variance function and ϕ is the dispersion parameter that is either known, in some cases such as Bernoulli and Poisson, or can be factorised out of the estimating equations if it is unknown. Clearly, neither A_i nor $R(\alpha)$ has a downweighting effect on irregular observations with large variation, since they are either the same for all clusters or do not depend on the residual z in any way. As a result, the generalised estimating equation diverges at a linear rate as $\|z\| \to \infty$. This is why, without downweighting influential cases, the generalised estimating equation estimator can be badly biased in the presence of outliers.

We simulated 50 clusters, each with cluster size of 10, from a simple linear model $y_i = \beta x_i + \varepsilon_i$, where the true $\beta = 1$, the covariate x_i is $(0{\cdot}1, 0{\cdot}2, \ldots, 1{\cdot}0)'$ and the errors ε_i were generated jointly from a 10-variate normal distribution with mean 0, marginal variance 1 and correlation structure either first-order autoregressive, AR(1), or exchangeable. We created one contaminated cluster in which the values of the response were assigned as $bx + \varepsilon$, where b varies from -100 to 100, and ε has the same distribution as ε_i above. These outliers are specified according to the well-known location-shift violation model, in which the true model corresponds to $b = 1$. Figure 1(a) displays the bounded quadratic inference function in (5), Fig. 1(b) displays the bounded estimating functions in (8) derived from the derivative of the quadratic inference function, and Fig. 1(c) shows the unbounded generalised estimating equation in (2).

The redescending property of the M-estimator is very attractive computationally since the solution of (8) is always unique because the estimating function never reaches zero as $|b| \to \infty$.

4. Simulation

We conducted a simple simulation study to demonstrate how the generalised estimating equation and the quadratic inference function estimators are affected by the proportion of contaminated data, where contaminated clusters contain only a single outlying observation.

We used the configuration that led to Fig. 1. In particular, we let the parameter involved in the above two correlation structures be 0·5. Results based on other choices of the parameter show similar patterns to those in Table 1. To contaminate a cluster, one outlier on one subject's response variable was introduced by using $100y_{it}$. The proportions of contaminated clusters were chosen to be 0%, 10%, 20%, 50% and 100%. The results listed in Table 1 are based on 1000 replications.

Table 1: *Simulation study. Average bias of estimates and standard errors, in parentheses, for the quadratic inference function and the generalised estimating equation methods, and the power of the goodness-of-fit test for the model assumption using the quadratic inference function. There are* 1000 *simulations*; $\rho = 0{\cdot}5$.

	AR(1)			Exchangeable			Unspecified		
%	GEE	QIF	Test	GEE	QIF	Test	GEE	QIF	Test
				True correlation is AR(1)					
0	0·001	0·002	0·054	0·004	0·001	0·045	0·004	0·001	0·053
	(0·11)	(0·11)		(0·12)	(0·11)		(0·48)	(0·12)	
10	2·56	0·02	0·235	2·54	0·02	0·350	3·17	0·02	0·235
	(1·19)	(0·13)		(1·18)	(0·22)		(1·60)	(0·13)	
20	5·15	0·03	0·758	5·07	0·03	0·815	7·30	0·03	0·758
	(1·64)	(0·14)		(1·60)	(0·24)		(2·82)	(0·14)	
50	12·91	0·09	0·996	12·35	0·005	0·998	23·35	0·09	0·996
	(2·68)	(0·15)		(2·45)	(0·27)		(6·12)	(0·15)	
100	25·63	0·27	1·00	23·01	0·05	1·00	53·90	0·27	1·00
	(3·58)	(0·18)		(2·76)	(0·33)		(9·78)	(0·179)	
				True correlation is exchangeable					
0	0·007	0·006	0·053	0·005	0·005	0·047	0·013	0·005	0·047
	(0·14)	(0·14)		(0·09)	(0·09)		(0·18)	(0·10)	
10	2·62	0·03	0·241	2·63	0·01	0·309	3·19	0·03	0·241
	(1·23)	(0·17)		(1·23)	(0·29)		(1·63)	(0·17)	
20	5·23	0·08	0·797	5·19	0·11	0·845	7·33	0·08	0·797
	(1·65)	(0·18)		(1·62)	(0·31)		(2·77)	(0·18)	
50	12·78	0·25	0·995	12·38	0·39	0·997	23·00	0·25	0·995
	(2·63)	(0·18)		(2·43)	(0·32)		(5·98)	(0·18)	
100	25·62	0·78	1·00	23·36	1·33	1·00	54·00	0·78	1·00
	(3·74)	(0·18)		(2·94)	(0·33)		(9·98)	(0·18)	

GEE, generalised estimating equation; QIF, quadratic inference function; %, percentage of contamination.

For the quadratic inference function method, we used the extended score in (4). It is known from Qu et al. (2000) that the exchangeable working correlation matrix corresponds to two basis matrices in (3), $M_1 = I$, the identity matrix, and M_2 a matrix with 0 on the diagonal and 1 off the diagonal. Similarly, the AR(1) working correlation corresponds to $M_1 = I$, M_2 with 1 on the two main off-diagonals and 0 elsewhere, and M_3 with 1 at the corners (1, 1) and (10, 10), and 0 elsewhere. In the simulation study and § 5, the quadratic inference function estimators were obtained by using only two basis matrices, M_1 and M_2, for the AR(1) correlation structure, to avoid the confounding of components in (4).

For the unspecified working correlation, we applied the conjugate gradient quadratic inference function method (Qu & Lindsay, 2003) with $M_1 = I$ and $M_2 = \hat{V}$, where $\hat{V} = (1/N) \sum (y_i - \mu_i)(y_i - \mu_i)'$ is a consistent estimator of the variance matrix of y. We obtained the quadratic inference function estimator of β by minimising $Q(\beta)$ in (5) and the standard errors by the square roots of the diagonal elements of $(\hat{D}'\hat{C}^{-1}\hat{D})^{-1}$, where the quadratic inference function estimator is used in $\hat{D}$ and $\hat{C}$. To yield the generalised estimating equation estimator and the standard errors, we applied Liang & Zeger's (1986) method for the chosen working correlation structures.

We calculated the bias of an estimator as the absolute difference between the estimator $\hat{\beta}$ and the true value $\beta_0 = 1$. From Table 1, it is easy to see that the bias of the quadratic inference function estimator appears marginal, being less than 1 in all cases. In contrast, the bias of the generalised estimating equation estimator is considerably higher and increases dramatically as the contamination proportion rises. In the case of light contamination, 10% say, the bias of the estimator from the quadratic inference function is so small as to be ignorable, being only 0·01–0·03, but that from the generalised estimation equations is as much as three times the true value of the parameter. In the worst case of 100% contamination, the bias of the generalised estimating equation estimator exceeds 50 and 20, respectively, under the unspecified and the AR(1) or exchangeable working correlations. Likewise, the standard error of the generalised estimating equation estimator increases quickly as more contamination is introduced. In such situations, the generalised estimating equation method is useless.

To detect whether or not the model assumption (1) is satisfied for different levels of data contamination, we applied the goodness-of-fit test (Hansen, 1982) introduced in § 2. Given that the test statistic $Q(\hat{\beta})$ follows a $\chi^2(1)$ distribution, we calculated the percentage of 1000 test statistics larger than the critical value 3·84 at the 5% significance level. With no contamination, the test sizes were all around 0·05. The power of the test increases as the contamination level rises. As shown, the power becomes very close to 1 when the contamination proportion is 50% or higher. In conclusion, the goodness-of-fit test is helpful for detecting violation of the model assumption when 10% or more of the clusters are contaminated.

5. DATA EXAMPLES

5·1. *Example of normal responses*

The drug chenodiol is known to be effective in dissolving gallstones in the kidney (Wei & Stram, 1988), but it might increase levels of serum cholesterol, which is a known risk factor for atherosclerotic disease. In the National Cooperative Gallstone Study, patients were assigned to take a high dose of chenodiol or a placebo. One question of major interest is whether or not chenodiol would increase cholesterol levels significantly.

The response variable is the serum cholesterol which was recorded at months 6, 12, 20 and 24, and covariates include treatment, with 0 for placebo and 1 for high dose, baseline cholesterol level and month. There were 67 subjects, 31 in the placebo group and 36 in the drug group.

For the continuous response, we used a linear model with the identity link,

$$\mu_{ij} = \beta_0 + \beta_1 \text{Drug} + \beta_2 \text{Baseline} + \beta_3 \text{Month} + \beta_4 (\text{Drug} \times \text{Month}).$$

Table 2 consists of two subtables, corresponding to the analyses under the original data and the contaminated data, respectively. The latter were created by altering the observed level of 387 to 5000 for the subject in the placebo group who had the largest baseline serum cholesterol level. Thus, in the contaminated data there is only one cluster contaminated by a single outlying response value.

Table 2: *National Cooperative Gallstone Study data. Estimators of the regression parameters using the generalised estimating equation and quadratic inference function methods for the original data and the contaminated data with a single high leverage case for a subject in the placebo group. The bold figures indicate significant changes of estimators or Z scores.*

	Original data			Contaminated data		
	Ind (Exch)	AR(1) (GEE)	AR(1) (QIF)	Ind (Exch)	AR(1) (GEE)	AR(1) (QIF)
Intercept	247·81	247·60	248·73	274·89	274·61	249·03
SE	4·44	4·50	4·16	22·96	22·71	4·17
Z	55·83	54·99	59·84	11·97	12·09	59·72
Drug	8·05	8·33	6·84	**−13·10**	**−12·87**	**6·44**
SE	5·94	6·14	5·68	21·90	21·68	5·71
Z	1·36	1·36	1·20	−0·60	−0·60	1·13
Base	0·66	0·66	0·66	1·96	1·95	0·66
SE	0·06	0·06	0·06	1·02	1·01	0·06
Z	11·44	11·15	11·36	**1·92**	**1·93**	**11·00**
Month	1·02	0·99	1·14	7·51	7·50	1·34
SE	0·27	0·26	0·28	6·39	6·38	0·34
Z	3·86	3·74	4·02	**1·18**	**1·18**	**3·94**
Drug × Month	−0·69	−0·65	−0·84	−7·17	−7·17	−1·08
SE	0·41	0·42	0·39	6·40	6·39	0·47
Z	−1·66	−1·55	−2·16	−1·12	−1·12	−2·30

Ind, independent working correlation; Exch, exchangeable working correlation; GEE, generalised estimating equation; QIF, quadratic inference function; SE, standard error.

The generalised estimating equation estimates are listed under three different working correlations, namely independence, exchangeable and AR(1). Note that the generalised estimating equation and the quadratic inference function estimators are indeed identical under independence or exchangeable correlation for balanced data, since the number of equations in (4) is the same as the number of parameters, because of confounding. We combine them in the second and fifth columns of Table 2 for the original data and the contaminated data respectively.

Comparing the two subtables in Table 2, we note that in the presence of a single high leverage point the generalised estimating equation method failed to give reasonable estimators: under all chosen working correlations the generalised estimating equation estimates differ widely between the settings of the original and the contaminated data. In particular, the sign of the drug effect switches from positive to negative, leading to a completely opposite interpretation of the relationship between the use of chenodiol and the serum cholesterol level. In addition, the baseline effect changes from highly significant to insignificant, as does the effect of month. In contrast, the quadratic inference function method under the AR(1) working correlation performs very robustly.

Figure 2 plots the residuals against the baseline cholesterol level under the AR(1) working correlation. Figure 2(a) indicates that the added high leverage point is very influential on the generalised estimating equation regression line as the residuals clearly show a downward trend, which implies that the modelled trend had to be pulled up towards the outlier. However, Fig. 2(b) does not show any trend, suggesting that the quadratic inference function regression line was not affected by the outlier.

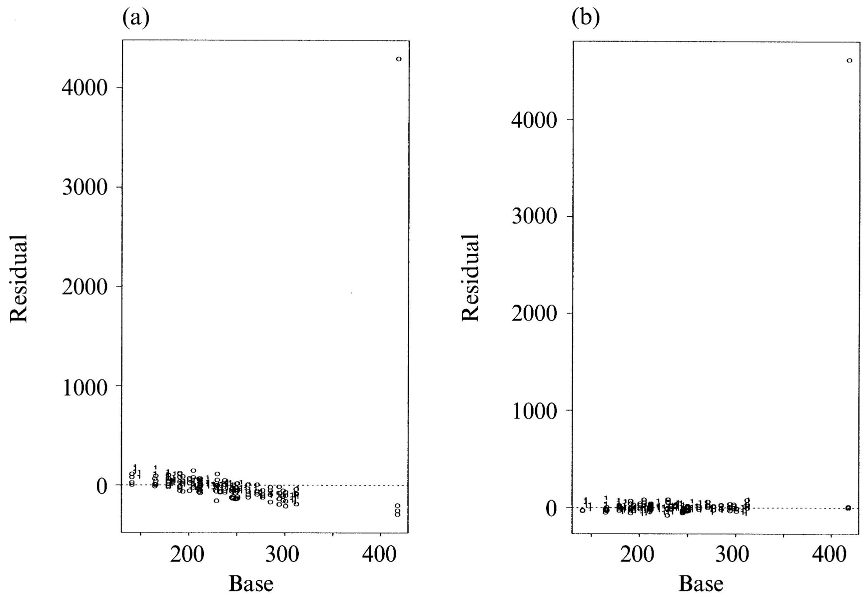

Fig. 2: Wei & Stram's (1988) data. Residual versus baseline cholesterol based on (a) the generalised estimating equation method using AR(1) working correlation, and (b) the quadratic inference function method using AR(1) working correlation.

Finally we tested the goodness of fit of the model assumption in (1) for the contaminated data, and obtained the test statistic $Q(\hat{\beta}) = 17{\cdot}513$. Based on the $\chi^2(5)$ distribution, the corresponding p-value is 0·004, indicating that this model assumption is not satisfied. Clearly, the quadratic inference function is more tolerant of the violation of the model assumption than is the generalised estimating equation.

5·2. *Example of binary responses*

We now consider the data example from Preisser & Qaqish (1999) on urinary incontinence. The response variable is binary, indicating whether or not the subject's daily life is bothered by accidental loss of urine with 1 corresponding to bothered and 0 otherwise. Subjects are correlated if they are from the same hospital practice. There are

137 patients from 38 practices, and each cluster contains at least 1 patient and at most 8 patients. There are five covariates, gender ('female'), age ('age'), daily leaking accidents ('dayacc'), severity of leaking ('severe') and number of times to use the toilet daily ('toilet').

The logistic link function is assumed for the marginal model, so that

$$\text{logit}\,(\mu_{ij}) = \beta_0 + \beta_1\text{female} + \beta_2\text{age} + \beta_3\text{dayacc} + \beta_4\text{severe} + \beta_5\text{toilet},$$

where μ_{ij} denotes the probability of being bothered for patient j in cluster i. For the logistic regression, the matrix A_i in (2) is a diagonal matrix with diagonal elements $A_{ij} = \mu_{ij}(1 - \mu_{ij})$, for $j = 1, \ldots, n_i$. The exchangeable working correlation is assumed to account for the within-cluster correlation for both the generalised estimating equation and the quadratic inference function.

Table 3: *Urinary incontinence data. Estimators and standard errors of the regression parameters obtained by the generalised estimating equation and quadratic inference function methods using exchangeable working correlation structure. The bold figures indicate that $|Z|$ scores are significant at test size* 0·05 *and are greater than* 1·96.

	GEE		QIF		GEE		QIF	
	Est	Z	Est	Z	Est	Z	Est	Z
	All observations				Remove 8th patient			
Intcpt	−3·05	**−3·18**	−3·66	**−2·81**	−2·66	**−3·23**	−3·13	**−2·69**
SE	0·96		1·31		0·82		1·17	
Female	−0·75	−1·24	−1·06	**−2·02**	−1·08	−1·92	−1·38	**−2·75**
SE	0·60		0·53		0·56		0·50	
Age	−0·68	−1·21	−0·66	−1·16	−0·91	−1·58	−0·81	−1·51
SE	0·56		0·56		0·58		0·53	
Dayacc	0·39	**4·20**	0·60	**3·59**	0·46	**4·65**	0·60	**4·20**
SE	0·09		0·17		0·10		0·14	
Severe	0·81	**2·26**	0·60	1·51	0·65	1·96	0·55	1·59
SE	0·36		0·40		0·33		0·35	
Toilet	0·11	1·09	0·25	**2·51**	0·14	1·23	0·23	**2·12**
SE	0·10		0·10		0·12		0·11	
	Remove 44th patient				Remove both patients			
Intcpt	−3·37	**−3·27**	−4·04	**−2·91**	−3·02	**−3·15**	−3·67	**−2·89**
SE	1·03		1·39		0·96		1·27	
Female	−0·76	−1·19	−1·02	−1·92	−1·14	−1·94	−1·48	**−2·80**
SE	0·64		0·53		0·59		0·53	
Age	−0·78	−1·32	−0·85	−1·44	−1·08	−1·74	−0·99	−1·78
SE	0·59		0·59		0·62		0·55	
Dayacc	0·39	**3·86**	0·61	**3·38**	0·47	**4·44**	0·59	**3·75**
SE	0·10		0·18		0·11		0·16	
Severe	0·72	**2·06**	0·64	1·54	0·53	1·76	0·61	1·63
SE	0·35		0·41		0·30		0·37	
Toilet	0·21	**2·08**	0·29	**2·54**	0·27	**2·64**	0·33	**2·80**
SE	0·10		0·11		0·10		0·12	

Est, estimate; SE, standard error; Z, Z-score; Intcpt, intercept; GEE, generalised estimating equation; QIF, quadratic inference function.

Patients 8 and 44 were identified as possible outliers by Preisser & Qaqish's (1996, 1999) diagnostics. Preisser & Qaqish (1999) applied two downweighting strategies of 'Mallows' and 'Schweppe' classes. Without the use of the downweighting strategies, the generalised estimating equation estimators are very sensitive to these two outliers. For example, in Table 3 both covariates 'dayacc' and 'severe' are significant based on the full data, and yet 'severe' becomes insignificant if patient 8 is removed, with or without patient 44, and 'toilet' becomes significant with patient 44 being removed, with or without patient 8.

In contrast, the quadratic inference function provides fairly robust estimates with or without patient 44 and/or patient 8. The only covariate causing doubt is 'female'. Its Z score is $-1{\cdot}92$ with p-value equal to 0·055 for the scenario without patient 44 but with patient 8, but it is clearly significant for the other three scenarios. Overall, 'female' may be concluded to be a significant factor.

Preisser & Qaqish (1999) applied a downweighting strategy based on the 'Mallows' class, and obtained results very similar to the generalised estimating equation results using the full data. They also applied the 'Schweppe' class downweighting strategy and found that 'age' and 'toilet' were significant but 'severe' became insignificant. The quadratic inference function method agrees on variables 'toilet' and 'severe' with their 'Schweppe' downweighting generalised estimating equation method; however, these two methods differ on variables 'age' and 'female'.

Acknowledgement

We appreciate many detailed suggestions and editorial attention from an associate editor and the editor. This research was supported in part by a Grant from the U.S. National Science Foundation to Qu and a Grant from the Natural Sciences and Engineering Research Council of Canada to Song. Part of this research was done when Song was visiting the Department of Biostatistics, University of Michigan. He acknowledges support from the department. The authors also thank C. Vecoli for his proof-reading.

References

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Am. Statist. Assoc.* **69**, 383–93.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. & STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* New York: John Wiley.

HANSEN, L. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–54.

HE, X., ZHU, Z.-Y. & FUNG, W.-K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **90**, 579–90.

HOLLAND, P. W. & WELSH, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Commun. Statist.* A **6**, 813–28.

HUBER, P. J. (1981). *Robust Statistics.* New York: John Wiley.

KÜNSCH, H. R., STEFANSKI, L. A. & CARROLL, R. J. (1989). Conditionally unbiased bounded-influence estimation in general regression models, with application to generalized linear models. *J. Am. Statist. Assoc.* **84**, 460–6.

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

MILLS, J. E., FIELD, C. A. & DUPUIS, D. J. (2002). Marginally specified generalized linear mixed models: A robust approach. *Biometrics* **58**, 727–34.

PREGIBON, D. (1982). Resistant fits for some commonly used logistic models with medical applications. *Biometrics* **38**, 485–98.

PREISSER, J. S. & QAQISH, B. F. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika* **83**, 551–62.

PREISSER, J. S. & QAQISH, B. F. (1999). Robust regression for clustered data with application to binary responses. *Biometrics* **55**, 574–9.

QU, A. & LINDSAY, B. G. (2003). Building adaptive estimating equations when inverse-of-covariance estimation is difficult. *J. R. Statist. Soc.* B **65**, 127–42.

QU, A., LINDSAY, B. G. & LI, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–36.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*. New York: John Wiley.

RONCHETTI, E. M. & TROJANI, F. (2001). Robust inference with GMM estimators. *J. Economet.* **101**, 37–69.

WEI, L. J. & STRAM, D. O. (1988). Analysing repeated measurements with possibly missing observations by modeling marginal distributions. *Statist. Med.* **7**, 139–48.

[*Received June* 2003. *Revised November* 2003]