

Robust Inference Progress from Independent to Longitudinal Setup

Brajendra C. Sutradhar

Abstract In practice one may not have always smooth data. When bulk of the data are smooth but the complete data set apparently contains a few contaminated observations or outliers, one encounters difficulties to choose an inference technique because of the fact that the traditional inference techniques developed for smooth data analysis may no longer provide unbiased and consistent estimates for the desired parameters such as regression parameters in linear or generalized linear models (GLMs) setup. In this paper, we first briefly review some of the widely used bias corrected techniques in linear model setup. But, as opposed to the linear models in normal or other continuous exponential family based variables, the robust inference for discrete data in the GLMs setup, such as for count and binary data, is, however, not adequately discussed in the literature. The advantages and drawbacks of an existing outliers resistant Mallows's type quasi-likelihood (MQL) estimation approach in GLMs setup are reviewed in brief. We then discuss a recently proposed fully standardized MQL (FSMQL) approach that provides almost unbiased estimates ensuring its higher consistency performance. One encounters further challenges when the data in GLMs setup are repeatedly collected over a period of time. This is mainly because one then requires to modify the FSMQL type estimation approaches such that the modified approach also accommodates the correlation structure of the repeated data. A recently proposed robust generalized QL (RGQL) approach is reviewed for the purpose.

1 Introduction

In a regression setup, the responses whether linear, count, or binary, are generated as a function of certain suitable covariates. If bulk of the responses appear to be close to the mean function of the responses with a few remaining responses appearing

B.C. Sutradhar (✉)
Memorial University, St. John's, NF, Canada, A1C5S7
e-mail: bsutradh@mun.ca

at a significant distance from the mean function, then these latter few responses are considered to be potential outliers. In general these outliers occur because of the corresponding covariates which may be contaminated in some ways, and they are referred to as the mean shifted outliers. In some situations, a response may be considered as an outlier because of its inflated variance as compared to the bulk of the responses. It is of main interest to understand the regression model appropriate for bulk of the good responses. But the use of few outlying responses may distort the inference for the bulk of the responses. There are at least two ways this inference problem has been tackled in the literature.

First, it is attempted to detect the outliers and exclude them for the overall inference. For some justifications on this, one may be referred to [Hampel et al. \(1986, Sect. 1.4\)](#) among others. For the purpose, many researchers have discussed the so-called maximum studentized residual (MSR) and maximum normed residual (MNR) tests for detection of outliers in a linear regression setup for independent data. For example, one may refer to the work of [Srikantan \(1961\)](#), [Stefansky \(1971, 1972\)](#), [Tietjen et al. \(1973\)](#), [Prescott \(1975\)](#), [Lund \(1975\)](#), [Bailey \(1977\)](#), [Johnson and Prescott \(1975\)](#), [Ellenberg \(1973, 1976\)](#), [Cook and Prescott \(1981\)](#), [Doornbos \(1981\)](#), and [Beckman and Cook \(1983, Sect. 4\)](#), among others. The powers of these two statistics in detecting outliers may also be affected by the ways the parameters of the regression models are estimated. For a discussion on this, see, for example, a relatively recent work by [Sutradhar et al. \(2007\)](#). In second approach, a robust weighted distance function is constructed such that the suspected outliers get smaller weights. Next the distance function is minimized for the estimation of the regression effects. Some of the existing widely used robust procedures are: Minimax estimation, M-estimation, L-estimation, and R-estimation. For details on these procedures, see, for example, [Hampel et al. \(1986\)](#), [Rousseeuw and Leroy \(1987\)](#), and [Huber \(2004\)](#), and the references therein.

In the independent setup, some authors such as [Cantoni and Ronchetti \(2001\)](#), among others, have suggested a Mallow's type quasi-likelihood (MQL) robust estimation approach to obtain a consistent estimate for the regression effects involved in the model. For the MQL construction, they use the Huber's robust function but did not use the inverse of the variance of such a function to make the MQL standardized. Recently, [Bari and Sutradhar \(2010a\)](#) have improved this estimating equation and introduced a fully standardized MQL (FSMQL) estimating equation that provides regression estimates with smaller bias. In this paper, we review these MQL and FSMQL approaches for the estimation of the regression effects involved in generalized linear models (GLMs), for example for binary and count data.

Also, there have been some studies using QL or generalized estimating equations (GEE) approaches for robust regression estimation in the longitudinal setup. For example, [Preisser and Qaqish \(1999\)](#) have used a resistant GEE (REGEE) approach, which was improved by [Cantoni \(2004\)](#) (see also [Sinha 2006](#) for a random effects approach) by using a semi-standardized MQL (SSMQL; see also [Bari and Sutradhar 2010b](#)) approach. In the second part of the paper, we review these approaches including the robust GQL (RGQL) approach discussed by [Bari and Sutradhar](#)

(2010b) and point out their advantages and drawbacks. Both count and binary longitudinal models are considered.

2 Robust Inference in Regression Models in Independent Setup

2.1 Inference for Linear Models

There exists a vast literature for robust inference in linear models for independent data in the presence of one or more outliers. See, for example, [Rousseeuw and Leroy \(1987\)](#), [Huber \(2004, Chap. 7\)](#), and a relatively recent paper by [Sutradhar et al. \(2007\)](#). These studies mainly deal with outliers in normal responses. For simplicity consider a simple linear regression model

$$y = X\beta + \varepsilon, \quad (1)$$

where $y = (y_1, \dots, y_i, \dots, y_K)'$ is a $K \times 1$ response vector, X is known design matrix of order $K \times p$, β is a $p \times 1$ vector of unknown parameters, and ε is an $K \times 1$ error variable distributed as $\varepsilon \sim N(0, \sigma^2 I_K)$, I_K being the $K \times n$ identity matrix. Usually, each observation in a realization (y, X) contributes to the evaluation of the regression coefficient β . The contribution of one observation, however, may be discordant to the point of sensibly determining the value of a regression parameter. Such an observation is said to be an outlier. To see how an outlier can perturb the linear model (1), two types of outliers are generally considered. They are (a) mean shifted outliers, also referred to as the additive outliers, and (b) variance inflated outliers, also referred to as the innovative or multiplicative outliers.

To construct an additive outlier model, one can perturb the linear model (1) and write

$$y = X\beta + \tilde{\varepsilon}, \quad (2)$$

where $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_i, \dots, \tilde{\varepsilon}_K)'$ is related to ε in (1) as

$$\tilde{\varepsilon}_j = \begin{cases} \varepsilon_j + \delta_1, & \text{for } j = i \\ \varepsilon_j, & \text{for } j \neq i, \end{cases} \quad (3)$$

where for $|\delta_1| > 0$, $y_i = x_i' \beta + \tilde{\varepsilon}_i$ is certainly a discordant observation when compared to the other $K - 1$ observations. It is clear from (1) and (3) that

$$\tilde{\varepsilon} \sim N(\delta, \sigma^2 I_K),$$

where

$$\delta = [01'_{i-1}, \delta_1, 01'_{K-i}]'.$$

To construct a variance inflated outlier model, one can perturb the model (1) as

$$y = X\beta + \varepsilon^*, \quad (4)$$

where $\varepsilon^* = (\varepsilon_1^*, \dots, \varepsilon_i^*, \dots, \varepsilon_K^*)'$ is related to ε in (1) as

$$\varepsilon_j^* = \begin{cases} \varepsilon_j / \sqrt{\omega}, & \text{for } j = i \\ \varepsilon_j, & \text{for } j \neq i, \end{cases} \quad (5)$$

where for $\omega \rightarrow 0$, the i th observation y_i will have large variance leading this observation to be an outlier. It is clear from (1) and (5) that

$$\varepsilon^* \sim N(0, V_\omega = \sigma^2 \text{diag}[1'_{i-1}, 1/\omega, 1'_{K-i}]).$$

Thus, under model (2), bulk $(K-1)$ of the error variables follow $N(0, \sigma^2)$ distribution and 1 follows $N(\delta_1, \sigma^2)$. This is equivalent to say that the $\tilde{\varepsilon}_i$ in model (2) are independent, identically distributed with the common underlying distribution

$$F(\tilde{\varepsilon}) = \left(1 - \frac{1}{K}\right) \Phi\left(\frac{\tilde{\varepsilon} - 0}{\sigma}\right) + \frac{1}{K} \Phi\left(\frac{\tilde{\varepsilon} - \delta_1}{\sigma}\right),$$

(Huber 2004, Example 1.1) where $\Phi(\cdot)$ is the standard normal cumulative. Similarly, one may say that ε_i^* under model (4) are independent, identically distributed with common underlying distribution

$$F(\varepsilon^*) = \left(1 - \frac{1}{K}\right) \Phi\left(\frac{\varepsilon^* - 0}{\sigma}\right) + \frac{1}{K} \Phi\left(\frac{\varepsilon^* - 0}{\sigma/\sqrt{\omega}}\right).$$

2.1.1 Robust Estimation of Regression Effects

It is understandable that the ordinary least square (LS) estimator

$$\hat{\beta}_{LS} = [X'X]^{-1}X'y \quad (6)$$

is biased for β under model (2)–(3) and will be unbiased but inefficient under model (4)–(5). There exist various robust approaches for the consistent estimation of β irrespective of the underlying model whether it is (2)–(3) or (4)–(5). Here we briefly describe two of the approaches, for example.

Huber's Robust Weights Based Iterative Re-weighted Least Square Approach

This estimate is obtained via an iterative re-weighted least squares (RWLS) method (Street et al. 1988). For p components of β , in this approach one solves the robust weights based estimating equation

$$\sum_{j=1}^K \xi_j x_{ju} (y_j - x'_j \beta) = 0, \quad u = 1, \dots, p, \quad (7)$$

where x_{ju} is the u th component of the x_j vector, and

$$\xi_j = \frac{\psi(r_j)}{r_j}, \quad (8)$$

with $\psi(r_j)$ as the Huber's bounded function of r_j given by

$$\psi(z) = \max[-a, \min(z, a)], \quad \text{with } a = 1.25,$$

where $r_j = (y_j - x'_j \beta_{r(0)}^*) / \tilde{\sigma}$ for $j = 1, \dots, n$, with $\beta_{r(0)}^*$ as an initial robust estimate of β which may be obtained by minimizing the L_1 distance $\sum_{j=1}^K |y_j - x'_j \beta|$, and $\tilde{\sigma}$ as a robust estimate of σ given by

$$\tilde{\sigma} = \text{Median} \left\{ \text{largest } K-p+1 \text{ of the } \frac{|y_j - x'_j \beta_{r(0)}^*|}{0.6745} \right\}.$$

Note that if $r_j = 0$, one uses $\xi_j = 1$. The solution to (7) may then be obtained as

$$\beta_{r(1)}^* = (X' \Omega X)^{-1} X' \Omega y, \quad (9)$$

where $\Omega = \text{diag}[\xi_1, \dots, \xi_K]$. This $\beta_{r(1)}^*$ replaces $\beta_{r(0)}^*$ and provides us with a new start and new weights for an improved estimate of β to be obtained by (9). This cycle of iterations continues until convergence. Let the final solution be denoted by $\hat{\beta}_{r(1)}$.

An Alternative Weights Based Iterative RWLS Approach

Rousseeuw and Leroy (1987, Chap. 5) suggest a least median of squares (LSM) approach where the scale parameter to compute the residual is estimated using robust weights different than Huber's weights used in the last section. In fact one can use the iterative least square approach discussed in the last section by replacing Huber's weights with these new weights suggested by Rousseeuw and Leroy (1987, p. 202). See, for example, Sutradhar et al. (2007) for a comparison

between RWLS approaches using Huber's and Rousseeuw and Leroy weights. To be specific, Rousseeuw and Leroy robust weights are defined as

$$\tilde{w}_j = \begin{cases} 1, & \text{if } |d_{j(\beta_{r(0)}^*)}/\tilde{s}_0| \leq 2.5 \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

where $d_{j(\beta_{r(0)}^*)} = y_j - x'_j \beta_{r(0)}^*$ and \tilde{s}_0 is given by

$$\tilde{s}_0 = 1.4826(1 + 5/(K - p)) \sqrt{\text{Median } d_{j(\beta_{r(0)}^*)}^2}.$$

These robust weights in (10) are then used to compute an $\tilde{\Omega}$ matrix as

$$\tilde{\Omega} = \text{diag}[\tilde{w}_1, \dots, \tilde{w}_j, \dots, \tilde{w}_K],$$

which is then used to obtain a first step improved robust estimate for β as

$$\beta_{r(1)}^{**} = (X' \tilde{\Omega} X)^{-1} X' \tilde{\Omega} y. \quad (11)$$

The cycle of iterations continues until convergence. Let this final RWLS estimate be denoted by $\hat{\beta}_{r(2)}$.

2.1.2 Robust Estimation of Variance Component

Note that in the linear model setup, the LS estimate of σ^2 is obtained by computing the residual sum of squares based on the least square estimate of β . That is, $\hat{\sigma}_{ls}^2 = \sum_{j=1}^K (y_j - x'_j \hat{\beta}_{ls})^2 / (K - p)$. Under the linear model in the presence of outliers, one may obtain LS estimate of σ^2 simply by replacing $\hat{\beta}_{ls}$ with $\hat{\beta}_{r(1)}$ or $\hat{\beta}_{r(2)}$ obtained in the last section. Thus the LS estimator for σ^2 has the formula

$$\tilde{\sigma}_{ls(1)}^2 = \sum_{j=1}^K (y_j - x'_j \hat{\beta}_{r(1)})^2 / (K - p), \quad (12)$$

or

$$\tilde{\sigma}_{ls(2)}^2 = \sum_{j=1}^K (y_j - x'_j \hat{\beta}_{r(2)})^2 / (K - p). \quad (13)$$

Huber's Robust Weights Based Iterative RWLS Estimator for σ^2

Following [Street et al. \(1988\)](#), one obtains this estimator as

$$\hat{\sigma}_{r(1)}^2 = \sum_{j=1}^K \xi_j [y_j - x'_j \hat{\beta}_{ls}]^2 / \left(\sum_{j=1}^K \xi_j - tr\{(X' \Omega^2 X)(X' \Omega X)^{-1}\} \right), \quad (14)$$

where ξ_j ($j = 1, \dots, K$) is the j th robust weight to protect the estimate against possible outliers, and $\Omega = diag(\xi_1, \dots, \xi_j, \dots, \xi_K)$. To be specific, ξ_j is defined as $\xi_j = \psi(r_j)/r_j$ with $r_j = (y_j - x'_j \hat{\beta}_{ls})/s^*$, where

$$s^* = \text{Median} \left\{ \text{largest } K-p+1 \text{ of the } \frac{|y_j - x'_j \hat{\beta}_{ls}|}{0.6745} \right\}.$$

Note that the ψ function involved in ξ_j in (14) is the same Huber’s robust function used in (8).

Rosseeuw and Leroy Weights Based Robust Estimator for σ^2

This robust estimator is computed following [Rousseeuw and Leroy \(1987, p. 202, Eq. \(1.5\)\)](#). More specifically, in this approach, robust weights are defined as

$$w_j = \begin{cases} 1, & \text{if } |d_{j(\hat{\beta}_{ls})}/s_0| \leq 2.5 \\ 0, & \text{otherwise,} \end{cases}$$

where $d_{j(\hat{\beta}_{ls})} = y_j - x'_j \hat{\beta}_{ls}$ and s_0 is given by

$$s_0 = 1.4826(1 + 5/(K - p)) \sqrt{\text{Median } d_{j(\hat{\beta}_{ls})}^2}.$$

Next, these weights are exploited to compute the estimator, say $\hat{\sigma}_{r(2)}^2$, as

$$\hat{\sigma}_{r(2)}^2 = \left(\sum_{j=1}^K w_j d_{j(\hat{\beta}_{ls})}^2 \right) / \left(\sum_{j=1}^K w_j - p \right). \quad (15)$$

2.1.3 Finite Sample Performance of the Robust Estimators: An Illustration

[Sutradhar et al. \(2007\)](#) conducted a simulation study to examine the performance of the robust methods as compared to the LS method in estimating the parameters in a linear model when the data contain a few variance inflated outliers. Here, we refer to some of the results of this study, for example. Consider a linear model with $p = 2$ covariates so that $\beta = (\beta_1, \beta_2)'$. For the associated $K \times 2$ design matrix X , consider their design configuration:

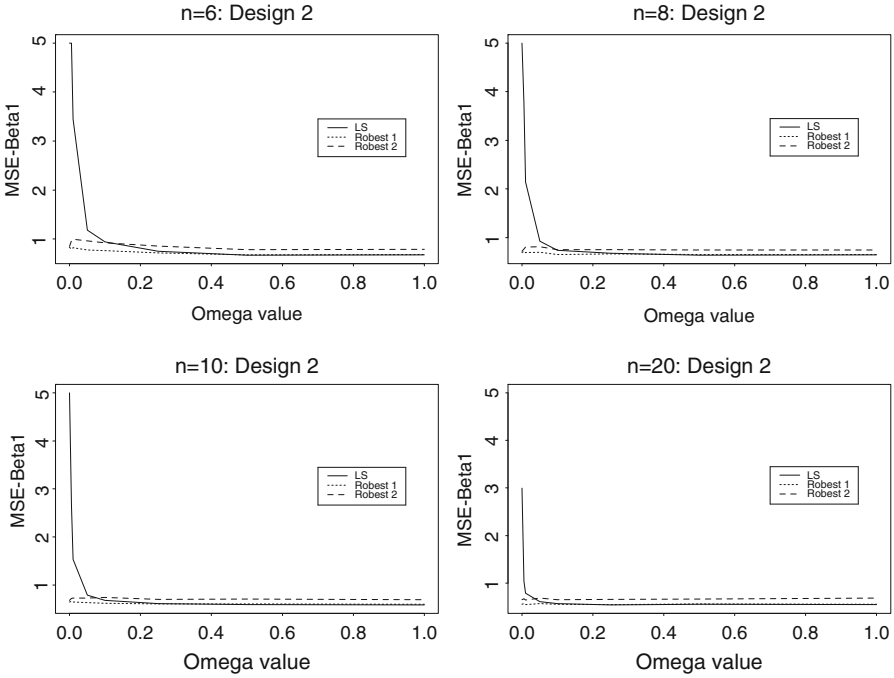


Fig. 1 Mean squared error (MSE) of $\hat{\beta}_{ls,1}$ (LS estimator of β_1), $\hat{\beta}_{r(1),1}$ (first robust estimator of β_2), and $\hat{\beta}_{r(2),1}$ (second robust estimator of β_1)

$$D_2 : x_1 = 1, x_2 = 0, \text{ all other } x(s) \text{ at } 0.5.$$

With regard to the sample size, consider $K(\equiv n) = 6, 8, 10,$ and 20 to examine the effect of small as well as moderately large samples on the estimation. Furthermore, select two locations for the possible outlier, namely locations at $i = 2$ and 3 for $K = 6$; $i = 2$ and 4 for $K = 8$; $i = 2$ and 6 for $K = 10$; and $i = 2$ and 11 for $K = 20$. Also, without any loss of generality, choose $\sigma^2 = 1$, $\beta_1 = 1$, and $\beta_2 = 0.5$. For variance inflation, eight values of ω_i , namely $\omega_i = 0.001, 0.005, 0.01, 0.05, 0.10, 0.25, 0.50,$ and 1.0 , were considered. Note that $\omega_i = 1.0$ represents the case where the data do not contain any outliers, whereas a small value of ω_i indicates that y_i is generated with a large variance implying that y_i can be an influential outlier. The data were simulated 10,000 times. Under each simulation, the LS estimate of β and σ^2 were obtained, which are denoted by $\hat{\beta}_{ls} = (\hat{\beta}_{ls,1}, \hat{\beta}_{ls,2})'$ and $\hat{\sigma}_{ls}^2$, respectively. As far as the robust estimation of β and σ^2 is concerned, these parameters were estimated by using two robust approaches. More specifically, $\hat{\beta}_{r(1)} = (\hat{\beta}_{r(1),1}, \hat{\beta}_{r(1),2})'$ is obtained by using (9), $\hat{\beta}_{r(2)} = (\hat{\beta}_{r(2),1}, \hat{\beta}_{r(2),2})'$ is obtained by using (11), and similarly $\hat{\sigma}_{r(1)}^2$ and $\hat{\sigma}_{r(2)}^2$ are obtained from (14) and (15), respectively. The mean squared errors (MSEs) of these estimators based on 10,000 simulations are displayed in Figs. 1–3, for the estimates of $\beta_1, \beta_2,$ and σ^2 , respectively.

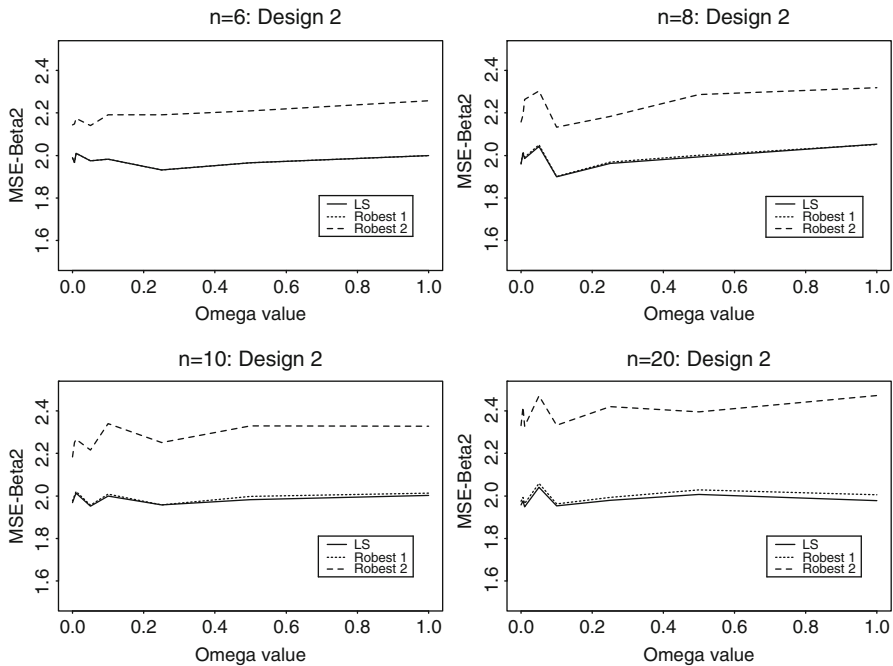


Fig. 2 Mean squared error (MSE) of $\hat{\beta}_{ls,2}$ (LS estimator of β_2), $\hat{\beta}_{r(1),2}$ (first robust estimator of β_2), and $\hat{\beta}_{r(2),2}$ (second robust estimator of β_2)

In summary, the results of this simulation study indicate that in the presence of a variance inflated outlier, the second robust approach performs worse as compared to the first robust and LS methods in estimating β_1 and β_2 . In estimating σ^2 , the LS method performs very poorly when compared with the robust methods.

2.2 Robust Estimation in GLM Setup For Independent Discrete Data

As opposed to the linear models in normal or other continuous exponential family based variables, the robust inference for discrete data in the GLMs setup, such as for count and binary data, is, however, not adequately discussed in the literature. For $i = 1, \dots, K$, let y_i be a discrete response, such as count or binary, collected from the i th individual, and $x_i = (x_{i1}, \dots, x_{iu}, \dots, x_{ip})'$ be the corresponding p -dimensional observed covariate vector. Note that when the data contain a single outlier, any of the K responses $y_1, \dots, y_i, \dots, y_K$ can be that outlier. Now, in the spirit of the mean shifted linear outlier model (2)–(3), suppose that we consider y_j , $j \neq i$, $i = 1, \dots, K$, for example, to be the outlier because of the covariate for the j th individual, namely x_j is contaminated. Note that if $\tilde{x}_i = (\tilde{x}_{i1}, \dots, \tilde{x}_{iu}, \dots, \tilde{x}_{ip})'$ denotes the p -dimensional

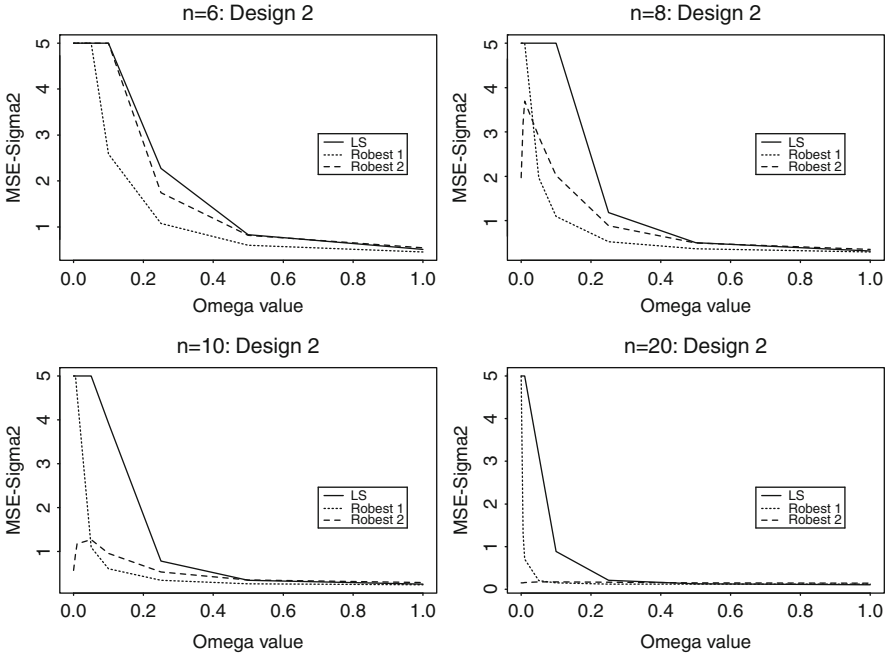


Fig. 3 Mean squared error (MSE) of $\hat{\sigma}_{LS}^2$ (LS estimator of σ^2), $\hat{\sigma}_{r(1)}^2$ (first robust estimator of σ^2), and $\hat{\sigma}_{r(2)}^2$ (second robust estimator of σ^2)

uncontaminated covariate vector corresponding to y_i for all $i = 1, \dots, K$, then for a positive vector δ , the observed covariates $\{x_i\}$ may be related to the uncontaminated covariates $\{\tilde{x}_i\}$ as

$$x_j = \tilde{x}_j + \delta,$$

$$\text{but } x_i = \tilde{x}_i, \text{ for } i \neq j, i = 1, \dots, K. \tag{16}$$

It is of primary interest to estimate $\beta = (\beta_1, \dots, \beta_u, \dots, \beta_p)'$, the effects of uncontaminated covariates \tilde{x}_i on the response y_i . But, as not all the \tilde{x}_i 's are observed, one cannot use them to estimate β , instead the observed contaminated x_i 's are used, which causes bias and hence inconsistency in the estimators.

2.2.1 Understanding Outliers in Count and Binary Data

K Count Observations with a Single Outlier

First assume that in the absence of outliers, $y_1, \dots, y_i, \dots, y_K$ are generated following the Poisson density $P(Y_i = y_i) = [\exp(-\mu_i)\mu_i^{y_i}]/y_i!$, with $\mu_i = \exp(\tilde{x}_i'\beta)$ with $\tilde{x}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})'$. Suppose that the values of these two covariates arise from

$$\tilde{x}_{i1} \stackrel{iid}{\sim} N(0.5, 0.25) \text{ and } \tilde{x}_{i2} \stackrel{iid}{\sim} N(0.5, 0.5),$$

respectively, for all $i = 1, \dots, K$. Suppose that j is the index for the outlying observation that takes a value between 1 and K .

Now, to consider y_j as an outlying value, that is, to have a data set of size K with one outlier, one may then shift the values of \tilde{x}_{j1} and \tilde{x}_{j2} as

$$x_{j1} = \tilde{x}_{j1} + \delta \text{ and } x_{j2} = \tilde{x}_{j2} + \delta, \delta > 0,$$

respectively, but retain $x_{i1} = \tilde{x}_{i1}$ and $x_{i2} = \tilde{x}_{i2}$, for all $i \neq j$. As far as the shifting is concerned, suppose that $\delta = 2.0$. Thus, y_1, \dots, y_K refer to a sample of K count observations with y_j as the single outlier.

K Binary Observations with a Single Outlier

Note that the existing literature (Copas 1988, p. 226; Carroll and Pederson 1993; Sinha 2004) does not provide a clear definition for the outliers in binary data. Remark that Cantoni and Ronchetti (2001) have suggested a practically useful MQL robust inference technique for independent data subject to outliers in GLM setup. However even though GLMs include count and binary models, since the concordant counts (bulk of the observations of similar nature) in the Poisson case and the concordant success numbers in the binomial case can be exploited in a similar way to recognize any possible outliers in the respective data sets, Cantoni and Ronchetti’s (2001) definitions of outliers are appropriate only for the Poisson and binomial cases. Thus, even though binary is a special case of the binomial setup, Cantoni and Ronchetti’s (2001) robust inference development does not appear to be appropriate for the binary data. In view of these difficulties with regard to the robust inferences for the binary case, Bari and Sutradhar (2010a) have provided a new definition for the outliers for the binary data. More specifically, they dealt with one and two sided outliers in the binary data. For convenience these definitions are summarized as follows.

One sided outlier For

$$Pr[Y_i = 1] = E[Y_i] = \mu_i = \frac{\exp(x'_i\beta)}{1 + \exp(x'_i\beta)},$$

and

$$p_{sb} = \max\{\mu_i\}, p_{lb} = \min\{\mu_i\},$$

suppose that the bulk $(K - 1)$ of the binary observations occur with small probabilities such that

$$Pr[Y_i = 1] = \begin{cases} \leq p_{sb} \text{ for } i \neq j, i = 1, \dots, K, \\ > p_{sb} \text{ for } i = j, \end{cases} \tag{17}$$

or, with large probabilities such that

$$Pr[Y_i = 1] = \begin{cases} \geq p_{lb} & \text{for } i \neq j, i = 1, \dots, K, \\ < p_{lb} & \text{for } i = j, \end{cases} \quad (18)$$

Here the binary y_j , whether 1 or 0, satisfying (17) is referred to as an upper sided outlier or satisfying (18) is referred to as a lower sided outlier, whereas the remaining $K - 1$ responses denoted by y_i for $i \neq j$ constitute a group of “concordant” observations.

Two sided outlier It may happen in practice that probabilities for the bulk of the observations lie in the range $p_{sb} \leq P(Y_i = 1) \leq p_{lb}$, leading to a situation where one may encounter a two sided outlier. To be specific, $y_j = 0$ or 1 will be an outlier if either $P(Y_j = 1) > p_{lb}$ or $P(Y_j = 1) < p_{sb}$.

Generation of K binary observations with an outlier We now illustrate the generation of K binary observations including one outlier. For the purpose one may first generate K binary responses $y_1, \dots, y_i, \dots, y_K$ assuming that they do not contain any outliers. To be specific, generate these K “good” responses following the binary logistic model $P(Y_i = 1) = [\exp(\tilde{x}_i' \beta)] / [1 + \exp(\tilde{x}_i' \beta)]$, with two covariates so that $\tilde{x}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})'$ and $\beta = (\beta_1, \beta_2)'$. As far as the covariate values are concerned, similar to the Poisson case, consider two covariates \tilde{x}_{i1} and \tilde{x}_{i2} as

$$\tilde{x}_{i1} \stackrel{iid}{\sim} N(-1.0, 0.25) \text{ and } \tilde{x}_{i2} \stackrel{iid}{\sim} N(-1.0, 0.5),$$

respectively, for $i = 1, \dots, K$.

Next, to create an outlier y_j where j can take any value between 1 and K , change the corresponding covariate values \tilde{x}_{j1} and \tilde{x}_{j2} as

$$x_{j1} = \tilde{x}_{j1} + \delta_1 \text{ and } x_{j2} = \tilde{x}_{j2} + \delta_2, \delta_1, \delta_2 > 0,$$

respectively. Note that for large positive δ_1 and δ_2 , these modified covariates will be increased in magnitude yielding larger probability for $y_j = 1$. One may then treat y_j as an outlier. For convenience, suppose that one uses $\delta_1 = 3.0$ and $\delta_2 = 4.0$. As far as the remaining covariates are concerned, they are kept unchanged. That is, for $i \neq j$ ($i = 1, \dots, K$), consider $x_{i1} = \tilde{x}_{i1}$ and $x_{i2} = \tilde{x}_{i2}$.

2.2.2 Naive and Existing Robust QL Estimation Approaches

Naive QL (NQL) Estimation of β

Had there been no outliers, one could have obtained the consistent estimate of β by solving the well-known QL (quasi-likelihood) estimating equation

$$\sum_{i=1}^K \left[\frac{\partial \tilde{\mu}_i}{\partial \beta} V^{-1}(\tilde{\mu}_i)(y_i - \tilde{\mu}_i) \right] = 0, \quad (19)$$

(see [Wedderburn 1974](#); [McCullagh and Nelder 1989](#); [Heyde 1997](#)) where, for example, $\tilde{\mu}_i = E[Y_i] = \exp(\tilde{x}_i' \beta)$ and $V(\tilde{\mu}_i) = \text{var}[Y_i] = \tilde{\mu}_i$ for Poisson count data; and $\tilde{\mu}_i = E[Y_i] = \exp(\tilde{x}_i' \beta) / [1 + \exp(\tilde{x}_i' \beta)]$ and $V(\tilde{\mu}_i) = \text{var}[Y_i] = \tilde{\mu}_i(1 - \tilde{\mu}_i)$ for binary data. But, as the uncontaminated \tilde{x}_i 's are unobserved, it is not possible to use (19) for the estimation of β . Now suppose that following (19) but by using the observed covariates $\{x_i\}$, one writes the naive quasi-likelihood (NQL) estimating equation for β given by

$$\sum_{i=1}^K \left[\frac{\partial \mu_i}{\partial \beta} V^{-1}(\mu_i)(y_i - \mu_i) \right] = 0, \quad (20)$$

where, for example, $\mu_i = \exp(x_i' \beta)$ and $V(\mu_i) = \mu_i$ for Poisson count data; and $\mu_i = \exp(x_i' \beta) / [1 + \exp(x_i' \beta)]$ and $V(\mu_i) = \mu_i(1 - \mu_i)$ for binary data. Since β is the effect of \tilde{x}_i on y_i for all $i = 1, \dots, K$, it then follows that the quasi-likelihood estimator obtained from (20) will be biased and hence inconsistent for β .

Partly Standardized Mallows Type QL (PSMQL) Estimation of β

As a remedy to the inconsistency of the quasi-likelihood estimator obtained from (20), [Cantoni and Ronchetti \(2001\)](#) (see also references therein), among others, have suggested a Mallows's type quasi-likelihood (MQL) robust estimation approach to obtain a consistent estimate for the regression effects β . For the purpose, for $r_i = \frac{y_i - \mu_i}{\sqrt{V(\mu_i)}}$, they first define the Huber robust function as

$$\psi_c(r_i) = \begin{cases} r_i, & |r_i| \leq c, \\ c \text{ sign}(r_i), & |r_i| > c, \end{cases} \quad (21)$$

where c is referred to as the so-called tuning constant. This robust function is then used to construct the MQL estimating equation given by

$$\sum_{i=1}^K \left[w(x_i) \frac{\partial \mu_i}{\partial \beta} V^{-\frac{1}{2}}(\mu_i) \psi_c(r_i) - a(\beta) \right] = 0, \quad (22)$$

where $a(\beta) = \frac{1}{K} \sum_{i=1}^K w(x_i) \frac{\partial \mu_i}{\partial \beta} V^{-\frac{1}{2}}(\mu_i) E[\psi_c(r_i)]$, with $\mu_i = E(Y_i)$, $V(\mu_i) = \text{var}(Y_i)$, and $w(x_i) = 1$ for the binomial data as in Huber's linear regression case, but $w(x_i) = \sqrt{(1 - h_i)}$ for the Poisson data, where h_i is the i th diagonal element of the hat matrix $H = X(X'X)^{-1}X'$, with $X = (x_1, \dots, x_i, \dots, x_K)'$ being the $K \times p$ covariate matrix.

Note that in order to minimize the robust distance function $\psi_c(r_i)$, the MQL estimating (22) was constructed by using the variance $V(\mu_i) = var(Y_i)$ as a weight function and $\frac{\partial \mu_i}{\partial \beta}$ as a gradient function, whereas a proper estimating equation should use $var(\psi_c(r_i))$ and $\frac{\partial \psi_c(r_i)}{\partial \beta}$ as the weight and gradient functions, respectively. One may therefore refer to the estimating (22) as a partly standardized MQL (PSMQL) estimating equation. This PSMQL estimating (22) provides regression estimates with smaller bias than the traditional maximum likelihood or NQL estimating (20). But, as discussed in Bari and Sutradhar (2010a), this improvement does not appear to be significant enough to recommend the use of the PSMQL estimation approach. Moreover, this PSMQL approach is not suitable for inferences in binary regression models.

FSMQL Estimation of β

As an improvement over the PSMQL estimation, Bari and Sutradhar (2010a) have proposed a FSMQL estimation approach where the regression effects β is obtained by solving the FSMQL estimating equation

$$\sum_{i=1}^K \left[w(x_i) \frac{\partial}{\partial \beta} \left\{ \psi_c(r_i) - \frac{1}{K} \sum_{i=1}^K E(\psi_c(r_i)) \right\} \{var(\psi_c(r_i))\}^{-1} \times \left\{ \psi_c(r_i) - \frac{1}{K} \sum_{i=1}^K E(\psi_c(r_i)) \right\} \right] = 0. \tag{23}$$

Note that this FSMQL estimating (23) is constructed by replacing the “working” variance and gradient functions $V(\mu_i)$ and $\frac{\partial \mu_i}{\partial \beta}$ in (22), with the true variance and gradient functions $var(\psi_c(r_i))$ and $\frac{\partial \psi_c(r_i)}{\partial \beta}$, respectively. Also, $w(x_i) = \sqrt{(1 - h_i)}$ is used in both binary and Poisson cases. Furthermore, the specific formulas for the true weight function $var(\psi_c(r_i))$ and the gradient function $\frac{\partial \psi_c(r_i)}{\partial \beta}$ for the count and binary cases are available from Bari and Sutradhar (2010a, Sects. 2.1 and 2.2).

Bari and Sutradhar (2010a) also considered another version of the FSMQL estimating (23), which was developed by using the deviance $\psi_c(r_i) - E(\psi_c(r_i))$ instead of $\psi_c(r_i) - \frac{1}{K} \sum_{i=1}^K E(\psi_c(r_i))$. This alternative FSMQL estimating equation has the form

$$\sum_{i=1}^K \left[w(x_i) \frac{\partial}{\partial \beta} \{ \psi_c(r_i) - E(\psi_c(r_i)) \} \{var(\psi_c(r_i))\}^{-1} \{ \psi_c(r_i) - E(\psi_c(r_i)) \} \right] = 0. \tag{24}$$

For convenience, one may refer to (23) and (24) as the FSMQL₁ and FSMQL₂ estimating equations, respectively.

Robust Function and Properties for Count Data

For the count data, consider the Huber robust function $\psi_c(r_i)$ as in (21). The expectation and variance of this function are available from [Cantoni and Ronchetti \(2001, Appendix A, p. 1028\)](#). The gradient of the robust function and its expectation may then be computed as follows (see also [Bari and Sutradhar 2010a, Appendix](#)):

$$\frac{\partial \psi_c(r_i)}{\partial \beta} = \begin{cases} -\frac{\mu_i}{V^{\frac{1}{2}}(\mu_i)} x_i, & |r_i| \leq c, \\ 0, & |r_i| > c, \end{cases} \tag{25}$$

and

$$\begin{aligned} \frac{\partial E(\psi_c(r_i))}{\partial \beta} &= -c \left[\frac{\partial}{\partial \beta} F_{Y_i}(i_2) + \frac{\partial}{\partial \beta} F_{Y_i}(i_1) \right] + \frac{\mu_i}{V^{\frac{1}{2}}(\mu_i)} \left[\left\{ x_i P(Y_i = i_1) + \frac{\partial}{\partial \beta} P(Y_i = i_1) \right\} \right. \\ &\quad \left. - \left\{ x_i P(Y_i = i_2) + \frac{\partial}{\partial \beta} P(Y_i = i_2) \right\} \right], \end{aligned} \tag{26}$$

where

$$\begin{aligned} \frac{\partial}{\partial \beta} P(Y_i = i_1) &= P(Y_i = i_1)(i_1 - \mu_i)x_i, \quad \frac{\partial}{\partial \beta} P(Y_i = i_2) = P(Y_i = i_2)(i_2 - \mu_i)x_i, \\ \frac{\partial}{\partial \beta} F_{Y_i}(i_1) &= \sum_{j=0}^{i_1} \frac{\partial}{\partial \beta} P(Y_i = j), \quad \text{and} \quad \frac{\partial}{\partial \beta} F_{Y_i}(i_2) = \sum_{j=0}^{i_2} \frac{\partial}{\partial \beta} P(Y_i = j). \end{aligned}$$

Robust Function and Properties for Binary Data

(a) Robust function in the presence of one sided outlier

Suppose that the bulk of the binary observations occur with small probabilities. In this case, the robust function $\psi_c(r_i)$ ($i = 1, \dots, n$) may be defined as

$$\psi_c(r_i) = \begin{cases} \frac{y_i - \mu_i}{V^{\frac{1}{2}}(\mu_i)}, & P(Y_i = 1) \leq p_{sb}, i \neq j, i = 1, \dots, K, \\ \frac{y_i - \mu_i^{(c_1)}}{V^{(c_1)\frac{1}{2}}(\mu_i^{(c_1)})}, & P(Y_i = 1) > p_{sb}, i = j, \end{cases} \tag{27}$$

where $\mu_i = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)}$, $V(\mu_i) = \mu_i(1 - \mu_i)$ for all $i = 1, \dots, K$, and $p_{sb} = \max\{\mu_i\}$, $i \neq j$, is a bound for all $K - 1$ small probabilities.

Note that as opposed to the case given in (27), if the bulk of the binary observations occur with large probabilities, then the robust function $\psi_c(r_i)$ ($i = 1, \dots, K$) is defined as

$$\psi_c(r_i) = \begin{cases} \frac{y_i - \mu_i}{V^{\frac{1}{2}}(\mu_i)}, & P(Y_i = 1) \geq p_{lb}, i \neq j, i = 1, \dots, K, \\ \frac{y_i - \mu_i^{(c_2)}}{V^{(c_2)\frac{1}{2}}(\mu_i^{(c_2)})}, & P(Y_i = 1) < p_{lb}, i = j, \end{cases} \quad (28)$$

where $p_{lb} = \min\{\mu_i\}$, $i \neq j$, is a bound for all $K - 1$ large probabilities.

(b) *Robust function in the presence of two sided outlier*

In this case, the robust function $\psi_c(r_i)$ ($i = 1, \dots, K$) may be defined as

$$\psi_c(r_i) = \begin{cases} \frac{y_i - \mu_i^{(c_1)}}{V^{(c_1)\frac{1}{2}}(\mu_i^{(c_1)})}, & P(Y_i = 1) > p_{lb}, i = j, \\ \frac{y_i - \mu_i}{V^{\frac{1}{2}}(\mu_i)}, & p_{sb} \leq P(Y_i = 1) \leq p_{lb}, i \neq j, i = 1, \dots, K, \\ \frac{y_i - \mu_i^{(c_2)}}{V^{(c_2)\frac{1}{2}}(\mu_i^{(c_2)})}, & P(Y_i = 1) < p_{sb}, i = j, \end{cases} \quad (29)$$

where $\mu_j^{(c_1)}$ and $V^{(c_1)}(\mu_j^{(c_1)})$ are defined as in (27), whereas $\mu_j^{(c_2)}$ and $V^{(c_2)}(\mu_j^{(c_2)})$ are defined as in (28).

(b(i)) *Basic properties of the robust function $\psi_c(r_i)$: Binary case*

It is convenient to write these properties for the two sided outlier case. The results for the one sided outlier may be obtained as a special case. The expectation, variance, and gradient of the robust function in the presence of a two sided outlier are available from [Bari and Sutradhar \(2010a, Appendix\)](#). For convenience, these properties are summarized as follows.

Let $\psi_c(r_i)$ denote the robust function defined as in (29). The expectation and variance of $\psi_c(r_i)$ are given by

$$E(\psi_c(r_i)) = \frac{\mu_i - \mu_i^{(c_1)}}{V^{(c_1)\frac{1}{2}}(\mu_i^{(c_1)})} P_1 + \frac{\mu_i - \mu_i^{(c_2)}}{V^{(c_2)\frac{1}{2}}(\mu_i^{(c_2)})} P_3, \quad (30)$$

and

$$\text{var}(\psi_c(r_i)) = \frac{(1 - 2\mu_i^{(c_1)})\mu_i + \mu_i^{(c_1)2}}{V^{(c_1)}(\mu_i^{(c_1)})} P_1 + P_2 + \frac{(1 - 2\mu_i^{(c_2)})\mu_i + \mu_i^{(c_2)2}}{V^{(c_2)}(\mu_i^{(c_2)})} P_3 - [E(\psi_c(r_i))]^2, \quad (31)$$

where P_1 , P_2 , and P_3 are the probabilities for a binary observation to satisfy the conditions $P(Y_i = 1) > p_{lb}$, $p_{sb} \leq P(Y_i = 1) \leq p_{lb}$, and $P(Y_i = 1) < p_{sb}$, respectively.

In practice, the probabilities P_1 , P_2 , and P_3 may be computed from the data by using the sample proportions given by, for example,

$$P_1 = \frac{\text{Number of observations satisfying } P(Y_i = 1) > p_{lb}}{\text{Total observation } (K)}.$$

The gradient of the robust function $\psi_c(r_i)$ [defined in (29)] and its expectation are given by

$$\frac{\partial \psi_c(r_i)}{\partial \beta} = \begin{cases} 0, & P(Y_i = 1) > p_{lb}, i = j, \\ \frac{-\mu_i(1-\mu_i)x_i}{V^{\frac{1}{2}}(\mu_i)}, & p_{sb} \leq P(Y_i = 1) \leq p_{lb}, i \neq j, i = 1, \dots, K, \\ 0, & P(Y_i = 1) < p_{sb}, i = j, \end{cases} \quad (32)$$

and

$$\frac{\partial E(\psi_c(r_i))}{\partial \beta} = \frac{(1 - \mu_i)\mu_i x_i}{V^{(c_1)\frac{1}{2}}(\mu_i^{(c_1)})} P_1 + \frac{(1 - \mu_i)\mu_i x_i}{V^{(c_2)\frac{1}{2}}(\mu_i^{(c_2)})} P_3. \quad (33)$$

To illustrate the finite sample based relative performance of the competitive robust approaches, namely PSMQL (22), FSMQL₁ (23), and FSMQL₂ (24) approaches, we refer to some of the simulation results from Bari and Sutradhar (2010a). In the presence of a single outlier, the count and binary data were generated as in Sect. 2.2.1. With $K = 60$ observations including an outlier, the relative bias (RB) of an estimator, for example, for β_k ($k = 1, \dots, p$) given by

$$RB(\hat{\beta}_k) = \frac{|\hat{\beta}_k - \beta_k|}{\text{s.e.}(\hat{\beta}_k)} \times 100, \quad (34)$$

were computed based 1,000 simulations. The results are shown in Table 1.

Table 1 (For count and binary data with one outlier) Simulated means (SM), simulated standard errors (SSE), and relative biases (RB) of the PSMQL, FSMQL₁, and FSMQL₂ estimates of the regression parameters $\beta_1 = 1.0$ and $\beta_2 = 0.5$, for sample size 60 and selected values of the tuning constant $c = 1.4$ under the Poisson model, and tuning constant $\mu^{c_1} = 0.9$ under the binary model, in the presence of one outlier

Model	K	Tuning constant	Statistic	Estimation method					
				PSMQL		FSMQL ₁		FSMQL ₂	
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
Count	60	$c = 1.4$	SM	0.507	0.600	0.899	0.517	0.893	0.488
			SSE	0.206	0.188	0.307	0.239	0.279	0.210
			RB	240	53	33	7	38	6
Binary	60	$\mu^{c_1} = 0.9$	SM	1.161	0.194	0.994	0.503	1.003	0.486
			SSE	0.777	0.760	0.782	0.777	0.779	0.764
			RB	21	40	1	0	0	2

The results of the table show that both fully standardized robust procedures FSMQL₁ and FSMQL₂ perform much better in estimating β as compared to the existing PSMQL robust approach.

3 Robust Inference in Longitudinal Setup

3.1 Existing GEE Approaches for Robust Inferences

Let $\mu_i(x_i) = E(Y_i) = (\mu_{i1}, \dots, \mu_{it}, \dots, \mu_{iT})'$ denote the mean, and $\Sigma_i(x_i, \rho) : T \times T$ be the true covariance matrix of the response vector y_i where x_i represents all true covariates, i.e., $x_i \equiv x_{i1}, \dots, x_{it}, \dots, x_{iT}$. For convenience, the covariance matrix $\Sigma_i(x_i, \rho)$ is often expressed as $\Sigma_i(x_i, \rho) = A_i^{\frac{1}{2}} C_i(\rho) A_i^{\frac{1}{2}}$, where $A_i = \text{diag}[\sigma_{i11}, \dots, \sigma_{iit}, \dots, \sigma_{iT T}]$ and $C_i(\rho)$ is the correlation matrix for repeated binary or count data. Note that if the longitudinal data do not contain any outliers, then one may obtain consistent and highly efficient estimate of β by solving the GQL estimating equation

$$\sum_{i=1}^K \left[\frac{\partial \mu_i'(x_i)}{\partial \beta} \Sigma_i^{-1}(x_i, \hat{\rho})(y_i - \mu_i(x_i)) \right] = 0, \tag{35}$$

(see Sutradhar 2003) where $\hat{\rho}$ is a suitable consistent, for example, a moment estimate of ρ .

Note that in practice it may, however, happen that a small percentage such as 1% of longitudinal observations are suspected to be outliers. Suppose that m of the KT responses are referred to as the outliers when their corresponding covariates are shifted by an amount δ , δ being a real valued vector. For convenience, we denote the new set of covariates as

$$\tilde{x}_{it} = \begin{cases} x_{it} & \text{for } (i, t) \not\equiv (i', t') \\ x_{it} + \delta & \text{for } (i, t) \equiv (i', t') \end{cases}$$

and use these observed covariates \tilde{x}_{it} for the estimation of β . It is, therefore, clear that since β is the effect of the true covariate x_{it} on y_{it} , the solution of the observed covariates \tilde{x}_i based naive GQL (NGQL) estimating equation

$$\sum_{i=1}^K \left[\frac{\partial \mu_i'(\tilde{x}_i)}{\partial \beta} \Sigma_i^{-1}(\tilde{x}_i, \hat{\rho})(y_i - \mu_i(\tilde{x}_i)) \right] = 0, \tag{36}$$

will produce biased and hence inconsistent estimate for β . To overcome this inconsistency problem, Preisser and Qaish (1999), among others, have proposed to solve a resistant generalized quasi-likelihood estimating equation (REGEE) given by

$$\sum_{i=1}^K \left[\frac{\partial \mu'_i(\tilde{x}_i)}{\partial \beta} V_i^{-1}(\tilde{x}_i, \alpha) (\psi_i^* - c_i) \right] = 0, \tag{37}$$

where ψ_i^* is a down-weighting function, $c_i = E(\psi_i^*)$, and $V_i(\tilde{x}_i, \alpha)$ is a “working” covariance matrix (Liang and Zeger 1986). Note that the REGEE in (37) does not appear to be a proper weighted estimating equation. This is because, first, $V_i(\tilde{x}_i, \alpha)$ is only a substitute of $\Sigma_i(\tilde{x}_i, \rho)$ matrix, whereas in the presence of outliers, one needs to use $\Omega_i^* = \text{var}(\psi_i^*)$ in order to obtain efficient regression estimates. Secondly, the REGEE (37) uses $\frac{\partial \mu'_i(\tilde{x}_i)}{\partial \beta}$ as the gradient function, whereas the consistency of the estimates may depend on the proper gradient function constructed by taking the derivative of the $\psi_i^* - c_i$ function with respect to β .

Cantoni (2004) has provided an improvement over the REGEE by introducing the proper gradient function in the estimating equation. To be specific, as compared to Preisser and Qaqish (1999) (see also Eq. (36)), Cantoni (2004) constructed an improved resistant generalized estimating equation (IREGEE) given by

$$\sum_{i=1}^K \left[E \left\{ \frac{\partial(\psi_i^* - c_i)}{\partial \beta'} \right\}' V_i^{-1}(\tilde{x}_i, \alpha) (\psi_i^* - c_i) \right] = 0, \tag{38}$$

where $E \left[\frac{\partial(\psi_i^* - c_i)}{\partial \beta'} \right]$ is a proper gradient of the robust function $\psi_i^* - c_i$, with

$$E \left[\frac{\partial(\psi_i^* - c_i)}{\partial \beta'} \right] = E \left[\frac{\partial(\psi_i^* - c_i)}{\partial \mu'_i(\tilde{x}_i)} \right] \frac{\partial \mu_i}{\partial \beta'} = \Gamma_i \frac{\partial \mu_i}{\partial \beta'}.$$

Note that the estimating (38) still uses a “working” covariance matrix $V_i(\tilde{x}_i, \alpha)$, whereas an efficient estimating equation (Sutradhar and Das 1999) should use the proper covariance matrix of the robust function, namely $\Omega_i^* = \text{var}(\psi_i^*)$. Further, similar to Cantoni (2004), Sinha (2006) has attempted to develop certain robust inferences to deal with outliers in the longitudinal data. But, Sinha (2006) has modeled the longitudinal correlations through random effects, which, therefore addresses a different problem than longitudinal data problems.

Recently, Bari and Sutradhar (2010b) has proposed an auto-correlation class based robust GQL (RGQL) approach for inferences in binary and count panel data models in the presence of outliers. This RGQL approach produces consistent and highly efficient regression estimates, and it is a generalization of the FSMQL approach for independent data to the longitudinal setup. The RGQL approach is summarized in the next section.

3.2 RGQL Approach for Robust Inferences in Longitudinal Setup

Note that when the covariates are stationary, that is, time independent, one may develop a general auto-correlation class based robust GQL estimation approach.

Bari and Sutradhar (2010b) have considered non-stationary covariates and exploited the most likely AR(1) type correlation structures for both count and binary data. These correlation structures are discussed in detail in Sutradhar (2010), see also Sutradhar (2011). For convenience we summarize these correlation structures as follows.

Recall that $x_{it} = (x_{it1}, \dots, x_{itu}, \dots, x_{itp})'$ is the $p \times 1$ vector of covariates corresponding to y_{it} when the data do not contain any outliers, and β denote the effects of the covariate x_{it} on y_{it} . The AR(1) correlation models for repeated responses $y_{i1}, \dots, y_{it}, \dots, y_{iT}$ based on the uncontaminated covariates $x_{i1}, \dots, x_{it}, \dots, x_{iT}$, for binary and count data are given below.

AR(1) model for repeated binary data

For $\mu_{it} = \frac{\exp(x'_{it}\beta)}{1+\exp(x'_{it}\beta)}$, for all $t = 1, \dots, T$, the AR(1) model for the binary data may be written as

$$y_{i1} \sim \text{bin}(\mu_{i1}) \text{ and } y_{it} | y_{i,t-1} \sim \text{bin}[\mu_{it} + \rho(y_{i,t-1} - \mu_{i,t-1})], \tag{39}$$

(Zeger et al. 1985; Qaqish 2003) where ρ is a correlation index parameter. The binary AR(1) model (39) has the auto-correlation structure given by

$$\text{corr}(Y_{iu}, Y_{it}) = \begin{cases} \rho^{t-u} \left[\frac{\sigma_{iuu}}{\sigma_{it}} \right]^{1/2}, & \text{for } u < t \\ \rho^{u-t} \left[\frac{\sigma_{itt}}{\sigma_{uu}} \right]^{1/2}, & \text{for } u > t \end{cases}, \tag{40}$$

where $\sigma_{uu} = \mu_{iu}(1 - \mu_{iu})$, for example, is the variance of y_{iu} . Note that ρ parameter in (39)–(40) must satisfy the range restriction

$$\max \left[-\frac{\mu_{it}}{1 - \mu_{i,t-1}}, -\frac{1 - \mu_{it}}{\mu_{i,t-1}} \right] \leq \rho \leq \min \left[\frac{1 - \mu_{it}}{1 - \mu_{i,t-1}}, \frac{\mu_{it}}{\mu_{i,t-1}} \right]. \tag{41}$$

AR(1) model for repeated count data

As opposed to the binary AR(1) model (39), the AR(1) model for the count data is defined as

$$y_{i1} \sim \text{Poisson}(\mu_{i1}) \text{ and } y_{it} = \rho * y_{i,t-1} + d_{it}, \quad t = 2, \dots, T, \tag{42}$$

(see McKenzie 1988; Sutradhar 2003), where $y_{i,t-1} \sim \text{Poisson}(\mu_{i,t-1})$ and $d_{it} \sim \text{Poisson}(\mu_{it} - \rho\mu_{i,t-1})$, with $\mu_{it} = E(Y_{it}) = \exp(x'_{it}\beta)$. In (42), d_{it} and $y_{i,t-1}$ are assumed to be independent. Also, for given count $y_{i,t-1}$,

$$\rho * y_{i,t-1} = \sum_{j=1}^{y_{i,t-1}} b_j(\rho),$$

where $b_j(\rho)$ stands for a binary variable with $P[b_j(\rho) = 1] = \rho$ and $P[b_j(\rho) = 0] = 1 - \rho$. The AR(1) model (42) for count data has the auto-correlation structure given by

$$corr(Y_{iu}, Y_{it}) = \rho^{t-u} \sqrt{\frac{\mu_{iu}}{\mu_{it}}}, \tag{43}$$

with ρ satisfying the range restriction

$$0 < \rho < \min \left[1, \frac{\mu_{it}}{\mu_{i,t-1}} \right], t = 2, \dots, T. \tag{44}$$

3.2.1 RGQL Estimating Equation

For $\xi_i = [\psi_c(r_{i1}), \dots, \psi_c(r_{it}), \dots, \psi_c(r_{iT})]'$, its expectation λ_i is available from [Cantoni and Ronchetti \(2001\)](#) for the count data, and from Sect. 2.2.2 for the binary case. Recall from (38) that based on “working” covariance of the responses ([Liang and Zeger 1986](#)), [Cantoni \(2004\)](#) has suggested an IREGEE approach for estimating β in the presence of outliers. One may obtain consistent β estimate by solving a slightly different equation than (38) given by

$$\sum_{i=1}^K \left[W_i \frac{\partial}{\partial \beta} \left\{ \xi_i - K^{-1} \sum_{i=1}^K \lambda_i \right\}' V_i^{-1}(\alpha) \left\{ \xi_i - K^{-1} \sum_{i=1}^K \lambda_i \right\} \right] = 0, \tag{45}$$

where $W_i = \text{diag}[w_{i1}, \dots, w_{it}, \dots, w_{iT}]$ is the $T \times T$ covariate dependent diagonal weight matrix so that covariates corresponding to the outlying response yield less weight for the corresponding robust function. To be specific, the t -th diagonal element of the W_i matrix is computed as $w_{it} = \sqrt{1 - h_{it}}$, h_{it} being the t -th diagonal element of the hat matrix $H_i = \tilde{X}_i(\tilde{X}_i' \tilde{X}_i)^{-1} \tilde{X}_i'$ with $\tilde{X}_i = [\tilde{x}_{i1}, \dots, \tilde{x}_{it}, \dots, \tilde{x}_{iT}]'$. See, for example, [Cantoni and Ronchetti \(2001\)](#). Also in (45), $V_i(\alpha) = \text{cov}(Y_i) = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}}$ is a “working” covariance matrix of y_i , with $R(\alpha)$ as the associated “working” correlation matrix. Note that there are twofold problems with this estimating equation. First, for efficiency increase, it would have been appropriate to use $\text{cov}(\xi_i) = \text{cov}[\psi_c(r_{i1}), \dots, \psi_c(r_{it}), \dots, \psi_c(r_{iT})]$ as the weight matrix instead of the true covariance matrix $\Sigma_i(\alpha) = \text{cov}(Y_i)$. Secondly, [Cantoni \(2004\)](#) did not even use Σ_i , rather has used a “working” covariance matrix $V_i(\alpha) = \text{cov}(Y_i)$.

To overcome this inefficiency problem encountered by Cantoni’s approach, [Bari and Sutradhar \(2010b\)](#) have suggested a robust function based GQL (RGQL) estimating equation for β as

$$\sum_{i=1}^K \left[W_i \frac{\partial}{\partial \beta} \left\{ \xi_i - K^{-1} \sum_{i=1}^K \lambda_i \right\}' \Omega_i^{-1} \left\{ \xi_i - K^{-1} \sum_{i=1}^K \lambda_i \right\} \right] = 0, \tag{46}$$

where

$$\Omega_i = cov(\xi_i) = (\omega_{iut}), \tag{47}$$

with

$$\omega_{iut} = E[\psi_c(r_{iu})\psi_c(r_{it})] - \{E(\psi_c(r_{iu}))E(\psi_c(r_{it}))\}, \tag{48}$$

where, as mentioned above, the formulas for $E[\psi_c(r_{it})]$ are available for both count and binary data.

Computation of Ω_i for the Binary Data

Note that the computation of the product moment $E[\psi_c(r_{iu})\psi_c(r_{it})]$ in (48) is manageable for the binary case, but it is extremely difficult for the count data. For example, suppose that y_{it} , $t = 1, \dots, T$, used in the robust functions $\psi_c(r_{it})$, follow an AR(1) type correlation structure given by (40), where $\mu_{it} = \frac{\exp(x'_{it}\beta)}{1+\exp(x'_{it}\beta)}$ and ρ is a correlation index parameter. Next, suppose that the binary data contain two sided outliers. One may then follow (29) and compute all nine combinations for the product term $\psi_c(r_{iu})\psi_c(r_{it})$ and compute the expectations of all these nine terms, and derive the formulas as

$$E[\psi_c(r_{iu})\psi_c(r_{it})] = \rho^{t-u} \sigma_{iut} a_{iut} + [E(\psi_c(r_{iu}))E(\psi_c(r_{it}))], \tag{49}$$

where

$$\begin{aligned} a_{iut} = & \frac{P_1^2}{\sqrt{\sigma_{it}^{(c_1)} \sigma_{iut}^{(c_1)}}} + P_1 P_2 \left[\frac{1}{\sqrt{\sigma_{it}^{(c_1)} \sigma_{iut}^{(c_1)}}} + \frac{1}{\sqrt{\sigma_{it}^{(c_1)} \sigma_{iut}^{(c_1)}}} \right] \\ & + P_1 P_3 \left[\frac{1}{\sqrt{\sigma_{it}^{(c_2)} \sigma_{iut}^{(c_1)}}} + \frac{1}{\sqrt{\sigma_{it}^{(c_1)} \sigma_{iut}^{(c_2)}}} \right] + P_2 P_3 \left[\frac{1}{\sqrt{\sigma_{it}^{(c_2)} \sigma_{iut}^{(c_1)}}} + \frac{1}{\sqrt{\sigma_{it}^{(c_1)} \sigma_{iut}^{(c_2)}}} \right] \\ & + \frac{P_2^2}{\sqrt{\sigma_{it} \sigma_{iut}}} + \frac{P_3^2}{\sqrt{\sigma_{it}^{(c_2)} \sigma_{iut}^{(c_2)}}}, \end{aligned}$$

for $u < t$. We may then easily compute ω_{iut} by using (49) and (48).

Further note that for the one sided outlier case, the $E[\psi_c(r_{iu})\psi_c(r_{it})]$ can be obtained from (49) as follows. For the one sided down-weighting function $\psi_c(r_{it})$ given in (28), one may compute the expectation of $\psi_c(r_{iu})\psi_c(r_{it})$ from (49) by

changing the limits obtained by replacing p_{lb} with 0. Similarly, the product moment based on the down-weighting function $\psi_c(r_{it})$ given in (27), can be obtained from (49) by changing the limits obtained by replacing p_{sb} with 1.

Under the AR(1) binary correlation structure (40), the outlier based moment estimation formula for ρ derived from (49), is given by

$$\hat{\rho}_M = \frac{\frac{\sum_{i=1}^K \sum_{u=1}^{T-1} [\psi_c(r_{iu}) - E(\psi_c(r_{iu}))][\psi_c(r_{i,u+1}) - E(\psi_c(r_{i,u+1}))] w_{iu} w_{i,u+1}}{K(T-1)}}{\frac{\sum_{i=1}^K \sum_{u=1}^T [\psi_c(r_{iu}) - E(\psi_c(r_{iu}))]^2 / \text{var}[\psi_c(r_{iu})]}{KT}}{\frac{\sum_{i=1}^K \sum_{u=1}^{T-1} \sigma_{iuu} \alpha_{iut} w_{iu} w_{i,u+1}}{K(T-1)}}}. \tag{50}$$

Alternatively, for any lag 1 dependent [irrespective of the correlation structure such as AR(1) or MA(1)] binary or count data with possible outliers, the lag 1 correlation index parameter ρ may be estimated as

$$\hat{\rho}_M = \frac{\sum_{i=1}^K \sum_{u=1}^{T-1} [\psi_c(r_{iu}) w_{iu} - \bar{\xi}_{u,w}] [\psi_c(r_{i,u+1}) w_{i,u+1} - \bar{\xi}_{u+1,w}]}{K(T-1)}{\frac{\sum_{i=1}^K \sum_{u=1}^T [\psi_c(r_{iu}) w_{iu} - \bar{\xi}_{u,w}]^2}{KT}}, \tag{51}$$

where $\bar{\xi}_{t,w} = \frac{1}{K} \sum_{i=1}^K \psi_c(r_{it}) w_{it}$.

Computation of Ω_i for Count Data

Note that as opposed to the binary case, the construction of the Ω_i matrix is difficult for the count data case. One may, however, alternatively compute this Ω_i matrix by using the general formula

$$\text{cov}(\xi_i) = \Omega_i = A_{i\xi}^{\frac{1}{2}} C_{i\xi} A_{i\xi}^{\frac{1}{2}}, \tag{52}$$

where $A_{i\xi} = [\text{var}(\psi_c(r_{i1})), \dots, \text{var}(\psi_c(r_{it})), \dots, \text{var}(\psi_c(r_{iT}))]$ and $C_{i\xi} = (c_{i\xi,ut})$, with $c_{i\xi,ut} = \text{corr}[\psi_c(r_{iu}), \psi_c(r_{it})]$ for $u, t = 1, \dots, T$. For (52), the formulas for $\text{var}[\psi_c(r_{it})]$ for the binary data are given in Sect. 2.2.2, and for the count data they are available from [Cantoni and Ronchetti \(2001, Appendix\)](#). As far as the computation of the $C_{i\xi}$ matrix is concerned, one may approximate this matrix by a constant matrix $C_{i\xi}^*$, say, by pretending that the covariates are stationary even though they are non-stationary (i.e., time dependent). Under this assumption, the (u, t) th component of the constant matrix $C_{i\xi}^*$ may be computed as

$$C_{i\xi}^* = (c_{i\xi,ut}^*),$$

where

$$c_{\xi,ut}^* = \frac{\frac{1}{K} \sum_{i=1}^K [\psi_c(r_{iu}) - \bar{\xi}_u] [\psi_c(r_{it}) - \bar{\xi}_t]}{\sqrt{\frac{1}{K} \sum_{i=1}^K [\psi_c(r_{iu}) - \bar{\xi}_u]^2 \frac{1}{K} \sum_{i=1}^K [\psi_c(r_{it}) - \bar{\xi}_t]^2}}, \quad (53)$$

with $\bar{\xi}_t = \frac{1}{K} \sum_{i=1}^K \psi_c(r_{it})$, for all $t = 1, \dots, T$.

Note that the REGEE approach encounters convergence problems and also this approach produces regression estimates with much larger relative biases than the RGQL approach. See, for example, the finite sample relative performance of the RGQL and REGEE approaches shown through intensive simulation studies reported in [Bari and Sutradhar \(2010b\)](#).

Acknowledgments The author fondly acknowledges the stimulating discussion by the audience of the symposium and wishes to thank for their comments and suggestions.

References

- Bailey, B.: Tables of the Bonferroni t -test. *J. Am. Stat. Assoc.* **72**, 469–478 (1977)
- Bari, W., Sutradhar, B.C.: On bias reduction in robust inference for generalized linear models. *Scand. J. Stat.* **37**, 109–125 (2010a)
- Bari, W., Sutradhar, B.C.: Robust inferences in longitudinal models for binary and count panel data in the presence of outliers. *Sankhya B* **72**, 11–37 (2010b)
- Beckman, R.J., Cook, R.D.: Outlier...s. *Technometrics* **25**, 119–163 (1983)
- Cantoni, E.: A robust approach to longitudinal data analysis. *Can. J. Stat.* **32**, 169–180 (2004)
- Cantoni, E., Ronchetti, E.: Robust inference for generalized linear models. *J. Am. Stat. Assoc.* **96**, 1022–1030 (2001)
- Carroll, R.J., Pederson, S.: On robustness in the logistic regression model. *J. R. Stat. Soc. B* **55**, 693–706 (1993)
- Cook, R.D., Prescott, P.: On the accuracy of Bonferroni significance levels for detecting outliers in linear models. *Technometrics*, **23**, 59–63 (1981)
- Copas, J.B.: Binary regression models for contaminated data (with discussion). *J. R. Stat. Soc. B* **50**, 225–265 (1988)
- Doombos, R.: Testing for a single outlier in a linear model. *Biometrics* **37**, 705–711 (1981)
- Ellenberg, J.H.: The joint distribution of the standardized least squares residuals from a general linear regression. *J. Am. Stat. Assoc.* **68**, 941–943 (1973)
- Ellenberg, J.H.: Testing for a single outlier from a general linear model. *Biometrics* **32**, 637–645 (1976)
- Hampel, F.R., Rousseeuw, P.J., Ronchetti, E.M., Stahel, W.A.: *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York (1986)
- Huber, P.J.: *Robust Statistics*. Wiley, New York (2004)
- Heyde, C.C.: *Quasi-likelihood and its applications*. Springer-Verlag, New York (1997)
- Johnson, B.A., Prescott, P.: Critical values of a test to detect outliers in factorial experiments. *Appl. Stat.* **24**, 56–59 (1975)
- Liang, K.-Y., Zeger, S.L.: Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22 (1986)
- Lund, R.E.: Tables for an approximate test for outliers in linear regression. *Technometrics* **17**, 473–476 (1975)

- McCullagh, P., Nelder, J.A.: *Generalized Linear Models*, 2nd edn. Chapman and Hall, London (1989)
- McKenzie, E.: Some ARMA models for dependent sequences of Poisson counts. *Adv. Appl. Probab.* **20**, 822–835 (1988)
- Preisser, J.S., Qaqish, B.F.: Robust regression for clustered data with applications to binary regression. *Biometrics* **55**, 574–579 (1999)
- Prescott, P.: An approximate test for outliers in linear models. *Technometrics* **17**, 129–132 (1975)
- Qaqish, B.F.: A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* **90**, 455–463 (2003)
- Rousseeuw, P.J., Leroy, A.M.: *Robust Regression and Outlier Detection*. Wiley, New York (1987)
- Sinha, S.K.: Robust analysis of generalized linear mixed models. *J. Am. Stat. Assoc.* **99**, 451–460 (2004)
- Sinha, S.K.: Robust inference in generalized linear model for longitudinal data. *Can. J. Stat.* **34**, 1–18 (2006)
- Srikantan, K.S.: Testing for a single outlier in a regression model. *Sankhya A* **23**, 251–260 (1961)
- Stefansky, W.: Rejecting outliers by maximum normed residual. *Ann. Math. Stat.* **42**, 35–45 (1971)
- Stefansky, W.: Rejecting outliers in factorial designs. *Technometrics* **14**, 469–479 (1972)
- Street, J.O., Carroll, R.J., Ruppert, D.: A note on computing robust regression estimates via iteratively reweighted least squares. *Am. Stat.* **42**, 152–154 (1988)
- Sutradhar, B.C.: An overview on regression models for discrete longitudinal responses. *Stat. Sci.* **18**, 377–393 (2003)
- Sutradhar, B.C.: Inferences in generalized linear longitudinal mixed models. *Can. J. Stat.* **38**, 174–196 (2010)
- Sutradhar, B.C.: *Dynamic Mixed Models for Familial Longitudinal Data*. Springer, New York (2011)
- Sutradhar, B.C., Chu, D.P.T., Bari, W.: Estimation effects on powers of two simple test statistics in identifying an outlier in linear models. *J. Stat. Comput. Simul.* **77**, 305–328 (2007)
- Sutradhar, B.C., Das, K.: On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika* **86**, 459–465 (1999)
- Tietjen, G.L., Moore, R.H., Beckman, R.J.: Testing for a single outlier in simple linear regression. *Technometrics* **15**, 717–721 (1973)
- Wedderburn, R.W.M.: Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439–447 (1974)
- Zeger, S.L., Liang, K.Y., Self, S.G.: The analysis of binary longitudinal data with time independent covariates. *Biometrika* **72**, 31–38 (1985)