ROBUST ESTIMATION IN FAMILIAL AND LONGITUDINAL MODELS

WASIMUL BARI

Robust Estimation in Familial and Longitudinal Models

by

©Wasimul Bari

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirement for the Degree of Doctor of Philosophy in Statistics

Department of Mathematics and Statistics

Memorial University

August, 2007

St. John's **Newfoundland** **Canada**

Abstract

There exists many studies on the robust estimation of the regression effects in a linear model set up for continuous such as Gaussian data possibly containing one or more outliers. The robust estimation of the regression effects in a generalized linear model (GLM) set up for the count and binary data in the presence of outliers is, however, relatively difficult. In this thesis, we deal with this difficult estimation issue and develop the robust estimation procedures under three scenarios. First, a fully standardized Mallows-type quasi-likelihood (FSMQL) estimation technique is developed to obtain consistent regression estimates in the GLM set up for both independent count and binary data. Secondly, we develop a robust generalized quasi-likelihood (RGQL) estimation procedure to deal with the outliers in the generalized linear mixed model (GLMM) set up for both count and binary data. Finally, we also develop the RGQL estimation procedure to deal with possible outliers in the GLM set up for the longitudinal count and binary data. The performances of the proposed robust estimators are examined through extensive simulation studies under all three set up: the GLM for the independent count and binary data; the GLMM for the familial count and binary data; and the GLM for the longitudinal count and binary data.

Acknowledgments

My supervisor, Professor B. C. Sutradhar, deserves a lot more of appreciation than what I am able to express by these few standard words on this page. With his vast knowledge and deep insight in the subject, with tremendous zeal and patience, he guided and inspired me throughout the course of my Ph. D. program. I would like to express my deep heart-felt gratitude to my supervisor, Professor B. C. Sutradhar, for his supervision, advice, encouragement, enduring patience, and constant support. It was undoubtedly a great privilege for me to work under his supervision for which I shall forever be grateful.

I sincerely acknowledge the financial support provided by the School of Graduate Studies, Department of Mathematics and Statistics, and Professor B. C. Sutradhar in the forms of Graduate Fellowship and Teaching Assistantship. Further, I wish to thank the Department for providing us a friendly atmosphere and the necessary facilities to complete my program in particular.

I sincerely thank the external examiner, Professor Mary E. Thompson, University of Waterloo, for her comments and suggestions. I also thank Drs. Zhaozhi Fan and J. C. Loredo-Osti, the internal examiners, for their comments.

I am also grateful to my mother and brothers for their eternal love, emotional support, and encouragement during this program.

It is my great pleasure to thank my friends and well-wishers who directly or indirectly encouraged and helped me in the Ph. D. program and contributed to this dissertation.

Contents

List of Tables

Chapter 1

Introduction

1.1 Motivation of the Problem

There exists a vast literature on the discrete such as count and binary data analysis in the clustered regression set up. In this set up, the responses in a cluster become correlated, but the correlation structure depends on the nature of the clusters. For example, when the data are collected from the members of a large number of independent families/clusters, they form a familial correlation structure. But, when the data are collected repeatedly over a small period of time from a large number of independent individuals, they form a longitudinal correlation structure. This type of clustered data whether familial or longitudinal has been widely discussed in the literature over the last two decades. The main thirst of these studies is to understand the effects of the associated covariates on the response data. In some situations, it may also be important to understand the correlation structure of the responses. In the familial set up, the correlations are usually understood through the variance components of the random family effects, whereas in the longitudinal set up, the correlation structure is formed due to the stochastic time factors. Note however that most of these familial or longitudinal studies are done based on the assumption that the data do not contain any outliers or they are not subject to any non-responses or any other less likely disturbances. For convenience, we now review some of the existing leading works in the familial and longitudinal set up. First, we consider the familial case in the following subsection and then we discuss the longitudinal case in the section 1.1.2.

1.1.1 A brief review of the familial data analysis

In the generalized linear model (GLM) set up, the existing studies with familial data have mainly been done for the binary and count data. This type of data belongs to the well-known exponential family in the independence set up. Thus, a GLM with appropriate link function is used to deal with such data in the independence set up. See, for example, McCullagh and Nelder (1989) among others. Note that for count data, one traditionally uses the so-called 'log' link to obtain the linear model, whereas the so-called 'logit' link is used for the binary data to obtain such linear model. As opposed to the independence set up, the count or binary responses under a family/cluster become correlated. The correlations arise due to the fact that all members of the family share a common random family effect. To accommodate this common random effect, the GLM considered in the independence set up has been extended to the generalized linear mixed model (GLMM) set up by adding a random effect with the linear predictor of the GLM. Note that under the normality assumption for the random family effects present in the linear function under the GLMM, many authors have dealt with estimation of the regression effects as well as the variance component of the random family effects. We refer to two pioneer works with regard to such inferences, namely, the penalized quasi-likelihood (PQL) approach of Breslow and Clayton (1993) and the hierarchical likelihood (HL) approach due to Lee and Nelder (1996), both for the count and binary data. These approaches are developed following the so-called best linear unbiased prediction (BLUP) technique used in linear mixed model to estimate the regression effects and the variance component of random effects through the estimation of the random effects. To be specific, even though the random effects are unobservable, in the PQL and HL approaches, the random effects are pretended to be fixed effects and they are estimated along with the regression effects. These estimates of regression effects and random effects are then used to estimate the variance component. It is however known that they may produce biased and hence inconsistent estimates, specially for the variance component. See for example, Kuk (1995), Breslow and Lin (1995), Sutradhar and Qu (1998), Jiang (1998), among others.

As opposed to the BLUP analogue approaches, there also exists Monte Carlo (MC) based approaches to analyze the data in the GLMM. For example, see Markov chain Monte Carlo (MCMC) technique of Zeger and Karim (1991), Monte Carlo EM (MCEM) and Monte Carlo Newton-Raphson (MCNR) algorithms of McCulloch (1994, 1997). These approaches may however computationally burdensome [Jiang (1998)], specially when one deals with multi-dimensional random effects.

Recently, Jiang and Zhang (2001) have introduced an improvement over Jiang's (1998) simulated method of moments. But, as shown by Sutradhar (2004), this improved method of moments may also be inefficient, specially when in estimating the variance component of the model. Furthermore, Sutradhar (2004) has used a generalized quasi-likelihood (GQL) approach which produces both consistent and highly efficient estimates (as compared to other competitive moment estimates). For similar GQL inferences in the GLMM set up for both binary and count data, we refer to Sutradhar and Rao (2003).

Note however that all these approaches including those in Sutradhar and Rao (2003) and Sutradhar (2004) deal with the familial binary or count data, but in the absence of any possible outliers.

1.1.2 A brief review of the longitudinal data analysis

In the longitudinal set up, when the repeated data do not contain any outliers, there exists a vast literature mainly beginning from the pioneer work of Liang and Zeger (1986). As opposed to the familial set up, here repeated data are collected from a large number of independent individuals. The responses become correlated because of the time effects on the repeated responses of the same individual. In this set up, one is interested to estimate the associated regression effects consistently and efficiently, as well as the longitudinal correlations at least consistently.

Under the assumption that the longitudinal discrete responses have specified known forms for the marginal means and variances, Liang and Zeger (1986) proposed a generalized estimating equation (GEE) approach for the consistent and efficient estimation of the regression effects, which was developed by using a suitable 'working' stationary matrix for the underlying true unknown correlation structure. As discussed by Crowder (1995) and Sutradhar and Das (1999) [see also, Sutradhar (2003)], this GEE approach however has many pitfalls with regard to both consistency and efficiency. Sutradhar (2003) has relaxed the assumption about the correlation structure. To be specific, it has been assumed in Sutradhar (2003) that the repeated responses follow a stationary correlation structure which belongs to a class of autocorrelation models that accommodates any of the basic correlation structures such as autoregressive of order one [AR(1)], moving average of order one [MA(1)], and equi-correlation

(EQ). Thus, as opposed to Liang and Zeger (1986), the correlation structure is assumed to be known subject to the restriction that it belongs to the suggested class of autocorrelations. As far as the estimating equation is concerned, Sutradhar (2003) has suggested a generalization of the traditional mean and variance based quasi-likelihood (QL) approach [see Wedderburn (1974) and McCullagh (1983), for example] for the estimation of the associated regression effects. This approach by Sutradhar (2003) may, therefore, be referred to as the generalized QL (GQL) approach. If the responses really follow this class of correlations, it is then clear that this GQL approach would produce both consistent and highly efficient regression estimates as compared to the other competitive such as 'working' independence based regression estimates. Note however that none of these approaches: Liang and Zeger (1986), Sutradhar and Das (1999), and Sutradhar (2003) dealt with any possible outlying observations in the longitudinal data set.

Some authors such as Thall and Vail (1990) and Davis et al. (2000) have analyzed the longitudinal data based on a correlation structure generated through a common individual random effect among the repeated responses. As it is discussed in the last section, this approach takes care of the familial correlations instead of the longitudinal correlations. See also Jowaheer and Sutradhar (2002). Thus, the random effect based longitudinal correlation modelling approach will not be followed any further in the thesis. Note that the above random effect based longitudinal data analyses were developed for the cases where the longitudinal data do not contain any outliers.

1.1.3 Existing robust estimation in the independent set up

Note that the robust inference for the regression effects in a linear model set up, specially for the symmetric continuous data with possible outliers, has a long history.

The main objective of such inferences is to downweight the suspected outliers so that the parameters of the model may be estimated consistently. For the upto date discussion on this and other related topics, we, for example, refer to Huber (2004), Rousseeuw and Leroy (1987), Hampel et al. (1986), and the references therein.

As opposed to the robust inference for the linear and non-linear set up for the continuous data with possible outliers, there does not, however, appear adequate discussions on the robust inference for the discrete such as binary and count data. For some recent discussions on the robust inference topic for the discrete data, we refer to Cantoni and Ronchetti (2001) and the references therein. To be specific, Cantoni and Ronchetti (2001) [see also, Mallows (1975)] have proposed a Mallows-type QL (MQL) estimating equation to estimate the associated regression parameters involved in the Poisson and binomial regression models, which can be considered as an improvement over the traditional QL estimating equation in the presence of possible outliers. It is, however, demonstrated in Chapter 2 that this approach of Cantoni and Ronchetti (2001) may further be improved to obtain the estimates with smaller biases. This is done by using the proper variance and gradient functions in the estimating equations, whereas the MQL approach uses certain 'working' variance and gradient functions to construct the estimating equations. This additional bias correction, therefore, improves the consistency of the estimates of parameters involved in the model.

With regard to the robust inference for the binary data in the presence of possible outlying observations, Copas (1988) proposed a misclassification based outlier resistant estimation approach that yields approximate consistent estimates for the regression effects. Next, to obtain more consistent estimates of the parameters, Carroll and Pederson (1993) introduced an improvement over the Copas's (1988) approach based on the Mallows-type estimation technique. In the thesis, in defining the outliers for the binary data, we use a similar but different way than that of Copas (1988) and Carroll and Pederson (1993), which is discussed in Chapter 2.

1.1.4 Existing robust estimation in the familial set up

The brief discussion on the inferences for the familial mixed model given in the section 1.1.1 indicates that obtaining the consistent estimates for the parameters under such models is difficult. This estimation problem naturally gets more complex if the familial data contain any outliers. Nevertheless, there have been a few attempts in the literature to use the familial mixed model to analyze the so-called longitudinal data in the presence of possible outliers. For example, we refer to Mills et al. (2002) and Sinha (2006). Note that these studies appear to encounter both modelling and estimation problems. To be specific, as pointed out by Jowaheer and Sutradhar (2002), the familial models (i.e., GLMM) are not able to accommodate the longitudinal correlation structure of the repeated data. With regard to the estimation of the parameters, even if one applies their familial model based estimation approach to analyze the familial data with possible outlying observations, the weighted likelihood approach of Mills et al. (2002) appear to produce the biased estimates irrespective of the situations whether the data contain outliers or not. See, for example, Table 1, 2, and 3 in Mills et al. (2002). These biases are generally produced because of the use of certain numerical techniques to drive out the random effects. Sinha's (2006) approach appears to be similar to that of Cantoni and Ronchetti (2001) where the estimating equations are developed by using certain 'working' gradient functions and covariance matrix constructed ignoring the outliers in the data.

In an earlier study, Sinha (2004) has correctly used a familial mixed model to analyze the familial data with possible outliers. As far as the estimation of the parameters is concerned, Sinha (2004) used the following steps: (1) a likelihood equation is written for the outlier free case conditional on the random effects; (2) a robust version of this equation is written to downweight the possible outliers that may be present in the data; and (3) the random effects are driven out by taking average over the Monte-Carlo based conditional distributions. Note, however, that this estimation approach of Sinha (2004) is quite cumbersome even if the data do not contain any outliers, that is, when only steps (1) and (3) are used. See, for example, Jiang (1998) and Jiang and Zhang (2001), where outlier free cases are dealt with by using the method of moments as opposed to the likelihood method. The alternative moment approach certainly produces the consistent estimates for the parameters of the model. In view of these studies for non-outlier cases, it seems that appropriate modification to the moment approach for the outliers would have been useful to the practitioners for simplicity. Moreover, the moment approaches are easily extendable to the situations where one cannot write the score equations because of the difficulty in constructing the likelihood function.

Further note that as one may obtain the consistent as well as more efficient estimates (as compared to the moment approach) by using the GQL approach [Sutradhar (2004)] in the absence of outliers, it appears to be much more appealing to modify the outlier free GQL approach to accommodate the possible outliers. This will be done in Chapter 3. We emphasize here that similar to the moment approach, the GQL approach may also be extendable to the situations where the construction of the likelihood function is not possible.

1.1.5 Existing robust estimation in the longitudinal set up

In the longitudinal set up, if the repeated data contain one or more outlying observations, the traditional GEE or GQL approach produces inconsistent estimates for the regression effects as well as for the associated correlation parameters. Some authors such as Preisser and Qaqish (1999), Mills et al. (2002), Cantoni (2004), and Sinha (2006) have attempted to develop the estimation approach to obtain estimates with small or no biases. But, these procedures use either improper gradient functions or 'working' covariance matrix or both in constructing the estimating equations for parameters. Note that it is well known by now that the use of the 'working' covariance matrix may produce the inconsistent and/or inefficient estimates [Sutradhar and Das (1999)] even if the data do not contain any outliers. In view of this result, there is no reason to extend the 'working' covariance matrix based approaches to analyze the longitudinal data in the presence of possible outliers. Further note that the use of an improper or 'working' gradient function may also result to biased estimates.

To be specific, Preisser and Qaqish (1999, eq. 1, p. 575) have used an estimating equation where a gradient function for non-outlying situations is used. In constructing the estimating equations, these authors also have used a 'working' covariance matrix which bypasses both the true correlation structure as well as the presence of possible outliers. Cantoni (2004) has improved this approach by using a proper gradient function constructed for the data containing possible outliers. She has, however, used the same 'working' covariance matrix as in Preisser and Qaqish (1999) in constructing the estimating equation.

Mills et al. (2002) and Sinha (2006) also have analyzed the longitudinal data with possible outliers. These authors unlike Preisser and Qaqish (1999) and Cantoni (2004) have modelled the longitudinal correlations by using the random effects approach.

Note that as argued by Jowaheer and Sutradhar (2002), their random effects based approaches appear to be quite inappropriate to model the longitudinal correlations. Further note that even if their approaches are used for the analysis of the familial data in the presence of outliers, their estimating equations would still contain 'working' gradients functions [Sinha (2006)].

In the thesis, unlike the above existing studies, we assume that the longitudinal data in the absence of outliers follow a true non-stationary correlation structure which may accommodate the AR(1), MA(1), and EQ correlation structures. We then construct the estimating equations not only by modifying the gradient functions due to the outliers, but also by modifying the covariance matrix to reflect the outliers in the data. Thus, the new modified estimation approach would be a proper generalization of the Mallows-type GQL approach [see also Sutradhar (2003)]. This new development is discussed in Chapter 4 including some simulation studies.

1.2 Objectives of the Thesis

The discrete clustered data analysis in the presence of possible outliers encounters two types of problems. First, it is in general difficult to model the correlation structure of the data even if the data is outlier free. Secondly, it is quite difficult to model outliers for such correlated data. The main objective of the thesis is to develop a valid estimation procedure through proper modelling for both correlation structure and possible outliers. To achieve this goal, we undertake the following issues in sequence.

1. Since the robust inference for the discrete such as count and binary independent data is not adequately discussed in the literature, we first make an improvement

over a recent Mallows-type quasi-likelihood (MQL) approach discussed by Cantoni and Ronchetti (2001). Note that the proposed improvement occurs because of the use of a proper gradient function and a proper variance structure in the construction of the estimating equations. This improvement is discussed in details in Chapter 2.

2. In Chapter 3, we generalize the robust estimating equations developed for the independent case to the familial set up for both count and binary data. Note that in a familial set up, one deals with a large number of independent families/clusters each with one or more members. As the members of a family share a common random family effect, the count or binary responses for the members of a given family become correlated. This correlation structure along with certain modelling for one or more outliers is exploited to analyze the familial data in the presence of possible outliers.

3. In Chapter 4, we provide a similar robust estimation approach as in Chapter 3. But, we do this for the longitudinal discrete (count and binary) data that may contain one or more outliers. Note that in the longitudinal set up, the correlation structures of the clustered data are quite different than those under the familial set up discussed in Chapter 3. In the longitudinal studies, a small number of repeated responses are collected from a large number of independent individuals over a period of time. As the repeated responses are made on the same individual, it is likely that these responses are correlated. The specifications of the proper correlation structure and gradient function for the repeated data in the presence of possible outliers are considered in constructing the estimating equations to obtain the consistent and efficient estimates of the

parameters involved in the longitudinal model.

4. The thesis is concluded in Chapter 5. We provide some remarks on the possibilities of extending the robust approaches discussed in Chapter 3 and 4 to the combined familial-longitudinal set up in the presence of possible outliers.

Chapter 2

Improved Robust Estimation for Independent Count and Binary Data

In generalized linear model (GLM) set up [McCullagh and Nelder (1989)], when the data do not contain any outliers, it is customary to analyze discrete such as count and binary data collected from a sample of independent individuals. Let K be the size of the sample under study. Suppose that y_i is a discrete response collected from the ith $(i = 1, \ldots, K)$ individual of the sample. Also suppose that $x_i = (x_{i1}, \ldots, x_{iu}, \ldots, x_{ip})'$ be the p-dimensional covariate vector corresponding to y_i. Let $\beta = (\beta_1, \ldots, \beta_u, \ldots, \beta_p)'$ denote the effects of covariates x_i on the response y_i for all $i = 1, \ldots, K$. This type of data may be analyzed by using the maximum likelihood (ML) approach where the marginal density is assumed to belong to the well-known exponential family. That is,

$$f(y_i) = k_0 exp\left[\{y_i\theta_i - a(\theta_i)\}\,\phi + b(y_i, \phi)\right], \tag{2.1}$$

where $\theta_i = h(\eta_i)$ with $\eta_i = x_i'\beta$, $a(\cdot)$, $b(\cdot)$, $h(\cdot)$ are of known functional forms, ϕ is possibly a known scale parameter such as $\phi = 1$ for both count and binary data, and k_0 is a normalizing constant. Here it is of primary interest to estimate the regression effect β. It is well known that the ML approach produces consistent and highly efficient estimates for this regression effect β. Alternatively, one may use the simpler moment or quasi-likelihood (QL) approaches to obtain consistent estimators of β, but these estimators may be slightly less or equally efficient as compared to the ML estimators. For various studies using the ML approach we, for example, refer to Pregibon (1982); Stefanski et al.(1986); Künsch et al. (1989); Morgenthaler (1992); and Ruckstuhl and Welsh (1999). Similarly, one may refer to Wedderburn (1974); McCullagh and Nelder (1989); and Heyde (1997) for studies applying moment or QL approaches.

Note that in practice, it may however happen that the data may contain one or more outlying observations. For example, in count data analysis, it may happen that the bulk of the observations follow the Poisson distribution with means close to each other, whereas a few outlying count observations may arise from the Poisson distribution but with inflated or deflated means due to the contaminated covariates. This type of outliers is referred to as the contaminated covariates based outliers and the outlying responses may be generally understood based on the magnitudes of their differences from their corresponding means. There exists some studies, see for example, the recent studies by Cantoni and Ronchetti (2001) and Sinha (2004) with regard to the robust estimation of the regression effects β in the presence of such contaminated covariates based outliers.

In the binary case, similar robust inferences have been studied by some authors over the last two decades. For example, we refer to a pioneer work by Copas (1988) mainly for a misclassification based robust estimation procedure that produces approximate consistent estimates for the regression effects. To be specific, it has been assumed by Copas (1988) that the bulk of the binary observations follow a logistic distribution with a specified success probability, but the outlying binary observations are generated following a different logistic distribution characterized by a misclassified success probability different than that of the logistic distribution defined for the bulk of the observations. Later on, Carroll and Pederson (1993) have suggested an improvement over the estimation approach considered by Copas (1988), that produces better consistent estimates for the regression effects. The estimation procedure suggested by Carroll and Pederson (1993) belongs to the so-called Mallows class.

To reflect possible contamination in the covariate which may cause y_i (count or binary) to be an outlier, we now denote the observed covariate vector as

$$\tilde{x}_i = \begin{cases} x_i & \text{when } ith \text{ respnse is not an outlier} \\ x_i + \delta & \text{when } ith \text{ response is an outlier} \end{cases},$$

where δ is a vector indicating possible amount of contamination in the covariate.

Note that in the count data set up, the study by Cantoni and Ronchetti (2001) also deals with Mallows-type robust estimation. More specifically, for the robust estimation of the regression parameters involved in Poisson and binomial regression models, Cantoni and Ronchetti (2001) have used a Mallows-type QL (MQL) estimating equation (MQLEE) given by

$$\sum_{i=1}^{K} \left[w(\tilde{x}_i) \frac{\partial \tilde{\mu}_i}{\partial \beta} \tilde{V}^{-\frac{1}{2}}(\tilde{\mu}_i) \psi_c(r_i) - a(\beta) \right] = 0, \tag{2.2}$$

where $a(\beta) = \frac{1}{K} \sum_{i=1}^{K} w(\tilde{x}_i) \frac{\partial \tilde{\mu}_i}{\partial \beta} \tilde{V}^{-\frac{1}{2}}(\tilde{\mu}_i) E[\psi_c(r_i)]$, with $\tilde{\mu}_i = E(Y_i|\tilde{x}_i)$, $\tilde{V}(\tilde{\mu}_i) = var(Y_i|\tilde{x}_i)$, $w(\tilde{x}_i) = \sqrt{(1-h_i)}$, where h_i is the ith diagonal element of the hat matrix $H =$

$\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$, with $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_i, \ldots, \tilde{x}_K)'$ being an $K \times p$ covariate matrix, and $\psi_c(r_i)$ is the so-called Huber function defined by

$$\psi_c(r_i) = \begin{cases} r_i, & |r_i| \leq c, \\ \\ c \text{ sign}(r_i), & |r_i| > c, \end{cases} \tag{2.3}$$

where $r_i = \frac{y_i - \tilde{\mu}_i}{\sqrt{\tilde{V}(\tilde{\mu}_i)}}$ and c is referred to as the tuning constant. Note that the MQL estimation by (2.2) provides an improvement over the traditional ML or QL estimation of the parameters in the presence of outliers. This MQL estimating equation is still not unbiased. See, for example, section 2.2.1 for the amount of biases under specific models.

In this chapter, we however demonstrate that the MQL estimating equation (2.2) of Cantoni and Ronchetti (2001) produces estimates with large biases, thus, their estimates may not be consistent. We further provide an improvement over this MQL procedure. As far as the definition of outliers for the count data is concerned, we use the same definition as those of Cantoni and Ronchetti (2001) and Sinha (2004). For the outliers in the binary data, our definition of outliers is similar but different than that of Copas (1988) and Carroll and Pederson (1993). To be specific, similar to Copas (1988), we also assume that the outlying binary observations follow a different logistic distribution than that for the bulk of the observations. As far as the success probabilities for these two different logistic distributions are concerned, Copas (1988) has modelled the success probability for the binary outliers as a function of an additional contaminated parameter, whereas in our definition, the success probability of the outliers gets inflated or deflated due to the contaminated covariates [Copas (1988, p. 226)].

Note that the MQLEE (2.2) is constructed to avoid the outlier effects when outliers are assumed to arise due to the contamination of the covariates. If the data do not contain any outliers, then one would use $w(\tilde{x}_i) = 1$ and $\psi_c(r_i) = r_i$, and the estimating equation (2.2) will appear the same as the well-known QL estimating equation (QLEE)

$$\sum_{i=1}^{K} \left[\frac{\partial \mu_i}{\partial \beta} V^{-1}(\mu_i)(y_i - \mu_i) \right] = 0, \tag{2.4}$$

[Wedderburn (1974) and McCullagh and Nelder (1989)] which produces consistent estimator of β. In (2.4), $\mu_i = E(Y_i|x_i) = a'(x_i'\beta)$ and $V(\mu_i) = var(Y_i|x_i) = a''(x_i'\beta)$, where $a(x_i'\beta)$ is a known function based on the identity link function h, defined as in (2.1). Here, $a'(\cdot)$ and $a''(\cdot)$ are the first and second derivatives. This QLEE based estimators however become inconsistent in the presence of outliers. To have a feel for this inconsistency of the regression estimators, we report a simulation study in section 2.3.

In spite of the deductibility property of the MQLEE (2.2) of reducing to the QLEE (2.4), the MQLEE (2.2) may, however, still produce biased estimate for β, making the MQL estimator inconsistent. This is because, in minimizing the robust distance function $\psi_c(r_i)$, the MQLEE uses variance $\tilde{V}(\tilde{\mu}_i) = var(Y_i|\tilde{x}_i)$ as a weight function and $\frac{\partial \tilde{\mu}_i}{\partial \beta}$ 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. Note that as the MQLEE (2.2) does not use the true variance and gradient functions, we refer to this MQLEE based estimation approach as a 'working' MQL (WMQL) approach, which may yield biased estimate for β because of using the β dependent 'working' functions $var(Y_i|\tilde{x}_i)$ and $\frac{\partial \tilde{\mu}_i}{\partial \beta}$, instead of the true functions.

In a longitudinal set up, recently Cantoni (2004) has constructed an estimating equation by incorporating the proper gradient functions, namely, $\frac{\partial \psi_c(r_i)}{\partial \beta}$, but as far as the weights are concerned , she used a 'working' covariance matrix of the responses.

We refer to this estimation approach as the semi-standardized 'working' MQL (SS-WMQL) approach, as it accommodates the proper gradient functions but still uses the 'working' variance as the weight function. Furthermore, Sinha (2004) [see also He et al. (2005) for extended semiparametric set up] has considered similar robust estimation problems in the clustered regression set up, but these approaches of Sinha (2004), Cantoni (2004), and He et al. (2005) do not yield the true variance and gradient functions based proper estimating equations (as in the present Chapter) in the independence set up.

To be specific, we consider a fully standardized MQL (FSMQL) estimation approach where the estimating equation is constructed by using the proper robust weight and gradient functions. This construction is given in section 2.1 both for the count and binary data subject to one or more outliers. In section 2.2, we demonstrate that the estimating functions used by Cantoni and Ronchetti (2001) to construct the WMQL estimating equations are not unbiased for zero. The biasness of the estimating functions for the FSMQL approach is also studied in the same section. Furthermore, this section contains the asymptotic properties of the WMQL and FSMQL estimators. In section 2.3, we conduct an extensive simulation study to examine the relative consistency performance of the proposed FSMQL estimators as compared to those of the WMQL and SSMQL approaches. We conclude this chapter in section 2.4 with some remarks.

2.1 An Improved Robust Quasi-Likelihood Estimation

Note that the WMQL (2.2) estimating equation may still produce highly biased estimate for the regression parameter β, leading this estimate to be inconsistent. As a remedy, we now propose to solve a fully standardized MQL (FSMQL) estimating equation to obtain the consistent estimate for the β parameter. This we do by replacing the 'working' variance and gradient functions $\tilde{V}(\tilde{\mu}_i)$ and $\frac{\partial \tilde{\mu}_i}{\partial \beta}$ in the estimating equation (2.2), with the true variance and gradient functions $var(\psi_c(r_i))$ and $\frac{\partial \psi_c(r_i)}{\partial \beta}$, respectively. For example, by using this replacement operation in the WMQLEE (2.2), one obtains a proper standardized estimating equation given by

$$\sum_{i=1}^{K}\left[w(\tilde{x}_i)\frac{\partial}{\partial \beta}\left\{\psi_c(r_i) - \frac{1}{K}\sum_{i=1}^{K} E\left(\psi_c(r_i)\right)\right\}\{var\left(\psi_c(r_i)\right)\}^{-1} \times\left\{\psi_c(r_i) - \frac{1}{K}\sum_{i=1}^{K} E\left(\psi_c(r_i)\right)\right\}\right] = 0, \tag{2.5}$$

which we refer to as the FSMQL$_1$ estimating equation (FSMQL$_1$EE) for β. Note that the formulas for the true weight function $var(\psi_c(r_i))$ and the gradient function $\frac{\partial \psi_c(r_i)}{\partial \beta}$ may be obtained based on the underlying model for the responses subject to one or more outliers.

We also propose to use a slightly different version of the estimating equation (2.5) by using the deviance function

$$\psi_c(r_i) - E(\psi_c(r_i))$$

instead of

$$\psi_c(r_i) - \frac{1}{K}\sum_{i=1}^{K} E(\psi_c(r_i)).$$

We then write a FSMQL_2EE given by

$$\sum_{i=1}^{K}\left[w(\tilde{x}_i)\frac{\partial}{\partial\beta}\left\{\psi_c(r_i)-E\left(\psi_c(r_i)\right)\right\}\left\{var\left(\psi_c(r_i)\right)\right\}^{-1}\left\{\psi_c(r_i)-E\left(\psi_c(r_i)\right)\right\}\right]=0, \tag{2.6}$$

and its solutions will be referred to as the FSMQL_2 estimates. Next, we also consider a semi-standardized estimating equation derived from (2.6) by using the proper gradient function only. Note however that in this version we use an expected gradient function instead of data based gradient function for the stability reason. The estimating equation has the form given by

$$\sum_{i=1}^{K}\left[w^2(\tilde{x}_i)E\left\{\frac{\partial}{\partial\beta}\left(\psi_c(r_i)-E\left(\psi_c(r_i)\right)\right)\right\}\left\{\psi_c(r_i)-E\left(\psi_c(r_i)\right)\right\}\right]=0, \tag{2.7}$$

which we refer to as the semi-standardized MQL (SSMQL) estimating equation. Note that this SSMQL estimating equation has been developed under the independence set up, whereas the SSWMQL approach of Cantoni (2004) was suggested for the longitudinal data with possible outliers.

Since the exponential family model (2.1) contains the count and binary data as practically important cases, in the next subsections, we introduce the outlier models suitable for these count and binary data subject to one or more outliers. We also show how to construct the robust functions $\psi_c(r_i)$ and their corresponding expectations $E(\psi_c(r_i))$, variances $var(\psi_c(r_i))$, and the gradient functions $\frac{\partial\psi_c(r_i)}{\partial\beta}$ under these two models.

2.1.1 Robust weight and gradient function for count data

In the count data set up, an outlying count may arise either due to a shift in the mean or due to an inflated variance for an individual. For example, suppose that in the absence of outliers, all K 'good' responses are assumed to be generated from a

Poisson distribution with mean $\mu_i = exp(x_i'\beta)$, $i = 1, \dots, K$. Also suppose that the Kth response became an outlier due to the contaminated covariate $\tilde{x}_K = x_K + \delta$, δ being a suitable real valued vector. The response means in such a case may be expressed as

$$\tilde{\mu}_i = \begin{cases} exp(x_i'\beta) & \text{for } i = 1, \dots, K-1 \\ \\ exp((x_i + \delta)'\beta) & \text{for } i = K \end{cases},$$

where x_i's are uncontaminated covariates. Under this model, y_K is referred to as a mean shifted outlier.

Alternatively, suppose that $K - 1$ counts follow a Poisson distribution with mean $\mu_i = exp(x_i'\beta)$ for $i = 1, \dots, K-1$ and the Kth response arises from a negative binomial distribution with the mean $\mu_K = exp(x_K'\beta)$ but with a different variance such as $var(Y_K|x_K) = \mu_K + \alpha\mu_K^2$ for a suitable scalar $\alpha > 0$. We refer this observation y_K to as a variance inflated outlier. In this case, it would be appropriate to define the robust function $\psi_c(r_i)$ with $r_i = \frac{y_i - \mu_i}{\sqrt{V(\mu_i, \alpha)}}$, where $\alpha > 0$ indicates an overdispersion for the outlying observation and for given α, variance is now a function of μ_i as well as α.

In the thesis, we however consider the mean shifted outlier models only, which received considerable attention in the literature [see for example, Cantoni and Ronchetti (2001), Sinha (2004)]. For this case, the robust function $\psi_c(r_i)$ may be expressed as in (2.3), where $r_i = \frac{y_i - \tilde{\mu}_i}{\sqrt{\tilde{V}(\tilde{\mu}_i)}}$. In order to construct the proposed FSMQL$_1$EE (2.5) and FSMQL$_2$EE (2.6), we now provide the formulas for the expectation, variance, and gradient function of $\psi_c(r_i)$ as in the following theorems.

Theorem 1: Let $\psi_c(r_i)$, with $r_i = \frac{y_i - \tilde{\mu}_i}{\sqrt{\tilde{V}(\tilde{\mu}_i)}}$, denote the robust function defined as in (2.3). Also let $i_1 = Int\left(\tilde{\mu}_i - c\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)\right)$ and $i_2 = Int\left(\tilde{\mu}_i + c\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)\right)$ be the nearest integer values of $\tilde{\mu}_i - c\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)$ and $\tilde{\mu}_i + c\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)$, respectively, with $\tilde{\mu}_i = exp(\tilde{x}_i'\beta)$ and $\tilde{V}(\tilde{\mu}_i) = \tilde{\mu}_i$. For the Poisson data, the expectation of the robust function $\psi_c(r_i)$ is given by

$$E(\psi_c(r_i)) = c\left[1 - F_{Y_i}(i_2) - F_{Y_i}(i_1)\right] + \frac{\tilde{\mu}_i}{\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)}\left[P(Y_i = i_1) - P(Y_i = i_2)\right], \qquad (2.8)$$

where $P(Y_i = i_1)$ and $F_{Y_i}(i_1)$, for example, are the probability and cumulative probability density functions of Y_i with

$$P(Y_i = i_1) = \frac{exp(-\tilde{\mu}_i)\tilde{\mu}_i^{i_1}}{i_1!};$$

and the variance of the same robust function $\psi_c(r_i)$ is given by

$$var(\psi_c(r_i)) = E(\psi_c^2(r_i)) - \left[E(\psi_c(r_i))\right]^2, \qquad (2.9)$$

where

$$\begin{aligned} E(\psi_c^2(r_i)) &= c^2[1 - F_{Y_i}(i_2) + F_{Y_i}(i_1)] + \frac{1}{\tilde{V}(\tilde{\mu}_i)}\Big[\tilde{\mu}_i^2(F_{Y_i}(i_2 - 2) - F_{Y_i}(i_1 - 2)) \\ &+ (\tilde{\mu}_i - 2\tilde{\mu}_i^2)(F_{Y_i}(i_2 - 1) - F_{Y_i}(i_1 - 1)) + \tilde{\mu}_i^2(F_{Y_i}(i_2) - F_{Y_i}(i_1))\Big]. \end{aligned}$$

Proof: The proof is available in Cantoni and Ronchetti (2001), Appendix A, p. 1028.

Theorem 2: For the Poisson data, the gradient of the robust function and its expectation are given by

$$\frac{\partial\psi_c(r_i)}{\partial\beta} = \begin{cases} -\frac{\tilde{\mu}_i}{\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)}\tilde{x}_i, & |r_i| \le c, \\ 0, & |r_i| > c, \end{cases} \qquad (2.10)$$

and

$$\begin{aligned} \frac{\partial E(\psi_c(r_i))}{\partial\beta} &= -c\left[\frac{\partial}{\partial\beta}F_{Y_i}(i_2) + \frac{\partial}{\partial\beta}F_{Y_i}(i_1)\right] + \frac{\tilde{\mu}_i}{\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)}\left[\left\{\tilde{x}_i P(Y_i = i_1) + \frac{\partial}{\partial\beta}P(Y_i = i_1)\right\}\right. \\ &\left.- \left\{\tilde{x}_i P(Y_i = i_2) + \frac{\partial}{\partial\beta}P(Y_i = i_2)\right\}\right], \qquad (2.11) \end{aligned}$$

where

$$\frac{\partial}{\partial\beta}P(Y_i=i_1)=P(Y_i=i_1)(i_1-\tilde{\mu}_i)\tilde{x}_i,\ \frac{\partial}{\partial\beta}P(Y_i=i_2)=P(Y_i=i_2)(i_2-\tilde{\mu}_i)\tilde{x}_i,$$

$$\frac{\partial}{\partial\beta}F_{Y_i}(i_1)=\sum_{j=0}^{i_1}\frac{\partial}{\partial\beta}P(Y_i=j),\ \text{and}\ \frac{\partial}{\partial\beta}F_{Y_i}(i_2)=\sum_{j=0}^{i_2}\frac{\partial}{\partial\beta}P(Y_i=j).$$

Proof: The proof is obvious from (2.3) and (2.8).

2.1.2 Robust weight and gradient function for binary data

Note that unlike the count data case, the construction of Mallows-type robust function for the binary data is complicated. In fact, Cantoni and Ronchetti (2001) considered the binomial model that can be analyzed in the manner similar to that of Poisson case. For the binary case, Sinha (2004) used the same robust function as that of Poisson data, where $y_j = 1$ is referred to as an outlier if the corresponding covariate values lead to a small probability such as $P(Y_j = 1|\tilde{x}_j) < 0.4$. Similarly, $y_j = 0$ is referred to as an outlier if the corresponding probability $P(Y_j = 1|\tilde{x}_j)$ is large, say $P(Y_j = 1|\tilde{x}_j) > 0.6$. This definition of an outlier for the binary data does not however appear to interpret the real nature of an outlier. This is because generally an observation is treated to be an outlier when it is quite different from the bulk of the observations in the sample. To be specific, suppose that in a sample of size of K, the covariate values of $K-1$ individuals lead to small probabilities such as $P(Y_i = 1|\tilde{x}_i) \leq 0.3$ for all $i \neq j$, $i = 1, \ldots, K$. In this case, $y_j = 0$ or 1 will be an outlier if $P(Y_j = 1|\tilde{x}_j)$ is large. But an observation $y_i = 1$ for $i \neq j$, $i = 1, \ldots, K$ with small probability should not be referred to as an outlier. This fact leads us to define the robust functions for the binary data under two following categories.

One sided outlier

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

$$\psi_c(r_i) = \begin{cases} \frac{y_i - \tilde{\mu}_i}{\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)}, & P(Y_i = 1|\tilde{x}_i) \leq p_{sb}, i \neq j, i = 1, \ldots, K, \\ \frac{y_i - \mu_i^{(c_1)}}{V^{(c_1)\frac{1}{2}}(\mu_i^{(c_1)})}, & P(Y_i = 1|\tilde{x}_i) > p_{sb}, i = j, \end{cases} \tag{2.12}$$

where $\tilde{\mu}_i = \frac{exp(\tilde{x}_i'\beta)}{1+exp(\tilde{x}_i'\beta)}$, $\tilde{V}(\tilde{\mu}_i) = \tilde{\mu}_i(1 - \tilde{\mu}_i)$ for all $i = 1, \ldots, K$, and $p_{sb} = max\{\tilde{\mu}_i\}$, $i \neq j$, is a bound for all $K - 1$ small probabilities. Note that in (2.12), y_j, whether 1 or 0, is treated as an outlier, whereas $K - 1$ responses denoted by y_i for $i \neq j$ constitute a group of 'good' observations. Further note that as p_{sb} depends on β, one may choose $p_{sb} = 0.4$ initially provided that the data contain more zero's than one's. To reflect this initial situation of $p_{sb} = 0.4$, we start with a suitable initial value of β so that $\tilde{\mu}_i$'s are small. Once a first step estimate of β is obtained, we then compute p_{sb} by using the given formula $p_{sb} = max\{\tilde{\mu}_i\}$, $i \neq j$. Next, for the suspected outlying observation, we replace its probability, i.e., $P(Y_j = 1|\tilde{x}_j) = \tilde{\mu}_j$, with an appropriate tuning constant related probability, say $\mu_j^{(c_1)}$ and by the same token $var(Y_j|\tilde{x}_j) = \tilde{\mu}_j(1 - \tilde{\mu}_j)$ should be replaced by $V^{(c_1)}(\mu_j^{(c_1)}) = \mu_j^{(c_1)}(1 - \mu_j^{(c_1)})$. Here, by using the similar argument as in the count data case, one may choose $\mu_j^{(c_1)}$ so that it is nearer to p_{sb} such as 0.5, 0.6, but less than $\tilde{\mu}_j$ (the outlying probability).

Note that as opposed to the case given in (2.12), if the bulk of the binary observations (i.e. 'good' observations) occur with large probabilities, then the robust function $\psi_c(r_i)$ $(i = 1, \ldots, K)$ is defined as

$$\psi_c(r_i) = \begin{cases} \frac{y_i - \tilde{\mu}_i}{\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)}, & P(Y_i = 1|\tilde{x}_i) \geq p_{lb}, i \neq j, i = 1, \ldots, K, \\ \frac{y_i - \mu_i^{(c_2)}}{V^{(c_2)\frac{1}{2}}(\mu_i^{(c_2)})}, & P(Y_i = 1|\tilde{x}_i) < p_{lb}, i = j, \end{cases} \tag{2.13}$$

where $p_{lb} = min\{\tilde{\mu}_i\}$, $i \neq j$, is a bound for all $K - 1$ large probabilities. To reveal the scenario that the data contain more one's than zero's, one may choose $p_{lb} = 0.6$ and start with a suitable value of β so that $\tilde{\mu}_i$'s are large. After getting the first step estimate for β, we then turn back to the formula $p_{lb} = min\{\tilde{\mu}_i\}$, $i \neq j$ to compute p_{lb}. In this case, to compute $\psi_c(r_j)$, the mean and the variance of the suspected outlying observation y_j are computed as $\mu_j^{(c_2)}$ and $V^{(c_2)}(\mu_j^{(c_2)}) = \mu_j^{(c_2)}(1 - \mu_j^{(c_2)})$, respectively, where $\mu_j^{(c_2)}$ is an appropriate tuning constant related probability. Note that to select a value for $\mu_j^{(c_2)}$, one may consider a value nearer to p_{lb} such that 0.4, 0.5, but greater than $\tilde{\mu}_j$ (the outlying probability).

Two sided outlier

It may happen in practice that probabilities for the bulk of the observations lie in the range $p_{lb} \leq P(Y_i = 1|\tilde{x}_i) \leq p_{sb}$, 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 $P(Y_j = 1|\tilde{x}_j) > p_{sb}$ or $P(Y_j = 1|\tilde{x}_j) < p_{lb}$. In this case, the robust function $\psi_c(r_i)$ $(i = 1, \ldots, K)$ may be defined as

$$\psi_c(r_i) = \begin{cases} \frac{y_i - \mu_i^{(c_1)}}{V^{(c_1)\frac{1}{2}}(\mu_i^{(c_1)})}, & P(Y_i = 1|\tilde{x}_i) > p_{sb}, i = j, \\ \frac{y_i - \tilde{\mu}_i}{\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)}, & p_{lb} \leq P(Y_i = 1|\tilde{x}_i) \leq p_{sb}, i \neq j, i = 1, \ldots, K, \\ \frac{y_i - \mu_i^{(c_2)}}{V^{(c_2)\frac{1}{2}}(\mu_i^{(c_2)})}, & P(Y_i = 1|\tilde{x}_i) < p_{lb}, i = j, \end{cases} \tag{2.14}$$

where $\mu_j^{(c_1)}$ and $V^{(c_1)}(\mu_j^{(c_1)})$ are defined as in (2.12), whereas $\mu_j^{(c_2)}$ and $V^{(c_2)}(\mu_j^{(c_2)})$ are defined as in (2.13). In general, p_{lb} and p_{sb} are considered to be 0.4 and 0.6, respectively [see Sinha (2004), for example]. The formulas of the expectation and variance of $\psi_c(r_i)$ for the binary data with a two sided outlier are given in Theorem 3, whereas the formulas for the gradient of $\psi_c(r_i)$ and its expectation in the two sided outlier case are given in Theorem 4. Note that the formulas for the expectation, variance and gradient of the robust function $\psi_c(r_i)$ in the one sided outlier case may be obtained as special cases following Theorems 3 and 4. Thus, they are not given in the form of any theorems.

Theorem 3: Let $\psi_c(r_i)$ denote the robust function defined as in (2.14). The expectation and variance of $\psi_c(r_i)$ are given by

$$E(\psi_c(r_i)) = \frac{\tilde{\mu}_i - \mu_i^{(c_1)}}{V^{(c_1)\frac{1}{2}}(\mu_i^{(c_1)})}P_1 + \frac{\tilde{\mu}_i - \mu_i^{(c_2)}}{V^{(c_2)\frac{1}{2}}(\mu_i^{(c_2)})}P_3, \tag{2.15}$$

and

$$var(\psi_c(r_i)) = \frac{(1-2\mu_i^{(c_1)})\tilde{\mu}_i + \mu_i^{(c_1)^2}}{V^{(c_1)}(\mu_i^{(c_1)})}P_1 + P_2 + \frac{(1-2\mu_i^{(c_2)})\tilde{\mu}_i + \mu_i^{(c_2)^2}}{V^{(c_2)}(\mu_i^{(c_2)})}P_3 - [E(\psi_c(r_i))]^2, \tag{2.16}$$

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

Proof: The proof follows by taking expectation of the $\psi_c(r_i)$ function given in (2.14). Remark that 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,

$$P_1 = \frac{\text{Number of observations satisfying } P(Y_i = 1|\tilde{x}_i) > p_{sb}}{\text{Total observation } (K)}.$$

Theorem 4: The gradient of the robust function $\psi_c(r_i)$ [defined in (2.14)] and its expectation are given by

$$\frac{\partial \psi_c(r_i)}{\partial \beta} = \begin{cases} 0, & P(Y_i = 1|\tilde{x}_i) > p_{sb}, i = j, \\ \frac{-\tilde{\mu}_i(1-\tilde{\mu}_i)\tilde{x}_i}{\tilde{V}^{\frac{1}{2}}(\tilde{\mu}_i)}, & p_{lb} \le P(Y_i = 1|\tilde{x}_i) \le p_{sb}, i \ne j, i = 1, \ldots, K, \\ 0, & P(Y_i = 1|\tilde{x}_i) < p_{lb}, i = j, \end{cases} \tag{2.17}$$

and

$$\frac{\partial E(\psi_c(r_i))}{\partial \beta} = \frac{(1-\tilde{\mu}_i)\tilde{\mu}_i\tilde{x}_i}{V^{(c_1)\frac{1}{2}}(\mu_i^{(c_1)})} P_1 + \frac{(1-\tilde{\mu}_i)\tilde{\mu}_i\tilde{x}_i}{V^{(c_2)\frac{1}{2}}(\mu_i^{(c_2)})} P_3 \tag{2.18}$$

Proof: The proof is immediate by taking the derivatives of the functions given in (2.14) and (2.15).

As mentioned earlier, the expectation, variance, gradient of the robust function $\psi_c(r_i)$ and its expectation for the one sided outlier case can easily be obtained from Theorems 3 and 4. For example, for $\psi_c(r_i)$ given in (2.12), one may compute its expectation from (2.15) by changing the limits obtained by replacing p_{lb} with 0. To be specific,

$$E(\psi_c(r_i)) = \frac{\tilde{\mu}_i - \mu_i^{(c_1)}}{V^{(c_1)\frac{1}{2}}(\mu_i^{(c_1)})} P_1.$$

Similarly, $var(\psi_c(r_i))$, $\frac{\partial \psi_c(r_i)}{\partial \beta}$, and $\frac{\partial E(\psi_c(r_i))}{\partial \beta}$ can be obtained from (2.16), (2.17), and (2.18), respectively, by modifying the limits obtained by replacing p_{lb} with 0. In the same fashion, for the binary outlying observations arising from the case given in (2.13), the formulas for $E(\psi_c(r_i))$, $var(\psi_c(r_i))$, $\frac{\partial \psi_c(r_i)}{\partial \beta}$, and $\frac{\partial E(\psi_c(r_i))}{\partial \beta}$ can be derived from (2.15), (2.16), (2.17), and (2.18), respectively, by replacing p_{sb} with 1 in the limits.

2.2 Asymptotics

2.2.1 Remarks on consistency

As pointed out earlier, when the data contain outliers, the traditional QL (unadjusted for outlier) approach may produce highly biased estimates for the regression effects. The Mallows-type QL approach constructed by downweighting the outlying observations reduces the biases of the regression estimates as compared to the traditional QL approach. As a further improvement, Cantoni and Ronchetti (2001) suggested a Fisher consistency type adjustment that leads to the WMQL estimating equation given in (2.2). Note however that this WMQLEE (2.2) still may produce significantly biased estimates. Thus, these WMQL estimators may not be consistent.

To examine the consistency of the WMQL estimators, we conduct a simulation study in section 2.3. The simulation results in section 2.3.1 for the Poisson case and in section 2.3.2 for the binary case appear to exhibit the inconsistency quite clearly. We also examine this inconsistency issue in the present section by carrying out some exact asymptotic computations for certain special cases. To be specific, we consider a special Poisson regression model with one outlier and compute the expectation of the WMQL estimating function $[E(WMQLEF)]$ from (2.2). The consistent WMQL estimators would be guaranteed only when $E(WMQLEF) = 0$. The consistency performance of the proposed FSMQL estimators may be examined in the manner similar to that of the WMQL estimators. The FSMQL approach appears to produce almost unbiased estimators. To avoid any duplications, in this section, we consider the FSMQL_1 estimators only which is obtained from (2.5) and examine its consistency performance. The consistency of the FSMQL_2 estimators obtained from (2.6) may be examined similarly, which is however not shown here for convenience.

Computation of $E(WMQLEF)$ for a special outlying Poisson model

Recall that the WMQL estimating equation used by Cantoni and Ronchetti (2001) has the form

$$\frac{1}{K}\sum_{i=1}^{K}[q(y_i) - a(\beta)] = 0, \tag{2.19}$$

where $q(y_i) = g(\tilde{x}_i)\psi_c(r_i)$ and $a(\beta) = \frac{1}{K}\sum_{i=1}^{K} g(\tilde{x}_i)E(\psi_c(r_i))$ with $g(\tilde{x}_i) = w(\tilde{x}_i)\frac{\partial\tilde{\mu}_i}{\partial\beta}\tilde{V}^{-\frac{1}{2}}(\tilde{\mu}_i)$, and $E(\psi_c(r_i))$ as given in (2.8) for the count data. The left-hand side of the WMQL estimating equation (2.19) is referred to as the WMQL estimating function (WMQLEF). That is,

$$WMQLEF \equiv \frac{1}{K}\sum_{i=1}^{K}[q(y_i) - a(\beta)].$$

To compute its expectation, suppose that the data contain one outlier. For the purpose, we first generate y_i from the Poisson distribution with parameter $\mu_i = exp(x_i\beta)$ with $x_i = 1.0$ for all $i = 1, \ldots, K$. Now, suppose that $\tilde{x}_i = x_i$ was observed for $i = 1, \ldots, K-1$, but the Kth observed covariate was contaminated as $\tilde{x}_K = x_K + \delta$ with a positive δ. In this case, the 'good' response y_K became an outlier.

Note that since y_K is now an outlier, in practice, y_K is downweighted for the consistent estimation of the parameters of the model. Consequently, the $\frac{1}{K}\sum_{i=1}^{K} q(y_i)$ function in (2.19) is written as

$$\frac{1}{K}\sum_{i=1}^{K} q(y_i) = \frac{1}{K}\left[\sum_{i=1}^{K-1} g(\tilde{x}_i)\frac{y_i - \tilde{\mu}_i}{\sqrt{\tilde{V}(\tilde{\mu}_i)}} - g(\tilde{x}_K)c\right],$$

and $a(\beta)$ in (2.19) is a constant which is calculated by using the formula for $E(\psi_c(r_i))$ from (2.8) for the Poisson data. As $y_i \sim \text{Poisson}(\tilde{\mu}_i)$, it is clear that

$$E\left[\frac{1}{K}\sum_{i=1}^{K} q(y_i)\right] = -\frac{g(\tilde{x}_K)c}{K} = b(\beta), \text{ say},$$

yielding

$$E(WMQLEF) \equiv E\left[\frac{1}{K}\sum_{i=1}^{K}\{q(y_i) - a(\beta)\}\right] = \frac{1}{K}\left[-g(\tilde{x}_K)c - \sum_{i=1}^{K} a(\beta)\right]. \tag{2.20}$$

Now to have a feel for the magnitude of this expectation, let us consider $\beta = 1.0$, $\delta = 0.0,\ 1.0,\ 2.0,\ 3.0,\ 4.0$, and $K = 20$ and 100. For selected values of c, namely, for $c = 1.6,\ 1.4$, and 1.2, we compute $E(WMQLEF)$ by (2.20) and report these expected values in Table 2.1.

Computation of $E(FSMQLEF)$ for a special outlying Poisson model

To examine the consistency performance of the proposed FSMQL$_1$ estimators, we now compute the expectation of the FSMQL$_1$ estimating function from (2.5) for the same outlying Poisson model as considered in the last subsection. Note that the FSMQL$_1$EE (2.5) can be re-expressed as

$$\frac{1}{K}\sum_{i=1}^{K} z(\tilde{x}_i)\left[u_i - \xi_i\right] = 0, \tag{2.21}$$

where $u_i = \psi_c(r_i)$, $z(\tilde{x}_i) = w(\tilde{x}_i)\frac{\partial}{\partial\beta}(u_i - \xi_i)\left\{var(u_i)\right\}^{-1}$, and $\xi_i = \frac{1}{K}\sum_{i=1}^{K} E\left(\psi_c(r_i)\right)$. One may then write the FSMQL$_1$ estimating function from (2.21) as

$$FSMQL_1EF \equiv \frac{1}{K}\sum_{i=1}^{K} z(\tilde{x}_i)\left[u_i - \xi_i\right].$$

The expectation of this function can be obtained as

$$E(FSMQL_1EF) \equiv E\left[\frac{1}{K}\sum_{i=1}^{K} z(\tilde{x}_i)\left\{u_i - \xi_i\right\}\right] = -\frac{1}{K}\left[\sum_{i=1}^{K} z(\tilde{x}_i)\xi_i + z(\tilde{x}_K)c\right]. \tag{2.22}$$

The values of $E(FSMQL_1EF)$, for the same parameter values used in the WMQL approach, are also reported in Table 2.1 for the comparison with the values of $E(WMQLEF)$.

It is clear from Table 2.1 that none of the two approaches yields unbiased estimates, the proposed FSMQL$_1$ approach being much better than the existing WMQL approach. This is because of the following two reasons: (i) the amount of biases produced by the FSMQL$_1$ approach, i.e. $|E(FSMQL_1EF) - 0|$ is in general close to zero, whereas the WMQL approach produced large values for $|E(WMQLEF) - 0|$; (ii) the amount of biases produced by the FSMQL$_1$ approach appears to be almost the same and insignificant irrespective of the values of δ and c. This finding indicates that it may be possible to improve the FSMQL$_1$ estimators by constructing a bias adjustment under this approach for a given c, which is however not a serious issue (as biases are quite small) and it is beyond the scope of the present chapter. As far as the ranges of the biases produced by the WMQL approach are concerned, the variation in the amount of biases appears to be quite large. Thus, in many cases, this approach may be useless.

To be more specific, $E(WMQLEF)$ appears to be larger than $E(FSMQL_1EF)$ for $\delta \geq 1.0$. The amount of bias is quite large under the WMQL approach as compared to the FSMQL$_1$ approach, specially for large c such as $c = 1.6$. For example, when $c = 1.6$ and $\delta = 2.0$, the expected values of the $WMQLEF$ are -0.820 and -0.142 for $K = 20$ and 100, respectively, whereas the corresponding expected values of the $FSMQL_1EF$ are -0.071 and -0.075, respectively. For $\delta = 4.0$, for example, WMQL appears to perform quite poorly. Note that when $\delta = 0.0$, that is, the data do not contain any outlier, the WMQL approach appears to produce estimates with lower biases. Note that the FSMQL$_1$ approach pays some price when the data do not contain any outliers. This is because this approach is constructed to downweigh the outliers properly which not necessarily produces zero bias when $\delta = 0.0$.

We have considered another scenario of outlier, where we assume that for all $i = 1, \ldots, K$, observation y_i arises from the Poisson distribution with large mean

$\mu_i = exp(x_i\beta)$ with $x_i = 4$, but the covariates are observed as

$$\tilde{x}_i = \begin{cases} x_i, & i = 1, \ldots, K-1 \\ x_i - \delta, & i = K \end{cases}.$$

with a positive δ.

For this set up, the expected values of the estimating functions involving $\tilde{x}_i$ under the WMQL and FSMQL_1 approaches are computed by (2.20) and (2.22), respectively, and reported in Table 2.2. In this case, the WMQL approach performs worse than the FSMQL_1 approach for small values of δ and large values of c, whereas for the large values of δ such as for $2 \leq \delta \leq 4$ and small value of c, the WMQL approach appears to perform better. Note however that when there is no outlier, that is, $\delta = 0.0$, the FSMQL_1 approach appears to perform better (but still produces some bias) than the WMQL approach, whereas in Table 2.1 the reverse was true. We further note that even though the WMQL approach appears to perform relatively well under the second scenario, the question arises how important this scenario is. This is because when Poisson data are generated with large means (i.e. large variance), the suspected outlying observation may very well belong to the group of 'good' observations. Hence, generating an outlier in this way does not appear to be of practical interest.

2.2.2 Asymptotic distribution

It has been demonstrated empirically in the last subsection that the WMQL estimator of β due to Cantoni and Ronchetti (2001) can be highly biased, whereas the proposed FSMQL_1 estimator appears to be slightly biased. Note that the biasness arises because of the fact that $E(WMQLEF)$ and $E(FSMQL_1EF)$ are not quite zero under the present model. We take this biasness into account and derive the asymptotic distribution of the FSMQL$_1$ estimator of β, followed by the distribution for the WMQL estimator. Recall that the FSMQL$_1$ estimator for β is obtained by solving the FSMQL$_1$ estimating equation (2.5) given by

$$M_K(\beta) = \frac{1}{K}\sum_{i=1}^{K}\left[w(\tilde{x}_i)\frac{\partial}{\partial\beta}(u_i - \xi_i)\sigma_i^{-1}(\psi)\left\{u_i - \xi_i\right\}\right] = 0, \tag{2.23}$$

where $u_i = \psi_c(r_i)$, $\xi_i = \frac{1}{K}\sum_{i=1}^{K} E(\psi_c(r_i))$, and $\sigma_i(\psi) = var(u_i - \xi_i)$. Since $E[M_K(\beta)] \neq 0$, for convenience, we first derive the asymptotic distribution of a fully consistent estimator obtained by solving

$$M_K^*(\beta) = M_K(\beta) - \Delta_c = 0, \tag{2.24}$$

where $\Delta_c = E[M_K(\beta)]$. Let $\hat{\beta}_u^*$ be the solution of the unbiased estimating equation $M_K^*(\beta) = 0$, whereas the solution of the biased estimating equation $M_K(\beta) = 0$ will be denoted by $\hat{\beta}_b^*$. The asymptotic distribution of $\hat{\beta}_u^*$ is given in Theorem 5, whereas the asymptotic distribution of $\hat{\beta}_b^*$ is provided in Theorem 6.

Theorem 5: Let β_0^* be the true value of β. Also, let the following conditions c1 to c6 be satisfied [see, Amemiya (1985), section 4.1, p. 105-114]: **c1.** Let B be an open subset of the Euclidean $p-$space. Thus, the true value β_0^* is an interior point of B; **c2.** $M_K^*(\beta)$ exists and is continuous in an open neighborhood $N_1(\beta_0^*)$ of β_0^*. Also, $\int M_K^*(\beta)\partial\beta$ is a measurable function of the response for all $\beta \in B$. This implies that $\int M_K^*(\beta)\partial\beta$ is continuous for $\beta \in N_1$; **c3.** There exists an open neighborhood $N_2(\beta_0^*)$ of β_0^* such that $\frac{1}{K}\int M_K^*(\beta)\partial\beta$ converges to a non-stochastic function $M^*(\beta)$ in probability uniformly in β in N_2; **c4.** $\frac{\partial}{\partial\beta'}M_K^*(\beta)$ exists and is continuous in an open, convex neighborhood of β_0^*; **c5.** $\frac{1}{K}\frac{\partial}{\partial\beta'}M_K^*(\beta)|_{\beta^{**}}$ converges to a finite non-singular matrix $A(\beta_0^*) = limE\frac{1}{K}\frac{\partial}{\partial\beta'}M_K^*(\beta)|_{\beta_0^*}$ in probability for any sequence β^{**} such that $plim\beta^{**} = \beta_0^*$; **c6.** Asymptotically $\frac{1}{\sqrt{K}}M_K^*(\beta)|_{\beta_0^*} \sim N[0, C(\beta_0^*)]$, where

$C(\beta_0^*) = limE\frac{1}{K}M_K^*(\beta)|_{\beta_0^*} \times M_K^{*\prime}(\beta)|_{\beta_0^*}$. It then follows that as $K \rightarrow \infty$

$$\sqrt{K}\ (\hat{\beta}_u^* - \beta_0^*) \sim N[0, A(\beta_0^*)^{-1}C(\beta_0^*)A(\beta_0^*)^{-1}].$$

Proof: Assume that conditions c1-c6 are satisfied. Let $\tilde{B}$ be the set of solutions of the estimating equation $M_K^*(\beta) = 0$ for β. Also, let $\{\hat{\beta}_u^*\}$ be a sequence obtained by choosing one element from $\tilde{B}$ such that $plim\hat{\beta}_u^* = \beta_0^*$. Then, following Taylor expansion, one may write

$$M_K^*(\beta)|_{\hat{\beta}_u^*} = M_K^*(\beta)|_{\beta_0^*} + \frac{\partial}{\partial\beta'}M_K^*(\beta)|_{\beta^{**}}(\hat{\beta}_u^* - \beta_0^*),$$

where β^{**} lies between $\hat{\beta}_u^*$ and β_0^*. Since the left-hand side is equal to zero,

$$\sqrt{K}(\hat{\beta}_u^* - \beta_0^*) = -\left[\frac{1}{K}\frac{\partial}{\partial\beta'}M_K^*(\beta)|_{\beta^{**}}\right]^{-1}\frac{1}{\sqrt{K}}M_K^*(\beta)|_{\beta_0^*}.$$

Now, by (c5), (c6), and Slutsky theorem, one may asymptotically write

$$\sqrt{K}\ (\hat{\beta}_u^* - \beta_0^*) \sim N[0, A(\beta_0^*)^{-1}C(\beta_0^*)A(\beta_0^*)^{-1}].$$

We now provide the asymptotic distribution of $\hat{\beta}_b^*$ (solution of $M_K(\beta) = 0$) in Theorem 6.

Theorem 6: Let $\beta_b^* = \beta_u^* + \delta_K^*$, where $\sqrt{K}\delta_K^*$ is assumed to converge in probability to $\tilde{\mu}_c = -E[\frac{\partial}{\partial\beta'}M_K(\beta)]^{-1}K^{\frac{1}{2}}\Delta_c$. It then follows that asymptotically (as $K \rightarrow \infty$)

$$\sqrt{K}\ (\hat{\beta}_b^* - \beta_0^*) \sim N[\tilde{\mu}_c, \tilde{V}_c],$$

where $\tilde{V}_c = A^\dagger(\beta_0^*)^{-1}C^\dagger(\beta_0^*)A^\dagger(\beta_0^*)^{-1}$ with $A^\dagger(\beta_0^*) = limE\frac{1}{K}\frac{\partial}{\partial\beta'}[M_K(\beta) - \Delta_c]|_{\beta_0^*}$ and $C^\dagger(\beta_0^*) = limE\frac{1}{K}[M_K(\beta) - \Delta_c]|_{\beta_0^*} \times [M_K(\beta) - \Delta_c]'|_{\beta_0^*}$.

Proof: The theorem follows by using the transformation $M_K^*(\beta) = M_K(\beta) - \Delta_c$ in Theorem 5.

Note that the asymptotic distribution of the WMQL estimator ($\hat{\beta}^*_{b^*}$) may be derived in the manner similar to that of $\hat{\beta}^*_b$, where $\hat{\beta}^*_{b^*}$ is the solution of the estimating equation (2.2) given by

$$Q_K(\beta) = \frac{1}{K}\sum_{i=1}^{K} Q_i(\beta) = 0, \tag{2.25}$$

where $Q_i(\beta) = P_i(\beta) - a(\beta)$ with $P_i(\beta) = w(\tilde{x}_i)\frac{\partial \tilde{\mu}_i}{\partial \beta}\tilde{V}^{-\frac{1}{2}}(\tilde{\mu}_i)u_i$ and $a(\beta) = \frac{1}{K}\sum_{i=1}^{K} E[P_i(\beta)]$. Now, for $\Delta^*_c = E[Q_K(\beta)]$, $\tilde{\mu}^*_c = -E[\frac{\partial}{\partial \beta'}Q_K(\beta)]^{-1}K^{\frac{1}{2}}\Delta^*_c$, and $\tilde{V}^*_c = A^{\dagger *}(\beta^*_0)^{-1}C^{\dagger *}(\beta^*_0)A^{\dagger *}(\beta^*_0)^{-1}$ with $A^{\dagger *}(\beta^*_0) = lim E\frac{1}{K}\frac{\partial}{\partial \beta'}[Q_K(\beta) - \Delta^*_c]|_{\beta^*_0}$ and $C^{\dagger *}(\beta^*_0) = lim E\frac{1}{K}[Q_K(\beta) - \Delta^*_c]|_{\beta^*_0} \times [Q_K(\beta) - \Delta^*_c]'|_{\beta^*_0}$, it follows by the calculations similar to that of Theorem 6 that $\sqrt{K}(\hat{\beta}^*_{b^*} - \beta^*_0)$ has a multivariate normal distribution with mean $\tilde{\mu}^*_c$ and variance $\tilde{V}^*_c$, i.e.

$$\sqrt{K}(\hat{\beta}^*_{b^*} - \beta^*_0) \sim N[\tilde{\mu}^*_c, \tilde{V}^*_c].$$

Note that even though β^*_0 has been used as the true value of β, in rest of the chapter, for convenience, we simply use β for β^*_0. By the same token, $\hat{\beta}^*_b$ under the FSMQL$_1$ approach will be expressed as $\hat{\beta}_{FSMQL_1}$ and similarly $\hat{\beta}^*_{b^*}$ under the WMQL approach will be expressed as $\hat{\beta}_{WMQL}$.

2.3 A Simulation Study

Recall from section 2.2 that both the FSMQL$_1$ and WMQL estimating functions were found to be biased, the FSMQL$_1$ estimating function being the better. More specifically, the expectation of the FSMQL$_1$ estimating function was found to be closer to zero as compared to the expected value of the WMQL estimating function for selected values of parameters under the Poisson model. This indicates that the estimators to be obtained from their estimating equations will be biased. Note that when an estimator is expected to be highly biased, a small variance for this estimator indicates that the estimator may not converge to the true parameter value in general. Consequently, it is reasonable to examine the relative performance of the competitive estimators by comparing their relative biases, where, for example, the relative bias (RB) of the WMQL estimator for a component of β, say β_k $(k = 1, \ldots, p)$ is defined by

$$\text{RB }(\hat{\beta}_{k,WMQL}) = \frac{|\hat{\beta}_{k,WMQL} - \beta_k|}{\text{s.e. }(\hat{\beta}_{k,WMQL})} \times 100. \tag{2.26}$$

In this section, the relative biases of the components of $\hat{\beta}_{WMQL}$ (Cantoni and Ronchetti (2001)) and those of the proposed estimators $\hat{\beta}_{FSMQL_1}$ and $\hat{\beta}_{FSMQL_2}$ will be compared through a simulation study. For the sake of completeness, we also examine the relative biases of the components of the SSMQL estimator of β, where $\hat{\beta}_{SSMQL}$ is obtained by solving the estimating equation (2.7). Recall that this estimating equation (2.7) was a special case of the estimating equation for β under the longitudinal set up considered by Cantoni (2004). As far as the data are concerned, we consider both Poisson and binary cases in the simulation study. Under each of these models we consider two scenarios. First, the data contain a single outlier and secondly, the data contain two outliers. Note that no simulation studies were under taken by Cantoni and Ronchetti (2001) to examine the performance of their WMQL estimators. The present simulation study, however, reveals that these WMQL estimators are in fact highly biased which contradicts the claims with regard to the consistency made by these authors, namely Cantoni and Ronchetti (2001).

We now turn back to the proposed simulation study and provide the simulation design and selected parameter values for each of the two scenarios under each of the two models. The performances of the WMQL, SSMQL, FSMQL_1, and FSMQL_2 estimators under both Poisson and binary outlier models are examined for small as well as large samples such as $K = 20,\ 30,\ 60,$ and 100. As far as the regression parameters are concerned, we consider $p = 2$ with $\beta = (\beta_1, \beta_2)' \equiv (1.0, 0.5)'$. Recall that for $k = 1,\ 2$, $\hat{\beta}_{k,WMQL}$, $\hat{\beta}_{k,SSMQL}$, $\hat{\beta}_{k,FSMQL_1}$, and $\hat{\beta}_{k,FSMQL_2}$ denote the WMQL, SSMQL, FSMQL_1, and FSMQL_2 estimators of β_k $(k = 1,\ 2)$, respectively. The main purpose of the simulation study is to compare the simulation means (SM), standard errors (SSE), and relative biases (RB) of these four estimators under both Poisson and binary models in the presence of one or two outliers. We make this comparison based on 1000 simulations. The simulation design and the corresponding simulation results are described below, under both Poisson and binary models.

Note that in section 2.1.2, we have provided the robust functions and their basic properties for both one sided and two sided binary outliers. In the simulation study, we however consider these cases with one and two sided outliers under both Poisson and binary models. Further note that the one sided outlier cases will be dealt with by generating a single outlier (different from the bulk of the observations), whereas without any loss of generality, the two sided outlier cases will be dealt with by generating two outliers in two opposite directions from the bulk of the observations.

2.3.1 Poisson model with one or two outliers

(a) Poisson model with a single outlier

To generate K count observations with one outlier, we first assume that in the absence of outliers, $y_1, \ldots, y_i, \ldots, y_K$ follow a Poisson model with two covariates x_{i1} and x_{i2}, namely

$$P(Y_i = y_i|x_i) = \frac{exp(-\mu_i)\mu_i^{y_i}}{y_i!},$$

where $\mu_i = exp(x_i'\beta)$ with $x_i = (x_{i1}, x_{i2})'$. Suppose that the values of these two covariates are chosen from

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

respectively, for all $i = 1, \ldots, K$. Suppose that i' takes a value between 1 and K. Now, to consider $y_{i'}$ as an outlying value, that is, to have a data set of size K with one outlier, we then shift the values of $x_{i'1}$ and $x_{i'2}$ as

$$\tilde{x}_{i'1} = x_{i'1} + \delta \text{ and } \tilde{x}_{i'2} = x_{i'2} + \delta,\ \delta > 0,$$

respectively, but retain

$$\tilde{x}_{i1} = x_{i1} \text{ and } \tilde{x}_{i2} = x_{i2},$$

for all $i \neq i'$. As far as the shifting is concerned, we, for convenience, use $\delta = 2.0$. Thus, $y_1, \ldots, y_K$ refer to a sample of K count observations with $y_{i'}$ as the single outlier.

Estimation performance in a single outlier case

Note that if there were no outliers in the count data, one would have used the estimating equation (2.4) to obtain the QL estimates of β_1 and β_2. These estimates are consistent. To have a feel how an outlier can affect the estimation of the parameters, we conduct a simulation study by generating the count data with an outlier as discussed above and estimate the parameters by using the QL estimating equation (2.4). The results obtained from 1000 simulations are reported in Table 2.3. It is clear from this table that the estimates are highly biased and hence inconsistent. This shows the necessity for a robust estimation technique such that the outlier can have no or little effect on the estimates.

For the robust estimation of the regression parameter β in the presence of outlier(s), one requires to take the tuning constant parameter c into account. Here, we consider three different values for the tuning constant $c = 1.6,\ 1.4,$ and 1.2. We now proceed to examine the performance of the robust estimation techniques discussed earlier in the presence of an outlier. For the purpose, by using the responses generated above with an outlier along with their corresponding covariates, we first compute the expectation and the variance of the robust function $\psi_c(r_i)$ as well as the gradients of the robust functions following section 2.1.1 for the Poisson model. We then use these results in the estimating equations (2.2), (2.7), (2.5), and (2.6) and obtain the estimates of β_1 and β_2. For $k = 1, 2$, these estimates are referred to as $\hat{\beta}_{k,WMQL}$, $\hat{\beta}_{k,SSMQL}$, $\hat{\beta}_{k,FSMQL_1}$, and $\hat{\beta}_{k,FSMQL_2}$, respectively. Now, by using the 1000 simulated values of $\hat{\beta}_k$, we obtain the simulated means (SM), simulated standard errors (SSE), and percentages of relative biases (RB) under a selected a estimation approach. The simulation results obtained under the Poisson model with a single outlier are reported in Table 2.4.

Note that when the sample size is small such as $K = 20$, the WMQL approach did not yield any convergent estimates, whereas the SSMQL approach yielded estimates only in 118 simulations out of 1000, which are not reported. It is clear from Table 2.4 that all four approaches produce biased estimates for the regression parameter β, but the proposed FSMQL estimation approach always appears to produce estimates with negligible biases. For example, when $K = 60$ and $c = 1.4$, the estimates of β_1 are 0.507 and 0.567 with corresponding standard errors 0.206 and 0.211 under the WMQL and SSMQL approaches, respectively; whereas the proposed FSMQL_1 and FSMQL_2 approaches yield 0.899 and 0.893 with standard errors 0.307 and 0.279, respectively. Equivalently, the percentage of relative biases (RBs) of $\hat{\beta}_1$ are 240 and

206 under the WMQL and SSMQL approaches, respectively; whereas the percentage of RBs for the FSMQL_1 and FSMQL_2 approaches are only 33 and 38, respectively. Similarly, the WMQL and SSMQL estimates of β_2 are 0.600 and 0.589 with standard errors 0.188 and 0.186, respectively; whereas these estimates are 0.517 and 0.488 with standard errors 0.239 and 0.210 under the proposed FSMQL_1 and FSMQL_2 approaches, respectively. These results yielded RBs 53 and 48 for the estimates of β_2 under the WMQL and SSMQL approaches, respectively; whereas the FSMQL_1 and FSMQL_2 approaches yielded RBs only 7 and 6, respectively.

(b) Poisson model with two outliers

For the Poisson model with two outlying observations, first, count responses are generated in the manner similar to that of the case of a single outlier, but two covariates x_{i1} and x_{i2} were chosen as

$$x_{i1} \stackrel{iid}{\sim} N(1.25, 0.25) \text{ and } x_{i2} \stackrel{iid}{\sim} N(2.25, 0.5),$$

respectively. Note that in the single outlier case, x_{i1} and x_{i2} were generated from normal distributions with a small common mean 0.5, whereas for the two outliers case, we have assigned large mean values for x_{i1} and x_{i2}. This we have done so that the two sided shifting of the covariates may help to identify two outliers. After generating K count observations from a Poisson model with these covariate values, we then create two outliers, namely $y_{i'}$ and $y_{i''}$, where i' and i'' both can take values between 1 and K, but $i \neq i', i''$, by shifting the covariate values $x_{i'1}$ and $x_{i'2}$ as

$$\tilde{x}_{i'1} = x_{i'1} + \delta \text{ and } \tilde{x}_{i'2} = x_{i'2} + \delta, \ \delta > 0,$$

respectively; and $x_{i''1}$ and $x_{i''2}$ as

$$\tilde{x}_{i''1} = x_{i''1} - \delta \text{ and } \tilde{x}_{i''2} = x_{i''2} - \delta, \ \delta > 0,$$

respectively. Once again we consider for convenience $\delta = 2.0$. The remaining covariates remain the same as those of x_{i1} and x_{i2}. That is,

$$\tilde{x}_{i1} = x_{i1} \text{ and } \tilde{x}_{i2} = x_{i2},$$

for all $i \neq i', i''$, $i = 1, \ldots, K$.

Estimation performance in two outliers case

As far as the tuning constant is concerned, we choose the same tuning constant $c = 1.6$, 1.4, and 1.2 as in the single outlier case. Now, we obtain the estimates of the regression parameter β under the WMQL, SSMQL, FSMQL_1, and FSMQL_2 approaches by using the formulas from section 2.1.1 in the robust estimating equations (2.2), (2.7), (2.5), and (2.6), respectively. Table 2.5 displays the SM, SSE, and RB for $\hat{\beta}_1$ and $\hat{\beta}_2$ obtained from these four estimation approaches under a Poisson model in the presence of two outliers based on 1000 simulations. It is clear from Table 2.5 that the WMQL and SSMQL approaches produce much more higher relative biases than the FSMQL_1 and FSMQL_2 approaches, irrespective of sample sizes and the tuning constant values.

We now interpret some of the results reported in Table 2.5. For small sample size such as $K = 30$ and $c = 1.4$, the percentages of relative biases (RBs) of $\hat{\beta}_1$ are 466 and 360 under the WMQL and SSMQL approaches, respectively; whereas the proposed FSMQL_1 and FSMQL_2 approaches yield RBs only 8 and 20, respectively. For the estimate of β_2, the WMQL and SSMQL produce RBs 392 and 273, respectively; these are only 7 and 1 under the proposed FSMQL_1 and FSMQL_2 approaches. As the sample size increases, the estimates of the regression parameter β improve under the WMQL and SSMQL approaches, but still produce higher biases. For example, for

$K = 100$ and $c = 1.4$, the RBs of $\hat{\beta}_1$ under the WMQL and SSMQL approaches are 308 and 290, respectively; the proposed FSMQL_1 and FSMQL_2 approaches, however, produce RBs only 8 and 9, respectively. For the estimate of β_2, the WMQL and SSMQL approaches yield RBs 267 and 252, respectively, whereas these are found to be only 17 and 10 under the proposed FSMQL_1 and FSMQL_2 approaches, respectively.

Note that the above discussion clearly indicates that for the estimation of the regression effects, the FSMQL_1 and FSMQL_2 approaches appear to perform much better than the WMQL and SSMQL approaches. Between the FSMQL_1 and FSMQL_2 estimators, the FSMQL_1 estimators appear to perform better than the FSMQL_2 estimators.

2.3.2 Binary model with one or two outliers

In this subsection, we conduct a simulation study to evaluate the estimation performances of the robust WMQL, SSMQL, FSMQL_1, and FSMQL_2 estimation approaches for the binary data in the presence of one or two outliers. The estimating equations are constructed based on the basic properties of the robust function provided in section 2.1.2, whereas similar properties for the Poisson case were given in section 2.1.1.

(a) Binary model with a single outlier

For the contaminated binary model with a single outlier, first, we generate K binary responses $y_1, \ldots, y_i, \ldots, y_K$ assuming that the data do not contain any outliers. We generate these responses from a binary logistic model

$$P(Y_i = 1|x_i) = \frac{exp(x_i'\beta)}{1 + exp(x_i'\beta)},$$

with $x_i = (x_{i1}, x_{i2})'$ and $\beta = (\beta_1, \beta_2)'$. As far as the covariate values are concerned, we consider two covariates x_{i1} and x_{i2} as

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

respectively, for $i = 1, \ldots, K$. Then, in creating an outlier $y_{i'}$ where i' can take any value between 1 and K, we change the corresponding covariate values $x_{i'1}$ and $x_{i'2}$ as

$$\tilde{x}_{i'1} = x_{i'1} + \delta_1 \text{ and } \tilde{x}_{i'2} = x_{i'2} + \delta_2,\ \delta_1, \delta_2 > 0,$$

respectively. Note that for large positive δ_1 and δ_2, these modified covariates will be increased in magnitude yielding larger probability for $y_{i'} = 1$. Thus, we treat $y_{i'}$ as an outlier. For convenience, we use $\delta_1 = 2.0$ and $\delta_2 = 3.0$. As far as the remaining covariates are concerned, they are kept unchanged. That is, for $i \neq i'$ $(i = 1, \ldots, K)$, we consider

$$\tilde{x}_{i1} = x_{i1} \text{ and } \tilde{x}_{i2} = x_{i2}.$$

Estimation performance in a single outlier case

Note that as expected, when an outlier is present in the data, the traditional QL approach was found to perform poorly in estimating the regression effects under a Poisson fixed model. To examine the performance of the QL approach for the estimation of the parameters of a binary model in the presence of an outlier, we now conduct a simulation study by generating the data as discussed above. Next, we use the QL estimating equation (2.4) to estimate the regression effects β_1 and β_2. The results based on the 1000 simulations are reported in Table 2.6. Similar to the results in Table 2.3, the results in Table 2.6 also reveal that the QL approach without any modification for outlier performs poorly in estimating the parameters of a binary fixed model.

In order to remove the biases (see Table 2.3 and 2.6) those arise due to the presence of possible outliers, various modifications to the QL approach, such as the robust WMQL, SSMQL, FSMQL_1, and FSMQL_2 approaches have been discussed earlier in this chapter. We now examine the performances of these robust approaches for the binary data generated with an outlier. With regard to the values of the tuning constant related probability μ^{c_1}, we choose $\mu^{c_1} = 0.5,\ 0.6,$ and 0.9 in the simulation study. Next, we compute the expectation and the variance of the robust function $\psi_c(r_i)$ along with the gradients of the robust functions following section 2.1.2. These results are then used in the estimating equations (2.2), (2.7), (2.5), and (2.6) to obtain the estimates of β_k $(k = 1, 2)$ under the WMQL, SSMQL, FSMQL_1, and FSMQL_2 approaches, respectively. For a given estimation approach, we now compute the SM, SSE, and RB of $\hat{\beta}_k$ from 1000 simulated estimates of β_k. These results are given in Table 2.7.

It is clear from Table 2.7 that in general, the WMQL estimation approach due to Cantoni and Ronchetti (2001) produces highly biased estimates for both regression parameters β_1 and β_2. When the proposed FSMQL_1 and FSMQL_2 approaches are compared with the SSMQL approach, it is found that all these three approaches produce estimates for both parameters with small biases in most of the cases. To have a feel for the performance of the proposed FSMQL approach in estimating both β_1 and β_2, we, for example, refer to the case with $K = 30$ and $\mu^{c_1} = 0.6$. In this case, the percentage of relative biases (RBs) yielded by the WMQL approach in estimating β_1 and β_2 are 24 and 24; whereas RBs due to the SSMQL and the proposed FSMQL_1 and FSMQL_2 approaches are only 1 and 4; 1 and 5; and 1 and 4, respectively.

(b) Binary model with two outliers

For the contaminated binary model with two outliers, we, first, generate K binary responses in the same fashion as in the case of a single outlier with two covariates x_{i1} and x_{i2} chosen from the normal distribution as

$$x_{i1} \overset{iid}{\sim} N(0, 0.25) \text{ and } x_{i2} \overset{iid}{\sim} N(0, 0.5),$$

respectively, for all $i = 1, \ldots, K$. Suppose that two outlying observations $y_{i'}$ and $y_{i''}$ (where i' and i'' can take any value between 1 and K) arise due to a shift in the covariate values $x_{i'1}$ and $x_{i'2}$ as

$$\tilde{x}_{i'1} = x_{i'1} + \delta_1 \text{ and } \tilde{x}_{i'2} = x_{i'2} + \delta_2,\ \delta_1, \delta_2 > 0;$$

and in the covariate values $x_{i''1}$ and $x_{i''2}$ as

$$\tilde{x}_{i''1} = x_{i''1} - \delta_1 \text{ and } \tilde{x}_{i''2} = x_{i''2} - \delta_2,\ \delta_1, \delta_2 > 0,$$

respectively. Consequently, for the cases with large values of δ_1 and δ_2, $y_{i'}$ and $y_{i''}$ become outliers. Here, we consider $\delta_1 = 2.0$ and $\delta_2 = 3.0$. The remaining covariates remain the same as before.

Estimation performance in two outliers case

For the values of the tuning constant related probabilities, we choose $\mu^{c_1} = 0.6$ and $\mu^{c_2} = 0.4$. Next, following section 2.1.2, we obtain the estimates of the regression effect $\beta = (\beta_1, \beta_2)'$ from the estimating equations (2.2), (2.7), (2.5), and (2.6) for the WMQL, SSMQL, FSMQL_1, and FSMQL_2, approaches, respectively. The SM, SSE, and RB of $\hat{\beta}$ obtained from 1000 simulations are reported in Table 2.8 under a binary model with two outlying observations.

It is clear from the results of Table 2.8 that the proposed FSMQL and SSMQL approaches definitely produce the estimates of β_1 and β_2 with smaller biases as compared to the other competitive approaches. For example, when $K = 60$, the WMQL approach produces the percentage of relative biases (RBs) of $\hat{\beta}_1$ and $\hat{\beta}_2$ as 21 and 13; whereas these are found to be only 4 and 5; 7 and 8; and 4 and 4 under the SSMQL, FSMQL_1, and FSMQL_2 approaches, respectively.

2.4 Some Remarks

There exists a vast literature for the robust inferences in the linear models in the presence of one or more outliers. See, for example, the references in Huber (2004) and Rousseeuw and Leroy (1987). When compared with the linear models, there however does not exist adequate discussions for the robust inferences in the non-linear regression models for discrete such as count and binary data. As the exact likelihood inferences in the presence of outliers are complicated, recently Cantoni and Ronchetti (2001) introduced a corrected (based on the Fisher consistency concept) Mallows-type quasi-likelihood (MQL) estimating equations approach to obtain consistent estimates for the parameters of non-linear regression models in the independence set up. More recently, among others, Cantoni (2004) and Sinha (2004) have dealt with similar estimating equations approach in the correlated set up for the longitudinal and familial data, respectively.

In the thesis, we have suggested a new quasi-likelihood estimating equations approach that unlike Cantoni and Ronchetti (2001) takes the correlations and gradients of the robust functions into account. It is demonstrated both asymptotically and by simulations that the proposed fully standardized MQL (FSMQL) approach definitely yields less biased estimates for the regression parameters as compared to the WMQL approach of Cantoni and Ronchetti (2001) in the independence set up. To be specific, in the Poisson case, the FSMQL approach appears to perform much better than the SSMQL approach. When the performances of the FSMQL_1 and FSMQL_2 are compared, it is found that in general, the FSMQL_1 approach performs better than the FSMQL_2 approach. In the binary case, the SSMQL, FSMQL_1, and FSMQL_2 approaches appear to produce small biases for the regression parameters. The SSMQL and FSMQL_2 approaches performs almost equally and slightly better than the FSMQL_1 approach.

Note that as the FSMQL_1 approach is almost always better than the other competitive approaches under the Poisson model, and also because this approach is only slightly worse than the SSMQL and FSMQL_2 approaches under the binary model, the FSMQL_1 approach is generally recommended for the estimation of the regression effects in the presence of possible outliers both for count and binary data.

Note that as far as the tuning constant (c) is concerned, this is typically chosen to achieve a certain level of asymptotic efficiency at the underlying distribution. For example, Cantoni and Ronchetti (2001) used $c = 1.6$ and $c = 1.2$ in their numerical illustrations for the analysis of count and binomial data, respectively in the presence of outliers. In the simulation studies for the count data, we have used three values of c, namely, $c = 1.2,\ 1.4,$ and 1.6, to see how these c values affect the performances of the estimates under three different estimation approaches. For the binary case, we have conducted an extensive simulation study by considering the tuning constant related probability $\mu^{c_1} = 0.5,\ 0.6,$ and 0.9, whereas the traditional tuning constant has been used in the existing literature. For almost all values of the tuning constant or the tuning constant related probability, in general, the proposed FSMQL approach was found to produce smaller/equal biases as compared to the other two competitive approaches.

Table 2.1: [**Checking the difference of the expectation of the estimating function from zero**] Expectation of the estimating functions under the WMQL and FSMQL_1 approaches for different sample sizes (K), tuning constant values (c), and amounts of contamination in the selected covariate (δ); Kth observation being an outlier arising through $\tilde{x}_K = 1.0 + \delta$.

		$E(WMQLEF)$			$E(FSMQL_1EF)$		
			c			c	
K	δ	1.6	1.4	1.2	1.6	1.4	1.2
20	0.0	-0.067	-0.054	0.005	-0.073	-0.077	-0.133
	1.0	-0.332	-0.281	-0.194	-0.071	-0.077	-0.128
	2.0	-0.820	-0.709	-0.555	-0.071	-0.075	-0.129
	3.0	-1.674	-1.458	-1.196	-0.071	-0.074	-0.130
	4.0	-3.133	-2.733	-2.291	-0.070	-0.074	-0.131
100	0.0	0.036	0.037	0.084	-0.076	-0.080	-0.152
	1.0	-0.022	-0.013	0.040	-0.075	-0.080	-0.150
	2.0	-0.142	-0.118	-0.049	-0.075	-0.080	-0.150
	3.0	-0.374	-0.322	-0.223	-0.075	-0.079	-0.150
	4.0	-0.806	-0.699	-0.547	-0.075	-0.079	-0.150

Table 2.2: **[Checking the difference of the expectation of the estimating function from zero]** Expectation of the estimating functions under the WMQL and FSMQL_1 approaches for different sample sizes (K), tuning constant values (c), and amounts of contamination in the selected covariate (δ); Kth observation being an outlier arising through $\tilde{x}_K = 4.0 - \delta$.

		$E(WMQLEF)$			$E(FSMQL_1EF)$		
		c			c		
K	δ	1.6	1.4	1.2	1.6	1.4	1.2
20	0.0	2.550	2.280	2.064	-0.285	-0.342	-0.505
	1.0	1.302	1.190	1.127	-0.294	-0.357	-0.525
	2.0	0.670	0.638	0.648	-0.304	-0.393	-0.529
	3.0	0.368	0.369	0.422	-0.335	-0.392	-0.616
	4.0	0.233	0.250	0.318	-0.322	-0.730	-0.824
100	0.0	0.721	0.681	0.695	-0.304	-0.362	-0.533
	1.0	0.464	0.457	0.502	-0.306	-0.366	-0.538
	2.0	0.336	0.345	0.405	-0.308	-0.373	-0.539
	3.0	0.275	0.290	0.360	-0.314	-0.373	-0.557
	4.0	0.248	0.267	0.339	-0.312	-0.445	-0.602

Table 2.3: [**For count data with a single outlier**] Simulated means (SM), simulated standard errors (SSE), and relative biases (RB) of the QL estimates of the regression parameters for different sample sizes under the Poisson model with $\beta_1 = 1.0$ and $\beta_2 = 0.5$ in the presence of a single outlier.

		Estimate	
K	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$
20	SM	0.361	0.242
	SSE	0.416	0.349
	RB	154	74
30	SM	0.233	0.522
	SSE	0.241	0.277
	RB	319	8
60	SM	0.230	0.667
	SSE	0.186	0.194
	RB	382	86
100	SM	0.486	0.597
	SSE	0.118	0.134
	RB	434	72

Table 2.4: [**For count data with a single outlier**] Simulated means (SM), simulated standard errors (SSE), and relative biases (RB) of the WMQL, SSMQL, $FSMQL_1$, and $FSMQL_2$ estimates of the regression parameters for different sample sizes and selected values of the tuning constant c under the Poisson model with $\beta_1 = 1.0$ and $\beta_2 = 0.5$ in the presence of a single outlier.

			Estimation Method							
			WMQL		SSMQL		$FSMQL_1$		$FSMQL_2$	
K	c	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
20	1.6	SM	-	-	-	-	0.808	0.430	0.802	0.431
		SSE	-	-	-	-	0.517	0.366	0.530	0.368
		RB	-	-	-	-	37	19	37	19
	1.4	SM	-	-	-	-	0.880	0.486	0.862	0.468
		SSE	-	-	-	-	0.481	0.362	0.426	0.313
		RB	-	-	-	-	25	4	32	10
	1.2	SM	-	-	-	-	0.896	0.489	0.887	0.477
		SSE	-	-	-	-	0.485	0.368	0.438	0.328
		RB	-	-	-	-	21	3	26	7
30	1.6	SM	0.180	0.586	0.300	0.559	0.896	0.404	0.892	0.404
		SSE	0.246	0.277	0.244	0.263	0.390	0.346	0.385	0.354
		RB	334	31	287	23	27	28	28	27
	1.4	SM	0.235	0.525	0.382	0.522	0.928	0.424	0.889	0.436
		SSE	0.210	0.246	0.250	0.252	0.404	0.320	0.361	0.275
		RB	364	10	247	9	18	24	31	23
	1.2	SM	0.251	0.526	0.428	0.522	0.957	0.435	0.937	0.447
		SSE	0.238	0.285	0.277	0.270	0.404	0.331	0.373	0.294
		RB	315	9	207	8	11	20	17	18

Contd........Table 2.4

			Estimation Method							
			WMQL		SSMQL		FSMQL_1		FSMQL_2	
K	c	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
60	1.6	SM	0.440	0.644	0.500	0.630	0.850	0.506	0.849	0.496
		SSE	0.228	0.198	0.230	0.196	0.328	0.229	0.322	0.225
		RB	246	73	217	66	46	3	47	2
	1.4	SM	0.507	0.600	0.567	0.589	0.899	0.517	0.893	0.488
		SSE	0.206	0.188	0.211	0.186	0.307	0.239	0.279	0.210
		RB	240	53	206	48	33	7	38	6
	1.2	SM	0.505	0.612	0.570	0.595	0.924	0.508	0.923	0.491
		SSE	0.221	0.197	0.226	0.195	0.363	0.271	0.335	0.246
		RB	224	57	190	49	21	3	23	4
100	1.6	SM	0.655	0.574	0.680	0.570	0.903	0.470	0.895	0.465
		SSE	0.152	0.134	0.153	0.134	0.228	0.167	0.222	0.163
		RB	227	55	208	53	43	18	48	21
	1.4	SM	0.695	0.548	0.721	0.544	0.936	0.474	0.929	0.457
		SSE	0.142	0.129	0.144	0.129	0.224	0.172	0.206	0.151
		RB	214	37	194	34	29	15	34	29
	1.2	SM	0.698	0.557	0.727	0.550	0.945	0.492	0.944	0.482
		SSE	0.152	0.135	0.153	0.135	0.275	0.204	0.253	0.180
		RB	199	42	178	37	20	4	22	10

Table 2.5: [**For count data with two outliers**] Simulated means (SM), simulated standard errors (SSE), and relative biases (RB) of the WMQL, SSMQL, FSMQL_1, and FSMQL_2 estimates of the regression parameters for different sample sizes and selected values of the tuning constant c under the Poisson model with $\beta_1 = 1.0$ and $\beta_2 = 0.5$ in the presence of two outliers.

			Estimation Method							
			WMQL		SSMQL		FSMQL_1		FSMQL_2	
K	c	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
20	1.6	SM	0.090	0.827	0.209	0.785	0.993	0.484	1.002	0.479
		SSE	0.307	0.132	0.227	0.110	0.218	0.115	0.211	0.113
		RB	297	246	349	259	3	14	1	19
	1.4	SM	0.140	0.795	0.218	0.782	0.993	0.485	0.967	0.490
		SSE	0.213	0.099	0.212	0.106	0.232	0.122	0.232	0.121
		RB	404	299	368	265	3	12	14	8
	1.2	SM	0.157	0.788	0.268	0.761	0.981	0.486	0.950	0.495
		SSE	0.200	0.096	0.217	0.109	0.239	0.126	0.233	0.123
		RB	421	300	337	238	8	11	21	4
30	1.6	SM	0.348	0.781	0.437	0.744	0.993	0.486	1.003	0.481
		SSE	0.140	0.072	0.136	0.070	0.197	0.105	0.197	0.105
		RB	465	392	413	347	4	13	2	18
	1.4	SM	0.374	0.770	0.464	0.731	0.983	0.493	0.959	0.501
		SSE	0.134	0.069	0.149	0.085	0.206	0.109	0.203	0.108
		RB	466	392	360	273	8	7	20	1
	1.2	SM	0.412	0.752	0.495	0.718	0.980	0.488	0.955	0.497
		SSE	0.142	0.073	0.141	0.073	0.208	0.110	0.207	0.111
		RB	414	343	358	297	10	11	22	3

Contd........Table 2.5

			Estimation Method							
			WMQL		SSMQL		$FSMQL_1$		$FSMQL_2$	
K	c	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
60	1.6	SM	0.594	0.687	0.633	0.670	0.989	0.494	0.994	0.491
		SSE	0.107	0.057	0.109	0.058	0.168	0.089	0.164	0.087
		RB	381	310	337	293	6	7	4	10
	1.4	SM	0.610	0.678	0.650	0.662	0.985	0.497	0.970	0.502
		SSE	0.108	0.057	0.109	0.058	0.178	0.096	0.175	0.094
		RB	360	313	321	280	8	4	17	2
	1.2	SM	0.633	0.670	0.670	0.653	0.979	0.496	0.965	0.500
		SSE	0.110	0.058	0.111	0.059	0.192	0.103	0.188	0.101
		RB	333	290	298	260	11	4	19	0
100	1.6	SM	0.745	0.620	0.760	0.613	1.000	0.490	1.003	0.487
		SSE	0.078	0.042	0.078	0.043	0.121	0.066	0.122	0.067
		RB	328	284	308	266	0	16	3	19
	1.4	SM	0.757	0.615	0.772	0.608	1.001	0.488	0.989	0.493
		SSE	0.079	0.043	0.079	0.043	0.131	0.071	0.129	0.070
		RB	308	267	290	252	8	17	9	10
	1.2	SM	0.773	0.607	0.786	0.601	1.004	0.485	0.989	0.491
		SSE	0.080	0.044	0.081	0.044	0.143	0.077	0.139	0.076
		RB	284	246	263	229	3	19	8	12

Table 2.6: [**For binary data with a single outlier**] Simulated means (SM), simulated standard errors (SSE), and relative biases (RB) of the QL estimates of the regression parameters for different sample sizes under the binary model with $\beta_1 = 1.0$ and $\beta_2 = 0.5$ in the presence of a single outlier.

		Estimate	
K	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$
20	SM	1.609	-0.171
	SSE	1.399	0.991
	RB	44	68
30	SM	1.819	-0.414
	SSE	1.209	1.099
	RB	68	83
60	SM	1.225	0.253
	SSE	0.803	0.734
	RB	28	34
100	SM	1.215	0.276
	SSE	0.548	0.462
	RB	39	49

Table 2.7: [**For binary data with a single outlier**] Simulated means (SM), simulated standard errors (SSE), and relative biases (RB) of the WMQL, SSMQL, FSMQL_1, and FSMQL_2 estimates of the regression parameters for different sample sizes and selected values of the tuning constant related probability μ^{c_1} under the binary model with $\beta_1 = 1.0$ and $\beta_2 = 0.5$ in the presence of a single outlier.

			Estimation Method							
			WMQL		SSMQL		FSMQL_1		FSMQL_2	
K	μ^{c_1}	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
20	0.5	SM	1.238	0.382	1.088	0.617	1.091	0.639	1.101	0.603
		SSE	1.141	1.273	1.027	1.252	1.034	1.265	1.029	1.225
		RB	21	9	9	9	9	11	10	8
	0.6	SM	1.237	0.298	1.181	0.592	1.188	0.620	1.192	0.581
		SSE	1.110	1.211	1.208	1.207	1.231	1.228	1.204	1.183
		RB	21	17	15	8	15	10	16	7
	0.9	SM	1.210	-0.086	1.021	0.460	1.048	0.495	1.028	0.446
		SSE	1.065	0.925	1.048	1.067	1.092	1.087	1.028	1.047
		RB	20	63	2	4	4	0	3	5
30	0.5	SM	1.286	0.221	1.025	0.560	1.018	0.580	1.022	0.565
		SSE	1.409	1.527	1.343	1.436	1.374	1.468	1.335	1.422
		RB	20	18	2	4	1	5	2	5
	0.6	SM	1.329	0.138	1.011	0.550	1.006	0.571	1.012	0.553
		SSE	1.385	1.497	1.340	1.434	1.366	1.462	1.324	1.411
		RB	24	24	1	4	1	5	1	4
	0.9	SM	1.341	-0.096	0.962	0.471	0.938	0.524	0.971	0.463
		SSE	1.139	1.147	1.319	1.404	1.340	1.427	1.312	1.395
		RB	30	52	3	2	5	2	2	3

Contd........Table 2.7

			Estimation Method							
			WMQL		SSMQL		FSMQL_1		FSMQL_2	
K	μ^{c_1}	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
60	0.5	SM	1.099	0.434	1.055	0.515	1.055	0.519	1.055	0.515
		SSE	0.797	0.792	0.795	0.789	0.804	0.798	0.795	0.788
		RB	13	8	7	2	7	2	7	2
	0.6	SM	1.105	0.407	1.047	0.513	1.046	0.519	1.049	0.512
		SSE	0.794	0.788	0.792	0.786	0.802	0.796	0.792	0.785
		RB	13	12	6	2	6	2	6	2
	0.9	SM	1.161	0.194	1.002	0.487	0.994	0.503	1.003	0.486
		SSE	0.777	0.760	0.779	0.765	0.782	0.777	0.779	0.764
		RB	21	40	0	2	1	0	0	2
100	0.5	SM	1.052	0.471	1.013	0.522	1.011	0.524	1.013	0.522
		SSE	0.541	0.490	0.522	0.466	0.526	0.469	0.522	0.466
		RB	10	6	2	5	2	5	3	5
	0.6	SM	1.064	0.447	1.008	0.520	1.007	0.523	1.009	0.520
		SSE	0.545	0.494	0.521	0.465	0.525	0.469	0.521	0.465
		RB	12	11	2	4	1	5	2	4
	0.9	SM	1.179	0.252	0.980	0.507	0.976	0.513	0.980	0.507
		SSE	0.582	0.538	0.513	0.458	0.518	0.462	0.513	0.457
		RB	31	46	4	2	5	3	4	2

Table 2.8: [**For binary data with two outliers**] Simulated means (SM), simulated standard errors (SSE), and relative biases (RB) of the WMQL, SSMQL, FSMQL_1, and FSMQL_2 estimates of the regression parameters for different sample sizes, $\mu^{c_1} = 0.6$, and $\mu^{c_2} = 0.4$ under the binary model with $\beta_1 = 1.0$ and $\beta_2 = 0.5$ in the presence of two outliers.

		Estimation Method							
		WMQL		SSMQL		FSMQL_1		FSMQL_2	
K	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
20	SM	0.780	0.377	1.056	0.558	1.132	0.613	1.044	0.613
	SSE	1.617	1.065	1.534	1.057	1.641	1.111	1.539	1.111
	RB	14	12	4	6	8	10	3	6
30	SM	0.941	0.368	1.102	0.559	1.164	0.589	1.105	0.558
	SSE	1.379	1.172	1.330	1.130	1.402	1.213	1.328	1.132
	RB	4	11	8	5	12	7	8	5
60	SM	0.758	0.424	1.039	0.526	1.079	0.545	1.038	0.525
	SSE	1.145	0.605	1.061	0.571	1.098	0.592	1.062	0.572
	RB	21	13	4	5	7	8	4	4
100	SM	0.905	0.436	1.037	0.510	1.060	0.522	1.038	0.510
	SSE	0.885	0.509	0.819	0.463	0.838	0.471	0.820	0.463
	RB	11	13	5	2	7	5	5	2

Chapter 3

Robust Estimation for Familial Count and Binary Data

Generalized linear mixed models (GLMMs) are useful for accommodating the overdispersion and correlations observed among the outcomes in a given cluster/family. These models are generated from the well-known generalized linear model (GLM) by adding random effects to the linear predictor. Let y_{ij} be the response obtained from the jth individual ($j = 1, \ldots, n_i$) of the ith cluster/family ($i = 1, \ldots, K$) and

$$\eta_{ij} = x'_{ij}\beta + \gamma_i^*, \tag{3.1}$$

be the corresponding linear mixed predictor such that conditional on the random effect γ_i^*, the response variable y_{ij} follows an exponential distribution given by

$$f(y_{ij}|\gamma_i^*) = k\ exp\left[\{y_{ij}\eta_{ij} - a(\eta_{ij})\}\,\phi + b(y_{ij}, \phi)\right], \tag{3.2}$$

where k is the normalizing constant, $a(\cdot)$ and $b(\cdot)$ are the known functional forms, and ϕ is possibly an unknown scale parameter. In (3.1), $x_{ij} = (x_{ij}, \ldots, x_{ijr}, \ldots, x_{ijp})'$ is the $p-$dimensional vector of fixed covariates for the jth member of the ith family and β is the corresponding vector of the regression effects. Furthermore, γ_i^*'s are assumed to be identically and independently distributed as $N(0, \sigma^2)$. It is then clear that y_{ij} unconditionally follows an exponential-family based mixed model. Note that the exponential family (3.2) contains two important cases such Poisson and binary distributions, where $\phi = 1$. In this chapter , we deal with both of these special cases intensively and hence assume that either ϕ is known or $\phi = 1$.

Note that even when the clustered data do not contain any outliers, the exact likelihood estimation for the regression parameter β and the variance component of the random effects σ^2 is complicated. This is because the derivation of the unconditional likelihood function by taking the integration of the conditional probability function in (3.2) over the normal distribution of the random effects is quite complex. There exists various numerical techniques [e.g. McCulloch (1997))] those maximize the likelihood function numerically without deriving the exact form of the likelihood function. Note that these numerical techniques, however, can be computationally cumbersome in certain situations [Jiang (1998)]. Moreover, it is not clear how one can numerically approximate a likelihood function when the discrete clustered data are assumed to follow certain auto-correlation models. See, for example, the negative binomial auto-correlation model in Jowaheer and Sutradhar (2002, appendix, p. 398) [see also Bockenholt (1999)].

As opposed to the numerical techniques, there also exists alternative approximations such as the penalized quasi-likelihood (PQL) approach of Breslow and Clayton (1993) [see also, Schall (1991), Breslow and Lin (1995), Kuk (1995), Lin and Breslow (1996)] and the hierarchical likelihood (HL) approach of Lee and Nelder (1996). In the context of Poisson mixed model, Sutradhar and Qu (1998) [see also, Jiang (1998)] have shown that the PQL approach may not yield the consistent estimates for the parameters of the mixed model, particularly for the variance component of the random effects, when the cluster sizes are small, which is typically the case in practice. Jiang (1998) [see also, Jiang and Zhang (2001)] has suggested a simulated method of moments (SMM) that always produces the consistent estimates for both β and σ^2. These SMM based estimators, however, can be inefficient. Sutradhar (2004) [see also, Sutradhar and Rao (2001, 2003)] has proposed a generalized quasi-likelihood (GQL) approach which produces consistent as well as highly efficient estimates as compared to those obtained from the SMM approach.

Note that in practice, it may however happen that among a set of large number of independent clusters, a small number of clusters contain one or more outlying observations. For example, in count data analysis, it may happen that the bulk of the observations in the whole cluster set up follow the Poisson distribution with means close to each other, whereas a few outlying count observations in a small number of clusters may arise from the Poisson distribution but with inflated or deflated means due to the contaminated covariates. Similar situations may arise in the binary case where covariates can be contaminated leading to the unusual higher or lower success probabilities. In section 3.1, we provide the definitions of outliers for both Poisson and binary cases. Note that with regard to an outlier in the binary case, we have defined the outliers in a new way, which reflects the true nature of the outliers as in the Poisson and continuous such as Gaussian cases. In section 3.2, we examine the effects of outliers defined in section 3.1 on the GQL and moment estimations for the parameters of the Poisson and binary mixed models. This we have done as this recent GQL approach was found to produce consistent and highly efficient estimates in the GLMM set up in the absence of outliers.

With regard to the estimation of the parameters in the presence of outliers, Sinha

(2004) [see also, Mills et al. (2002)] has discussed a robust numerical technique in the cluster set up, which may be treated as a generalization of the numerical technique proposed by McCulloch (1997) in the absence of outliers. Note however that for the reasons indicated above with regard to the difficulties that one may encounter with the numerical techniques, we do not follow such numerical technique any further in the present chapter. Instead, we follow a robust Mallows-type quasi-likelihood (MQL) approach discussed by Cantoni and Ronchetti (2001) in the independence set up and develop a robust GQL (RGQL) approach as an improvement over the MQL approach in estimating the regression effects for the clustered data in the presence of outliers. The proposed robust approach may also be treated as a generalization of the GQL approach due to Sutradhar (2004) which was, however, developed for the clustered data in the absence of outliers. We further provide a robust moment (RM) approach for the estimation of the variance component of the random effects in the presence of outliers. The development of the RGQL and RM approaches is given in section 3.3. A simulation study is conducted in section 3.4 to examine the performances of the RGQL and RM approaches under both count and binary mixed models. This chapter concludes in section 3.5.

3.1 Outliers in the Discrete Clustered Data

It may happen in practice that the data may contain one or more outliers. For $i = 1, \ldots, K$, let $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})'$ denote the $n_i \times 1$ vector of count or binary responses. Suppose that m of these $\sum_{i=1}^{K} n_i$ responses are referred to as the mean shifted outliers when their corresponding covariates are shifted by an amount

δ, where δ is a vector. For convenience, we denote these new set of covariates as

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

Note that the definition of outliers depends on the discrete nature of the data. In the following subsections, we discuss how the outliers arise in both count and binary data.

3.1.1 Outliers in the count data

In the count data set up, an outlying count may arise either due to a shift in the mean or due to an inflated variance for an individual. For example, suppose that conditional on the random effect $\gamma_i = \frac{\gamma_i^*}{\sigma} \overset{i.i.d.}{\sim} N(0,1)$, in a situation when data do not contain any outliers, the response y_{ij} follows a Poisson distribution with mean $\mu_{ij}^* = E(Y_{ij}|\gamma_i, x_{ij}) = exp(x_{ij}'\beta - \frac{\sigma^2}{2} + \sigma\gamma_i)$ for $i = 1,\ldots,K,\ j = 1,\ldots,n_i$, where x_{ij} is the covariate for the jth individual of the ith family. Now, suppose that the covariate of one of the observations, say $(K, n_K)th$ observation is contaminated by an invisible amount δ, δ being a suitable vector. For convenience, we denote the observed covariates as

$$\tilde{x}_{ij} = \begin{cases} x_{ij} & \text{for } i = 1,\ldots,K, j = 1,\ldots,n_i, j \neq n_K \\ \\ x_{ij} + \delta & \text{for } i = K, j = n_K \end{cases},$$

and write the conditional mean as

$$\tilde{\mu}_{ij}^* = E(Y_{ij}|\gamma_i, \tilde{x}_{ij}) = \begin{cases} exp(x_{ij}'\beta - \frac{\sigma^2}{2} + \sigma\gamma_i) & \text{for } i = 1,\ldots,K, j = 1,\ldots,n_i, j \neq n_K \\ \\ exp[(x_{ij} + \delta)'\beta - \frac{\sigma^2}{2} + \sigma\gamma_i] & \text{for } i = K, j = n_K \end{cases},$$

Under this model, y_{Kn_K} is referred to as a mean shifted outlier.

Alternatively, suppose that for $i = 1, \ldots, K$, $j = 1, \ldots, n_i$, $j \neq n_K$, conditional on γ_i, y_{ij} follows a Poisson distribution with mean

$$\mu^*_{ij} = exp(x'_{ij}\beta - \frac{\sigma^2}{2} + \sigma\gamma_i),$$

but the $(K, n_K)th$ response arises from a negative binomial (NB) distribution with mean

$$\mu^*_{Kn_K} = exp(x'_{Kn_K}\beta - \frac{\sigma^2}{2} + \sigma\gamma_i),$$

and variance

$$var(Y_{Kn_K}|\gamma_i, x_{Kn_K}) = \mu^*_{Kn_K} + \alpha{\mu^*}^2_{Kn_K},$$

for a suitable scalar $\alpha > 0$. Since the variance of this observation, as compared to the other $\sum_{i=1}^{K} n_i - 1$ good observations, gets larger when α increases, we refer to this observation y_{Kn_K} as a variance inflated outlier.

In the thesis, we, however, consider the mean shifted outliers only, which received considerable attention in the literature [see for example, Cantoni and Ronchetti (2001), Sinha (2004)].

3.1.2 Outliers in the binary data

Note that unlike the count data case, the definition of outliers for the binary data is more complicated. Cantoni and Ronchetti (2001) considered the outliers in the binomial data that can be defined in the manner similar to that of the Poisson case. Similarly, for the binary case, Sinha (2004) used the same definition as that of the Poisson case. More specifically, for $(i', j') \not\equiv (i, j)$, $i = 1, \ldots, K$, $j = 1, \ldots, n_i$, suppose that $y_{i'j'} = 1$, but its corresponding probability (based on its covariate value) is small such that $P(Y_{i'j'} = 1|\gamma_i, \tilde{x}_{i'j'}) < 0.4$. Some authors such as Sinha (2004) have considered such a binary response to be an outlier. Similarly, $y_{i'j'} = 0$ has been referred to as an outlier if the corresponding probability $P(Y_{i'j'} = 1|\gamma_i, \tilde{x}_{i'j'})$ is large, say $P(Y_{i'j'} = 1|\gamma_i, \tilde{x}_{i'j'}) > 0.6$.

Note however that the above definition of an outlier for the binary data considered by Sinha (2004), among others, does not appear to interpret the real nature of an outlier. This is because an observation should be treated as an outlier when it is quite different from the bulk of the observations in the sample, either in magnitude or in the sense of probability. To be specific, suppose that in a sample of size $\sum_{i=1}^{K} n_i$, the covariate values of $\sum_{i=1}^{K} n_i - 1$ individuals lead to small probabilities such as $P(Y_{ij} = 1|\gamma_i, \tilde{x}_{ij}) \leq 0.3$ for all $(i, j) \not\equiv (i', j')$, $i = 1, \ldots, K$, $j = 1, \ldots, n_i$. In this case, if $P(Y_{i'j'} = 1|\gamma_i, \tilde{x}_{i'j'})$ is large, $y_{i'j'}$ will be an outlier irrespective of its value 0 or 1. This fact leads us to consider one sided or two sided outliers in the binary set up. These cases are explained below.

One sided outlier in binary mixed model set up

Suppose that the bulk of the binary observations (i.e. 'good' observations) occur with small probabilities. Let y_{ij} denote these observations for $(i, j) \not\equiv (i', j')$, $i = 1, \ldots, K$, $j = 1, \ldots, n_i$. In this case, the observation $y_{i'j'}$ will be defined as an outlier if the observed $P(Y_{i'j'} = 1|\gamma_i, \tilde{x}_{i'j'}) = \tilde{\mu}^*_{i'j'} > p^*_{sb}$, where $p^*_{sb} = max\{\tilde{\mu}^*_{ij}\}$, with

$$\tilde{\mu}^*_{ij} = P(Y_{ij} = 1|\gamma_i, \tilde{x}_{ij}) = \begin{cases} \frac{exp(x'_{ij}\beta+\sigma\gamma_i)}{1+exp(x'_{ij}\beta+\sigma\gamma_i)} & (i, j) \not\equiv (i', j') \\ \frac{exp[(x_{i'j'}+\delta)'\beta+\sigma\gamma_i]}{1+exp[(x_{i'j'}+\delta)'\beta+\sigma\gamma_i]} & \text{for } (i, j) \equiv (i', j') \end{cases}.$$

Note that $\tilde{\mu}^*_{i'j'}$ differs from all other $\tilde{\mu}^*_{ij}$ because of a contamination of the covariate by a real valued vector δ. Note that when $y_{i'j'}$ is generated with a large probability

$\mu^*_{i'j'}$, $y_{i'j'}$ can still be 0 or 1. Thus, $y_{i'j'}$, whether 1 or 0, must be treated as an outlier.

As opposed to the upper sided outlier case given above, one may also encounter the lower sided outlier case. This happens if the bulk of the binary observations (i.e. 'good' observations) occur with large probabilities, but the outlying observation $y_{i'j'}$ (whether 1 or 0) satisfies the observed $P(Y_{i'j'} = 1|\gamma_i, \tilde{x}_{i'j'}) = \tilde{\mu}^*_{i'j'} < p^*_{lb}$, where $p^*_{lb} = min\{\tilde{\mu}^*_{ij}\}$.

Two sided outlier in binary mixed model set up

It may also happen in practice that the probabilities for the bulk of the observations lie in the range $p^*_{lb} \leq P(Y_{ij} = 1|\gamma_i, \tilde{x}_{ij}) \leq p^*_{sb}$, where p^*_{lb} and p^*_{sb} are defined as above. This leads to a situation where one may encounter a two sided outlier. To be specific, $y_{i'j'} = 0$ or 1 will be referred to as a two sided outlier if either the observed $P(Y_{i'j'} = 1|\gamma_i, \tilde{x}_{i'j'}) > p^*_{sb}$ or the observed $P(Y_{i'j'} = 1|\gamma_i, \tilde{x}_{i'j'}) < p^*_{lb}$.

3.2 Effects of Outliers on the Unmodified GQL Estimation

It is well known that in the GLMM set up, the joint GQL estimation approach [Sutradhar (2004)] produces consistent and highly efficient estimates for the regression effects (β) and the variance component of the random effects (σ^2). Alternatively, for computational simplicity, one may use the GQL approach for the consistent and efficient estimation of β parameter, whereas σ^2 may be computed consistently by the method of moments (MM). Note that when data contain a single or more outliers, whether one uses the joint GQL or the combined GQL and MM approach, these approaches without any modifications may not be suitable for the estimation of parameters in the GLMM set up. In order to verify the effect of the presence of outliers on the combined GQL and MM approach, in this section, we mainly conduct a simulation study by generating the count or binary data with possible outliers (as defined in the last section), but estimating the main parameters β and σ^2 by using the unmodified combined GQL and MM approach.

For the purpose, we first demonstrate below how one can generate count and binary data with possible outliers. We do this following the definition of outliers provided in the last section.

3.2.1 Generating clustered data with outliers

Generating clustered count data with outliers

To generate a set of clustered count data with a few outliers, we consider $K = 100$, $p = 2$, and $n_i = 4$ $(i = 1, \dots, K)$, for example. Thus, all together $\sum_{i=1}^{K} n_i = 400$ responses will be generated. Out of these 400 responses, we choose to contaminate only $m = 4$ responses, where m denotes the number of outliers in the data. To do this, we first generate $\sum_{i=1}^{K} n_i = 400$ 'good' count responses in the absence of any outliers by following the Poisson mixed model as explained in section 3.1.1 with the conditional mean parameter as $\mu_{ij}^{*} = exp(x_{ij1}\beta_1 + x_{ij2}\beta_2 - \frac{\sigma^2}{2} + \sigma\gamma_i)$, where i denotes the ith cluster $(i = 1, \dots, K)$ with size n_i and $j = 1, \dots, n_i$. Here, γ_i's are generated from the normal distribution with mean 0 and variance 1. As the effects of these covariates, we consider two versions of the regression parameters, namely $\beta = (1.0, 1.0)'$ and $\beta = (1.5, 0.75)'$. Next, we convert $m = 4$ of these $\sum_{i=1}^{K} n_i = 400$ count responses as outlying responses following the definition given in section 3.1.1. To do this, we do not change the values of these 4 responses (generated as 'good' observations), rather we change their corresponding covariates. To be specific, we simply shift the first covariate with a real value δ_1 and the second covariate with another real value δ_2. That is, to treat the jth 'good' response in the ith cluster, for example, as an outlier, we consider $x_{ij1} + \delta_1$ and $x_{ij2} + \delta_2$ as the covariate values instead of x_{ij1} and x_{ij2}, respectively. Note that this shifting changes the mean level of this response as compared to the bulk of the responses, but the original responses were not changed. As far as the magnitude of shift for each covariate is concerned, we consider $\delta = (\delta_1, \delta_2)' = (4.0, 4.0)'$. Next, with regard to the variance component of the random effects, we choose $\sigma^2 = 0.25, 0.5, \text{and } 0.75$.

As far as the covariates are concerned, we consider a design D_1 with two covariates as given by

$$x_{ij1} = \begin{cases} 1.0 & \text{for } i = 1, \ldots, K/2;\ j \leq n_i/2 \\ 0.0 & \text{for } i = 1, \ldots, K/2;\ j > n_i/2 \\ 1.0 & \text{for } i = K/2+1, \ldots, K \end{cases},$$

and

$$x_{ij2} = \begin{cases} 1.0 & \text{for } i = 1, \ldots, K/2;\ j \leq n_i/2 \\ 1.5 & \text{for } i = 1, \ldots, K/2;\ j > n_i/2 \\ -1.0 & \text{for } i = K/2+1, \ldots, K;\ j \leq n_i/2 \\ 0.0 & \text{for } i = K/2+1, \ldots, K;\ j > n_i/2 \end{cases},$$

respectively.

Generating clustered binary data with outliers

For simplicity, here, we consider the generation of one sided outliers. The two sided outliers may be generated similarly.

To generate $m = 4$ binary responses as one sided outliers out of $\sum_{i=1}^{n_i} = 400$ responses, we first generate 400 'good' responses from the binary mixed model with conditional mean $\mu_{ij}^* = \frac{exp(x_{ij1}\beta_1 + x_{ij2}\beta_2 + \sigma\gamma_i)}{1+exp(x_{ij1}\beta_1 + x_{ij2}\beta_2 + \sigma\gamma_i)}$, $i = 1, \ldots, K$, $j = 1, \ldots, n_i$, where γ_i's are independently and identically normally distributed with mean 0 and variance 1. Under this simulation study, we consider two different designs D_2 and D_3, each containing two covariates.

Under the design D_2, two covariates are selected as

$$x_{ij1} = \begin{cases} -2.0 & \text{for } i = 1, \ldots, K/2 \\ -1.0 & \text{for } i = K/2+1, \ldots, 3K/4 \\ -1.5 & \text{for } i = 3K/4+1, \ldots, K \end{cases},$$

and

$$x_{ij2} \sim N(-0.75, 0.25).$$

For the selection of the regression parameter values, we consider two sets of values, namely $\beta = (1.0, 1.0)'$ and $\beta = (1.5, 0.5)'$. As far as the values of K, m, n_i, and σ^2 are concerned, we use the same values as those of the count case in the last subsection.

Under the second design D_3, the two covariates are chosen as

$$x_{ij1} = \begin{cases} 1.0 & \text{for } i = 1, \ldots, K/2;\ j \le n_i/2 \\ 0.5 & \text{for } i = 1, \ldots, K/2;\ j > n_i/2 \\ 1.0 & \text{for } i = K/2+1, \ldots, K \end{cases},$$

and

$$x_{ij2} = \begin{cases} 1.0 & \text{for } i = 1, \ldots, K/2;\ j \leq n_i/2 \\ 1.25 & \text{for } i = 1, \ldots, K/2;\ j > n_i/2 \\ 0.75 & \text{for } i = K/2+1, \ldots, K;\ j \leq n_i/2 \\ 2.0 & \text{for } i = K/2+1, \ldots, K;\ j > n_i/2 \end{cases},$$

respectively. As far as the regression effects are concerned, unlike under design D_2, in this case, we consider two sets of regression effects as $\beta = (-1.0, -1.0)'$ and $\beta = (-0.5, -1.5)'$. Note that in both cases, under designs D_2 and D_3, the true values of the regression parameters are chosen such that μ^*_{ij} for i and j, are small. To be specific, by generating $K = 100$ γ_i from $N(0,1)$, when μ^*_{ij} were computed by using the selected covariates and the true regression parameter values, the probabilities μ^*_{ij} for all 'good' binary responses were found to lie, in general, between 0.007 and 0.395 with specific ranges as follows:

$D_2 : \beta_1 = \beta_2 = 1.0$		$\delta_1 = 2.0, \delta_2 = 3.0$
σ^2	Range of μ^*_{ij}	$\tilde{\mu}^*_{i'j'}$ values based on contaminated covariates
0.25	0.021-0.372	0.938, 0.872, 0.966, 0.932
0.50	0.013-0.390	0.949, 0.866, 0.970, 0.932
0.75	0.010-0.395	0.956, 0.860, 0.973, 0.932

$D_2 : \beta_1 = 1.5,\ \beta_2 = 0.5$		$\delta_1 = 2.0,\ \delta_2 = 3.0$
σ^2		
0.25	0.016-0.373	0.833, 0.709, 0.944, 0.859
0.50	0.012-0.380	0.859, 0.696, 0.950, 0.858
0.75	0.008-0.395	0.877, 0.686, 0.955, 0.858
$D_3 : \beta_1 = \beta_2 = -1.0$		$\delta_1 = -1.0,\ \delta_2 = -3.0$
σ^2		
0.25	0.021-0.346	0.750, 0.950, 0.832, 0.879
0.50	0.015-0.381	0.674, 0.962, 0.791, 0.867
0.75	0.011-0.393	0.608, 0.969, 0.754, 0.857
$D_3 : \beta_1 = -0.5,\ \beta_2 = -1.5$		$\delta_1 = -1.0,\ \delta_2 = -3.0$
σ^2		
0.25	0.013-0.356	0.891, 0.973, 0.938, 0.957
0.50	0.009-0.367	0.849, 0.979, 0.921, 0.953
0.75	0.007-0.381	0.809, 0.983, 0.905, 0.949

To treat $m = 4$ of the $\sum_{i=1}^{K} n_i = 400$ 'good' binary observations as one sided outliers, the response values, whether 0 or 1, are not changed, but we use $x_{ij1} + \delta_1$ and $x_{ij2} + \delta_2$ as the covariate values instead of x_{ij1} and x_{ij2}, respectively. For the values of δ_1 and δ_2, we choose $\delta = (\delta_1, \delta_2)' = (2.0, 3.0)'$ and $\delta = (-1.0, -3.0)'$ under designs D_2 and D_3, respectively. Now, by using the contaminated covariates, the probabilities, namely $\tilde{\mu}^*_{i'j'}$ were made large as shown above in the tabular form. It is clear that these values are quite large beyond the upper limit of μ^*_{ij} for the 'good' observations.

3.2.2 Unmodified GQL estimation for clustered data in the presence of outliers

Since y_{ij}'s are generated with 'good' covariates x_{ij}, it is clear that if the outliers are not accounted for, one may then estimate the parameters of model by exploiting y_{ij} and $\tilde{x}_{ij}$. Thus, when the outliers are neglected, one may follow Sutradhar (2004), for example, and estimate the regression effects β by solving the GQL estimating equation given by

$$\sum_{i=1}^{K}\left[\frac{\partial \tilde{\mu}_i'}{\partial \beta}\tilde{\Sigma}_i^{-1}(y_i - \tilde{\mu}_i)\right] = 0, \tag{3.3}$$

where $\tilde{\mu}_i = (\tilde{\mu}_{i1}, \ldots, \tilde{\mu}_{ij}, \ldots, \tilde{\mu}_{in_i})' = E(Y_i|\tilde{X}_i)$, $\tilde{\Sigma}_i = (\tilde{\sigma}_{ijk}) = cov(Y_i|\tilde{X}_i)$, and $\frac{\partial \tilde{\mu}_i'}{\partial \beta}$ is the $p \times n_i$ derivative matrix of mean vector $\tilde{\mu}_i$ with respect to β with $\tilde{X}_i = [\tilde{x}_{i1}, \ldots, \tilde{x}_{ij}, \ldots, \tilde{x}_{in_i}]'$. Note that for $\gamma_i = \frac{\gamma_i^*}{\sigma} \overset{i.i.d.}{\sim} N(0,1)$, the jth element of $\tilde{\mu}_i$ and the $(j,k)th$ element of $\tilde{\Sigma}_i$ for the count data have the formulas given by

$$\begin{aligned} \tilde{\mu}_{ij} &= E(Y_{ij}|\tilde{x}_{ij}) = E_{\gamma_i}E(Y_{ij}|\gamma_i, \tilde{x}_{ij}) = E_{\gamma_i}(\tilde{\mu}_{ij}^*) \\ &= E_{\gamma_i}[exp(\tilde{x}_{ij}'\beta - \frac{\sigma^2}{2} + \sigma\gamma_i)] = exp(\tilde{x}_{ij}'\beta), \end{aligned} \tag{3.4}$$

and

$$\tilde{\sigma}_{ijk} = \begin{cases} \tilde{\mu}_{ij} + c\tilde{\mu}_{ij}^2 & \text{for } j = k, \\ c\tilde{\mu}_{ij}\tilde{\mu}_{ik} & \text{for } j \neq k, \end{cases} \tag{3.5}$$

respectively, where c is the overdispersion parameter defined as $c = exp(\sigma^2) - 1$. The formula for $\tilde{\sigma}_{ijj}$ in (3.5) is derived from $\tilde{\sigma}_{ijj} = var(Y_{ij}|\tilde{x}_{ij}) = var_{\gamma_i}[E(Y_{ij}|\gamma_i, \tilde{x}_{ij})] + E_{\gamma_i}[var(Y_{ij}|\gamma_i, \tilde{x}_{ij})]$ and similarly, the formula for $\tilde{\sigma}_{ijk}$ is derived from $\tilde{\sigma}_{ijk} = cov(Y_{ij}, Y_{ik}| \tilde{x}_{ij}, \tilde{x}_{ik}) = cov_{\gamma_i}[E(Y_{ij}|\gamma_i, \tilde{x}_{ij}), E(Y_{ik}|\gamma_i, \tilde{x}_{ik})] + E_{\gamma_i}[cov(Y_{ij}, Y_{ik}|\gamma_i, \tilde{x}_{ij}, \tilde{x}_{ik})]$, $j \neq k$. Furthermore, for a large integer M such as $M = 5000$, $\tilde{\mu}_{ij}$ and $\tilde{\sigma}_{ijk}$ for the binary case have the formulas as

$$\begin{aligned}\tilde{\mu}_{ij} &= E(Y_{ij}|\tilde{x}_{ij}) = E_{\gamma_i}E(Y_{ij}|\gamma_i, \tilde{x}_{ij}) \\ &= E_{\gamma_i}(\tilde{\mu}^*_{ij}) = \frac{1}{M}\sum_{\ell=1}^{M}\tilde{\mu}^*_{ij,\ell},\end{aligned} \tag{3.6}$$

and

$$\tilde{\sigma}_{ijk} = \begin{cases} \tilde{\mu}_{ij}(1-\tilde{\mu}_{ij}) & \text{for } j = k, \\ \\ \frac{1}{M}\sum_{\ell=1}^{M}\tilde{\mu}^*_{ij,\ell}\tilde{\mu}^*_{ik,\ell} - \tilde{\mu}_{ij}\tilde{\mu}_{ik} & \text{for } j \neq k, \end{cases} \tag{3.7}$$

respectively, where $\tilde{\mu}^*_{ij,\ell} = \frac{exp(\tilde{x}'_{ij}\beta+\gamma_{i\ell}\sigma)}{1+exp(\tilde{x}'_{ij}\beta+\gamma_{i\ell}\sigma)}$ is obtained from $\tilde{\mu}^*_{ij} = E(Y_{ij}|\gamma_i, \tilde{x}_{ij}) = \frac{exp(\tilde{x}'_{ij}\beta+\gamma_i\sigma)}{1+exp(\tilde{x}'_{ij}\beta+\gamma_i\sigma)}$ by using $\gamma_i = \gamma_{i\ell}$, $\gamma_{i\ell}$ being the ℓth ($\ell = 1, \ldots, M$) simulated value of $\gamma_i \sim N(0,1)$. To be specific, the formula for $\tilde{\mu}_{ij}$ is obtained by solving the integral $\tilde{\mu}_{ij} = \int_{-\infty}^{\infty}\tilde{\mu}^*_{ij}\phi(\gamma_i)\partial\gamma_i$ [Sutradhar (2004) and Fahrmeir and Tutz (1994)], where $\phi(\cdot)$ denotes the standard normal distribution. Similarly, $E(Y_{ij}Y_{ik}|\tilde{x}_{ij}, \tilde{x}_{ik})$ is obtained as

$$E(Y_{ij}Y_{ik}|\tilde{x}_{ij}, \tilde{x}_{ik}) = \int_{-\infty}^{\infty}\tilde{\mu}^*_{ij}\tilde{\mu}^*_{ik}\phi(\gamma_i)\partial\gamma_i = \tilde{\mu}_{ijk}, \text{ say,} \tag{3.8}$$

leading to the covariance $\tilde{\sigma}_{ijk}$ as in (3.7).

Note that when outliers are avoided, one may estimate the variance of random effects σ^2 by solving the moment estimating equation

$$S - E(S) = 0, \tag{3.9}$$

where

$$S = \sum_{i=1}^{K}\sum_{j=1}^{n_i} y_{ij}^2 + \sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i} y_{ij}y_{ik}, \tag{3.10}$$

is a statistic written by combining all possible squares and distinct pairwise products of the responses [see, Sutradhar (2004) and Jiang (1998)] in the GLMM set up. As far as the formula for $E(S)$ is concerned, one may derive it easily by using the variances and covariances given in (3.5) or (3.7). To be specific, for the count data

$$E(S) = \sum_{i=1}^{K}\sum_{j=1}^{n_i}[\tilde{\sigma}_{ijj} + \tilde{\mu}_{ij}^2] + \sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}[\tilde{\sigma}_{ijk} + \tilde{\mu}_{ij}\tilde{\mu}_{ik}], \tag{3.11}$$

where $\tilde{\mu}_{ij}$ and $\tilde{\sigma}_{ijj}$, $\tilde{\sigma}_{ijk}$ are given as in (3.4) and (3.5), respectively. Now, by using S from (3.10) and $E(S)$ from (3.11), one may solve (3.9) and obtain a closed form moment estimator of σ^2 as given by

$$\hat{\sigma}^2 = log_e \left[1 + \frac{\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij}-\hat{\mu}_{ij})^2}{D_1} + \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}(y_{ij}-\hat{\mu}_{ij})(y_{ik}-\hat{\mu}_{ik})}{D_2} - \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\hat{\mu}_{ij}}{D_1}}{\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\hat{\mu}_{ij}^2}{D_1} + \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\hat{\mu}_{ij}\hat{\mu}_{ik}}{D_2}} \right], \tag{3.12}$$

where $D_1 = \sum_{i=1}^{K} n_i$ and $D_2 = \sum_{i=1}^{K} n_i(n_i-1)/2$. Similarly, under the binary set up, one may compute $E(S)$ as

$$E(S) = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\tilde{\mu}_{ij} + \sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\tilde{\mu}_{ijk}, \tag{3.13}$$

with $\tilde{\mu}_{ij}$ and $\tilde{\mu}_{ijk}$ as given in (3.6) and (3.8), respectively. Note that unlike (3.12) for the Poisson case, there is no explicit formula for the estimator of σ^2 in the binary case. More specifically, one may obtain the estimate of σ^2 by solving $S - E(S) = 0$ iteratively, where S and $E(S)$ are given as in (3.10) and (3.13), respectively. For example, given the value $\hat{\sigma}^2_{(r)}$ as an estimate of σ^2 at the rth iteration, the estimate at the $(r+1)st$ iteration will be obtained as

$$\hat{\sigma}^2_{(r+1)} = \hat{\sigma}^2_{(r)} + \left[\frac{\partial}{\partial\sigma^2}E(S)\right]_r^{-1}[S - E(S)]_r\,, \tag{3.14}$$

where $[\cdot]_r$ denotes the fact that the expression in the brackets is evaluated at $\sigma^2 = \hat{\sigma}^2_{(r)}$.

Let $\hat{\beta}$ and $\hat{\sigma}^2$ denote the GQL and moment estimators for β and σ^2, respectively, irrespective of the nature of the data whether count or binary, obtained from the GQL estimating equation (3.3) and the moment estimating equation (3.9). Note that in absence of outliers, the GQL estimating equation (3.3) becomes unbiased and hence produces consistent as well as highly efficient estimate for β. Similarly, when data do not contain any outliers, one may obtain consistent estimate for σ^2 by solving (3.9).

3.2.3 Performance of the unmodified GQL estimators: A simulation study

Count data case

To examine the effects of outliers on the unmodified GQL estimating equation (3.3) and moment estimating equation (3.9), we simulate the count responses with $m = 4$ outliers under design D_1, 500 times, following the procedure described in section 3.2.1. Under each simulation, we then record the estimates of β_1 and β_2 as a solution of (3.3), and the estimate of σ^2 computed by (3.12). The simulated mean (SM), simulated standard error (SSE), and mean squared error (MSE) computed from 500 values for each of these three estimates are reported in Table 3.1.

It is clear from this table that the GQL estimate of β and the MM estimate of σ^2 are not at all satisfactory. For example, for the true value $\beta_1 = \beta_2 = 1.0$ and $\sigma^2 = 0.5$, the results in Table 3.1 show that the GQL estimates for β_1 and β_2 are: 0.656 and 0.051, respectively with corresponding standard errors 1.062 and 0.721. Thus, the regression estimates are highly biased and hence, they are inconsistent. Similarly, the MM estimate for $\sigma^2 = 0.5$ is found to be 1.371 with standard error 0.594, which is again highly biased. Similar results hold for other selection of the parameter values. Thus, the unmodified GQL and MM estimators are inconsistent for the regression and the variance component parameters, respectively, when the data really contain outliers.

Binary data case

In this case, we simulate $\sum_{i=1}^{K} n_i = 400$ binary responses with $m = 4$ outliers as described in section 3.2.1 and apply the unmodified GQL estimating equation (3.3) and moment estimating equation (3.9) to compute the estimates for β_1, β_2, and σ^2, respectively. Since the convergence is quicker in the binary case, especially when robust estimation in section 3.3 is concerned, we use 1000 simulations under the binary mixed model, whereas 500 simulations are used under the Poisson mixed model. The simulation results with SM, SSE, and MSE under the designs D_2 and D_3 are given in Tables 3.2 and 3.3, respectively.

Once again, similar to those of Table 3.1, it is found from Tables 3.2 and 3.3 that the GQL and MM approaches produce inconsistent estimates for the parameters when the data contain outliers. For example, Table 3.2 reveals that under the binary mixed model, for $\beta_1 = 1.5$, $\beta_2 = 0.5$, and $\sigma^2 = 0.25$, the GQL approach produces estimates for β_1 and β_2 as 2.000 and -0.535, respectively with corresponding standard errors 0.259 and 0.336. These estimators are definitely highly biased. Similarly, the MM approach produces estimate for σ^2 as 0.076 with standard error 0.108, whereas the true value of σ^2 is 0.25. Thus, the unmodified moment approach grossly underestimates σ^2 and hence this estimator is also inconsistent. When D_3 is used, the simulation results in Table 3.3 show that for the true value $\beta_1 = \beta_2 = -1.0$ and $\sigma^2 = 0.75$, the simulated unmodified GQL estimates for β_1 and β_2 are given by -1.407 and -0.501 with corresponding standard errors 0.361 and 0.217. Similarly, the unmodified MM estimate for $\sigma^2 = 0.75$ is found to be 0.321 with standard error 0.322. These and other results in Table 3.2 and 3.3 clearly demonstrate that the unmodified GQL and MM approaches produce highly biased and hence inconsistent estimates for the parameters of the binary mixed model containing one or more outliers. Similar results were also found for the Poisson mixed model with outliers.

3.3 Robust GQL Estimation

It has been demonstrated in the last section through a simulation study that the unmodified GQL estimating equation (3.3) for the regression effect β and the moment estimating equation (3.9) for the variance component σ^2 do not provide consistent estimates, when the clustered data contain one or more outliers. In fact, in a non-clustered regression set up, that is when $\sigma^2 = 0$, the unmodified QL estimates of the regression parameters in the presence of outliers are also known to be inconsistent. See, for example, Cantoni and Ronchetti (2001), where as a remedy to the inconsistency problem, the traditional QL approach has been modified through the introduction of a robust Mallows-type quasi-likelihood (MQL) estimation. By the same token, as a modification of the GQL approach, we now introduce a robust GQL (RGQL) approach to obtain consistent estimates for both β and σ^2 under the clustered regression set up in the presence of outliers. Note that the proposed RGQL approach may be treated as a generalization of the MQL approach of Cantoni and Ronchetti (2001) to the clustered regression set up. We also remark here that the MQL estimating equations are not in so-called standardized forms, whereas in the present RGQL approach, we propose to use estimating equations in standardized forms that lead to higher consistency and efficiency. Furthermore, Cantoni and Ronchetti (2001) have dealt with independent count and binomial data with outliers, whereas we deal with clustered count and binary data with outliers.

As an aid to the construction of the proposed fully standardized RGQL estimating equations in the clustered set up, we first provide the formulas for the downweighting functions for the count and binary cases. These downweighting functions are then used to take the outliers into account as opposed to ignoring them completely, in estimating the main parameters of the model consistently.

3.3.1 Downweighting functions

Count data case

For $i = 1, \ldots, K$ and $j = 1, \ldots, n_i$, let $r_{ij} = \frac{y_{ij} - \tilde{\mu}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}}$, where y_{ij} is the count response of the jth member in the ith family; and $\tilde{\mu}_{ij}$ and $\tilde{\sigma}_{ijj}$ are the mean and variance of y_{ij} defined as in (3.4) and (3.5). Also, let $\psi_c(r_{ij})$ denote the downweighting function that helps to minimize the influence of the suspected outlying observations in the estimation without removing them completely. We now follow Huber (2004) [see, also Cantoni and Ronchetti (2001) and Sinha (2004)] and define the downweighting function for the Poisson case as

$$\psi_c(r_{ij}) = \begin{cases} r_{ij}, & |r_{ij}| \leq c, \\ c \ \mathrm{sign}(r_{ij}), & |r_{ij}| > c, \end{cases} \tag{3.15}$$

where c is the well-known tuning constant.

Binary data case

When the binary data contain outliers, as explained in section 3.1.2, it appears to be appropriate to define two types of outliers, namely one sided and two sided outliers. The downweighting functions for these outliers are defined as follows.

(a) One sided outlier

Recall that for the binary response y_{ij}, which may or may not be an outlier, the mean $\tilde{\mu}_{ij}$ and variance $\tilde{\sigma}_{ijj}$ are defined as in (3.6) and (3.7), respectively. For convenience, we denote the outlying observation by $y_{i'j'}$. When the bulk of the binary observations occur with small success probabilities, the downweighting function $\psi_c(r_{ij})$ $(i = 1, \ldots, K,\ j = 1, \ldots, n_i)$ may be defined as

$$\psi_c(r_{ij}) = \begin{cases} \frac{y_{ij}-\tilde{\mu}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}}, & \text{for } P(Y_{ij} = 1|\tilde{x}_{ij}) \leq p_{sb}, (i,j) \not\equiv (i',j'), \\ \frac{y_{ij}-\mu_{ij}^{(c_1)}}{\sqrt{\sigma_{ijj}^{(c_1)}}}, & \text{for } P(Y_{ij} = 1|\tilde{x}_{ij}) > p_{sb}, (i,j) \equiv (i',j'), \end{cases} \tag{3.16}$$

where following section 3.1.2, p_{sb} is defined as a critical probability which is the largest probability among the probabilities corresponding to the so-called 'good' observations. That is, $p_{sb} = max\{\tilde{\mu}_{ij}\}$, $(i,j) \not\equiv (i',j')$. Furthermore, in (3.16), $\mu_{i'j'}^{(c_1)}$ denotes an appropriate tuning constant related probability that replaces the probability $\tilde{\mu}_{i'j'}$ for the outlying observation in order to reduce its influence. For simplicity, this probability is referred to as the tuning probability and will be denoted by μ^{c_1} without any loss of generality. By the same token, the variance of the outlying observation $\tilde{\sigma}_{i'j'j'}$ should be replace by $\sigma_{i'j'j'}^{(c_1)}$, where $\sigma_{i'j'j'}^{(c_1)} = \mu_{i'j'}^{(c_1)}(1 - \mu_{i'j'}^{(c_1)})$.

Now, by using the similar argument as in the count data case, one may choose the tuning probability $\mu_{i'j'}^{(c_1)}$ so that it is nearer to p_{sb} such as 0.6, but less than $\tilde{\mu}_{i'j'}$ (the outlying probability). Note that as p_{sb} depends on β, one may choose $p_{sb} = 0.4$ initially provided the data contain more zero's than one's. To reflect this initial situation of $p_{sb} = 0.4$, we start with a suitable initial value of β so that $\tilde{\mu}_{ij}$'s are small. Once a first step estimate of β is obtained, we then compute p_{sb} by using the given formula $p_{sb} = max\{\tilde{\mu}_{ij}\}$, $(i,j) \not\equiv (i',j')$.

Further note that if the bulk of the binary observations however occur with large success probabilities, the downweighting function $\psi_c(r_{ij})$ $(i = 1, \ldots, K, j = 1, \ldots, n_i)$ for this case may be written by making slight modification to the definition given in (3.16). Thus, in this case, $\psi_c(r_{ij})$ is defined as

$$\psi_c(r_{ij}) = \begin{cases} \frac{y_{ij} - \tilde{\mu}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}}, & P(Y_{ij} = 1|\tilde{x}_{ij}) \geq p_{lb}, (i,j) \not\equiv (i',j'), \\ \frac{y_{ij} - \mu_{ij}^{(c_2)}}{\sqrt{\sigma_{ijj}^{(c_2)}}}, & P(Y_{ij} = 1|\tilde{x}_{ij}) < p_{lb}, (i,j) \equiv (i',j'), \end{cases} \tag{3.17}$$

where unlike in (3.16), $p_{lb} = min\{\tilde{\mu}_{ij}\}$, $(i,j) \not\equiv (i',j')$. Also, $\mu_{i'j'}^{(c_2)}$ is an appropriate tuning constant related probability similar to that of $\mu_{i'j'}^{(c_1)}$ in (3.16), and $\sigma_{i'j'j'}^{(c_2)} = \mu_{i'j'}^{(c_2)}(1 - \mu_{i'j'}^{(c_2)})$.

Note that to select a tuning constant value for $\mu_{i'j'}^{(c_2)}$, one may consider a value nearer to p_{lb}, but greater than $\tilde{\mu}_{i'j'}$ (the outlying probability). For example, the tuning probability $\mu_{i'j'}^{(c_2)}$, i.e., μ^{c_2} in brief, may be closer to be 0.4. To reveal the scenario that the data contain more one's than zero's, one may choose $p_{lb} = 0.6$ initially and start with a suitable value of β so that $\tilde{\mu}_{ij}$'s are large. After getting the first step estimate for β, we then turn back to the formula $p_{lb} = min\{\tilde{\mu}_{ij}\}$, $(i,j) \not\equiv (i',j')$ to compute p_{lb}.

(b) Two sided outlier

Note that in practice it may happen that the success probabilities for the bulk of the observations lie within a range of probabilities. Suppose that this range is given by $p_{lb} \leq P(Y_{ij} = 1|\tilde{x}_{ij}) \leq p_{sb}$. Now, one can combine two one-sided, that is, lower and upper sided outlier cases to define the downweighting function $\psi_c(r_{ij})$ $(i = 1, \ldots, K,$

$j = 1, \ldots, n_i$) for the present two sided outlier case. To be specific,

$$\psi_c(r_{ij}) = \begin{cases} \frac{y_{ij}-\mu_{ij}^{(c_1)}}{\sqrt{\sigma_{ijj}^{(c_1)}}}, & P(Y_{ij}=1|\tilde{x}_{ij}) > p_{sb}, (i,j) \equiv (i',j'), \\ \frac{y_{ij}-\tilde{\mu}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}}, & p_{lb} \leq P(Y_{ij}=1|\tilde{x}_{ij}) \leq p_{sb}, (i,j) \not\equiv (i',j'), \\ \frac{y_{ij}-\mu_{ij}^{(c_2)}}{\sqrt{\sigma_{ijj}^{(c_2)}}}, & P(Y_{ij}=1|\tilde{x}_{ij}) < p_{lb}, (i,j) \equiv (i',j'), \end{cases} \tag{3.18}$$

where $\mu_{i'j'}^{(c_1)}$ and $\sigma_{i'j'j'}^{(c_1)}$ are defined as in (3.16), whereas $\mu_{i'j'}^{(c_2)}$ and $\sigma_{i'j'j'}^{(c_2)}$ are defined as in (3.17). In general, p_{lb} and p_{sb} are considered to be 0.4 and 0.6, respectively. Further note that the definition of outliers for the binary data that appears in the existing literature [see, for example, Sinha (2004)] is very similar to the definition of our two sided outlier case only. Thus, one sided outlier concept is relatively new, and only two sided outlier case is not sufficient to interpret the outliers in the binary case. Moreover, if our two sided case is considered to be sufficient to define the outliers in the binary data, there is a difference between this two sided based outlier definition and the definition used in the literature. This is because existing definition for the outliers uses the tuning constant c directly, whereas we have used tuning constant based related probabilities to define the downweighting functions.

3.3.2 Robust GQL estimating equation for β

Recall from section 3.2 that the unmodified GQL approach yields highly biased and hence inconsistent estimates for the regression effects. As a remedy, by using the downweighting, i.e., so-called robust functions defined in the last subsection, we now provide a modification to the traditional GQL approach and refer to this as the robust GQL (RGQL) approach.

Let

$$\xi_i = [\psi_c(r_{i1}), \ldots, \psi_c(r_{ij}), \ldots, \psi_c(r_{in_i})]'$$

be the $n_i \times 1$ vector of downweighting or robust functions for the *ith* family. Also, let

$$\lambda_i = E(\xi_i) \text{ and } \Omega_i = cov(\xi_i),$$

be the expectation vector and covariance matrix for ξ_i, respectively. Now, by replacing the observation vector, its mean, and covariance matrix in (3.3) with the corresponding vector of robust functions, their mean and covariance matrix, we propose to use the RGQL estimating equation given by

$$\sum_{i=1}^{K} \left[W_i \frac{\partial}{\partial \beta} \left\{ \xi_i - K^{-1} \sum_{i=1}^{K} \lambda_i \right\}' \Omega_i^{-1} \left\{ \xi_i - K^{-1} \sum_{i=1}^{K} \lambda_i \right\} \right] = 0, \tag{3.19}$$

as a modification to the GQL estimating equation (3.3). In (3.19), $W_i = diag[w_{i1}, \ldots, w_{ij}, \ldots, w_{in_i}]$ is the $n_i \times n_i$ 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 *jth* diagonal element of the W_i matrix is computed as $w_{ij} = \sqrt{1 - h_{ijj}}$, h_{ijj} being the *jth* diagonal element of the hat matrix $H_i = \tilde{X}_i(\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i'$ with $\tilde{X}_i = [\tilde{x}_{i1}, \ldots, \tilde{x}_{ij}, \ldots, \tilde{x}_{in_i}]'$. The derivations for λ_i, $\frac{\partial}{\partial \beta}\xi_i'$, $\frac{\partial}{\partial \beta}\lambda_i'$, and Ω_i both for count and binary cases are given in the appendix.

Let $\hat{\beta}_{RGQL}$ be the solution of (3.19). Note that as opposed to the RGQL estimating equation (3.19), for $n_i = 1$ and $\sigma^2 = 0$, Cantoni and Ronchetti (2001) have considered a MQL estimating equation. In this MQL approach, the authors have used $\tilde{\Sigma}_i = cov(Y_i|\tilde{X}_i)$ as the weight matrix instead of the correct weight matrix Ω_i. Furthermore, the MQL approach used $\frac{\partial}{\partial \beta}\tilde{\mu}_i'$ as the gradient function instead of the true gradient function $\frac{\partial}{\partial \beta}\left\{\xi_i - K^{-1}\sum_{i=1}^{K}\lambda_i\right\}'$. Thus, it is clear that the present fully standardized GQL estimating equation (3.19) provides an improvement in the sense of consistency and efficiency. Further note that under certain regularity conditions, it may be shown that the equation (3.19) is asymptotically (as $K \to \infty$) an unbiased estimating equation. Consequently, in the fashion similar to that of chapter 2, one may study the asymptotic properties of the RGQL estimators of the regression effects obtained from (3.19), which is however not discussed in this chapter. As the small sample properties of the estimators are rather important for practical purpose, we conduct a simulation study in section 3.4 to examine the performance of the RGQL approach based regression estimators for both count and binary data. It should be mentioned that Cantoni and Ronchetti (2001) neither provided any such simulation results, nor they considered the binary case. Before we consider the simulation study in section 3.4, we now need to develop an estimation formula for the σ^2 parameter. This we do in the following subsection.

3.3.3 Robust moment estimation for σ^2

Recall that when the data contain one or more outliers, the unmodified moment estimating equation (3.9) does not produce consistent estimator for σ^2, the variance component of the random effects. In the manner similar to that of the estimating equation (3.19) for β, we now provide a robust moment (RM) estimating equation approach to obtain a consistent estimator for σ^2. For the purpose, similar to (3.19), we replace y_{ij} in (3.10) by $\psi_c(r_{ij})$ and write a preliminary statistic

$$S_\psi = \sum_{i=1}^{K}\sum_{j=1}^{n_i} \psi_c^2(r_{ij}) + \sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i} \psi_c(r_{ij})\psi_c(r_{ik}), \tag{3.20}$$

as a replacement for S in (3.10). Note that in the construction of the S statistic in (3.10), it was not needed to consider $\sum_{i=1}^{K}\sum_{j=1}^{n_i} y_{ij}$ for the following reasons. First, in the binary case, $y_{ij} \equiv y_{ij}^2$. Secondly, in the Poisson case, $E(Y_{ij}|\tilde{x}_{ij})$ (3.4) does not contain σ^2. In the contrary, $\psi_c(r_{ij}) \neq \psi_c(r_{ij})^2$ in the binary case. Furthermore, $E[\psi_c(r_{ij})]$ in the Poisson case contains σ^2. Consequently, we improve the S_ψ statistic in (3.20) by adding the information $\sum_{i=1}^{K}\sum_{j=1}^{n_i}\psi_c(r_{ij})$ and write this improved statistic as

$$\varphi = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\psi_c(r_{ij}) + S_\psi. \tag{3.21}$$

Now, for known β, the RM estimating equation for σ^2 can be written as

$$g(\sigma^2) = \varphi - E(\varphi) = 0, \tag{3.22}$$

where

$$E(\varphi) = \frac{1}{M}\left[\sum_{\ell=1}^{M}\sum_{i=1}^{K}\sum_{j=1}^{n_i} a^*_{ij,\ell} + \sum_{\ell=1}^{M}\sum_{i=1}^{K}\sum_{j=1}^{n_i} b^*_{ij,\ell} + \sum_{\ell=1}^{M}\sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i} a^*_{ij,\ell}a^*_{ik,\ell}\right], \tag{3.23}$$

where the formulas for $a^*_{ij,\ell}$ and $b^*_{ij,\ell}$ are given in the equations (A.3) and (A.10) in the Appendix A for the count data; and in the equations (A.16) and (A.21) in the Appendix A for the binary data, respectively. The estimating equation (3.22) can be solved for σ^2 by using the Newton-Raphson iterative procedure. Let $\hat{\sigma}^2_{(r+1)}$ be the solution for σ^2 at the $(r+1)st$ iteration, which can be obtained as

$$\hat{\sigma}^2_{(r+1)} = \hat{\sigma}^2_{(r)} - \left[\frac{g(\sigma^2)}{\frac{\partial}{\partial\sigma^2}g(\sigma^2)}\right]_{(r)}, \tag{3.24}$$

where $[\cdot]_{(r)}$ denotes the fact that the expression in the bracket is evaluated at $\sigma^2 = \hat{\sigma}^2_{(r)}$. In the Appendix B, we also show how to compute $g(\sigma^2)$ and $\frac{\partial}{\partial\sigma^2}g(\sigma^2)$ under count and binary cases.

Let $\hat{\sigma}^2_{RM}$ be the solution of (3.22), which is obtained by using the iterative equation (3.24). Once again under some regularity conditions, the asymptotic properties of $\hat{\sigma}^2_{RM}$ can be studied in the manner similar to that of the asymptotic properties of

$\hat{\beta}_{RGQL}$. There are however not discussed in this chapter. As far as the small sample properties of the estimator $\hat{\sigma}^2_{RM}$ is concerned, they are also discussed in section 3.4 along with the small sample properties of $\hat{\beta}_{RGQL}$, the RGQL estimator of β.

3.4 Performance of the RGQL Estimators: A Simulation Study

In section 3.2.3, a simulation study was conducted to examine the effects of the possible outliers on the so-called GQL estimation approach. It was found that both the unmodified GQL estimates for the regression effects, β, and the unmodified moment estimate for the variance of the random effects, σ^2, were highly biased and hence inconsistent. As a remedy, in the last section, we have developed a robust GQL approach that accommodates the presence of outliers in estimating β by solving the RGQL estimating equation (3.19), and σ^2 by solving the RM estimating equation (3.22). The purpose of this section is to examine the performances of $\hat{\beta}_{RGQL}$ and $\hat{\sigma}^2_{RM}$, where $\hat{\beta}_{RGQL}$ and $\hat{\sigma}^2_{RM}$ are the RGQL estimator of β and RM estimator of σ^2, respectively.

As far as the true value of $\beta = (\beta_1, \beta_2)'$ and σ^2 are concerned, we keep the same values for these parameters as in the last simulation conducted in section 3.2.3. Thus, we consider two sets of true values for β as $\beta = (\beta_1, \beta_2)' = (1.0, 1.0)'$ and $\beta = (\beta_1, \beta_2)' = (1.5, 0.75)'$ corresponding to the design D_1, for the Poisson mixed model; and also two sets of values for β as $\beta = (1.0, 1.0)'$ and $\beta = (1.5, 0.5)'$ corresponding to the design D_2, and $\beta = (-1.0, -1.0)'$ and $\beta = (-0.5, -1.5)'$ corresponding to the design D_3, for the binary mixed model. Note that these three designs D_1, D_2, and D_3 for the selection of the covariates were already explained in section 3.2.1.

Furthermore, as the true values of σ^2, we have chosen $\sigma^2 = 0.25,\ 0.5,$ and 0.75 as before.

Next, all together $\sum_{i=1}^{K} n_i = 400$ count responses with $m = 4$ outliers were generated as in the last simulation following section 3.2.1. Similarly, for the binary case, 400 responses with $m = 4$ outliers were generated following section 3.2.1. Note that these count and binary responses were generated in the cluster set up with $K = 100$ clusters, each with size $n_i = 4$. The simulated robust estimates for the clustered count and binary data are now discussed in the sections 3.4.1 and 3.4.2, respectively.

3.4.1 Poisson case

Similar to the binomial case discussed by Cantoni and Ronchetti [2001, Tables 1 and 2], we have chosen the tuning constant $c = 1.2$ for the robust estimation of the parameters under the Poisson mixed model. Next, by using the covariates and responses as generated in section 3.2.1, under each simulation, we have obtained $\hat{\beta}_{RGQL}$, the RGQL estimate of β as a solution of (3.19), and $\hat{\sigma}^2_{RM}$, the RM estimate of σ^2 as a solution of (3.22). Note that the details for the construction of (3.19) and (3.22) under the Poisson mixed model are given in the appendix. All together, we have conducted the experiment for 500 simulations. The simulated mean (SM), simulated standard error (SSE), and mean squared error (MSE) of the estimates of β and σ^2 obtained from these 500 simulations under the design D_1 are reported in Table 3.4.

The results in Table 3.4 show that, in general, the RGQL estimates for both β_1 and β_2 are very close to their corresponding true values. The performance, however, appears to be better under the Poisson mixed model with $\beta = (1.0, 1.0)'$ as compared to that of the model with $\beta = (1.5, 0.75)'$. For example, when $\beta_1 = \beta_2 = 1.0$ and

$\sigma^2 = 0.5$, the RGQL estimates of β_1 and β_2 are 0.961 and 0.983, respectively, which are very close to the true values. When $\beta_1 = 1.5$ and $\beta_2 = 0.75$, and $\sigma^2 = 0.5$, the RGQL estimates of β_1 and β_2 are 1.465 and 0.692, respectively, which are reasonably good estimates, even though they are not so close as compared to those under the model with $\beta_1 = \beta_2 = 1.0$.

As far as the performance of the RM estimator of σ^2 is concerned, the robust moment estimator $\hat{\sigma}^2_{RM}$ does not perform as good as $\hat{\beta}_{RGQL}$. For example, when $\sigma^2 = 0.75$ along with $\beta_1 = 1.5$ and $\beta_2 = 0.75$, the estimate of σ^2 was found to be 0.687 for its true value 0.75 with MSE 0.027, whereas MSEs of $\hat{\beta}_{1,RGQL}$ and $\hat{\beta}_{2,RGQL}$ estimates were found to be 0.005 and 0.013, respectively. Nevertheless, the RM estimates of σ^2 appear to be acceptable for the practical purpose, especially when it is known that the existing traditional approaches such as the moment and PQL [Breslow and Clayton (1993)] approaches may produce estimate with relatively large bias, even if there is no outliers in the data. See, for example, Sutradhar and Qu (1998) on the performance of the PQL approach. Thus, the biases in Table 3.4 produced by the RM estimator of σ^2 seem to be acceptable for the practical purpose, even though there is scope to reduce the amount of biases. To be specific, one perhaps could also use the RGQL estimation approach to estimate σ^2 in order to reduce the biasness, which is, however, beyond the scope of the present chapter.

Note that when the simulated standard errors are considered, both RGQL and RM approaches always produce estimates with small standard errors. For example, when $\beta_1 = \beta_2 = 1.0$, and $\sigma^2 = 0.75$, the SSEs of $\hat{\beta}_{1,RGQL}$, $\hat{\beta}_{2,RGQL}$, and $\hat{\sigma}^2_{RM}$ are found to be 0.070, 0.057, and 0.143, which are reasonably small. This leads to small MSEs for all of these estimates.

3.4.2 Binary case

Under the binary mixed model, we have chosen, for example, the upper sided outliers case to examine the performances of the proposed robust approaches. The other two, i.e., the lower and two sided outliers cases may be studied similarly. To deal with the upper sided outliers case, we need to consider a tuning probability, which ranges between the large outlying probability and the maximum of the probabilities for the so-called 'good' observations. The tuning probability $\mu^{c_1} = 0.5$ or 0.6 satisfies the above criterion in the context of our simulation design considered for the upper sided outliers. We have also considered a large value for μ^{c_1} such as $\mu^{c_1} = 0.9$ to see its effects on the estimation, even though it does not satisfy the above range criterion.

Now, to obtain the robust estimates for β and σ^2 under the binary mixed model, we have solved the RGQL estimating equation (3.19) and RM estimating equation (3.22), respectively, by using the formulas from the appendix, appropriate for the binary mixed model. Since the iteration procedure in the binary case was found to converge more quickly than the Poisson case, we have chosen 1000 simulations to obtain $\hat{\beta}_{RGQL}$ and $\hat{\sigma}^2_{RM}$ under the binary mixed model. The SM, SSE, and MSE of the estimates of β and σ^2 computed from these 1000 simulations are reported in Tables 3.5 and 3.6 under the design D_2 and D_3, respectively.

It is clear from Table 3.5, for example, that irrespective of the true value of β_1 and β_2, the RGQL estimates of β_1 and β_2 are almost unbiased, especially when $\mu^{c_1} = 0.5$ or 0.6. The standard errors of the RGQL estimates of β_1 and β_2 are, however, generally large, yielding large MSEs. They appear to decrease slightly, when the true value of σ^2 increases. The difference in the tuning probabilities does not create any significant change in the MSEs of the regression estimates. Thus, the RGQL approach appears to produce unbiased and hence consistent estimates for the regression effects under the binary mixed model, even though the estimates may not be highly efficient.

As far as the estimation of σ^2 parameter is concerned, the RM estimates appear to be slightly biased when appropriate tuning probability, namely $\mu^{c_1} = 0.5$ or 0.6 is used in the robust estimation. When $\mu^{c_1} = 0.9$ is considered, the bias becomes quite large, which is expected as this tuning probability is almost the same as the outlying probability causing problem in downweighting. Note that when the true value of σ^2 increases, the bias of the RM estimate of σ^2 also increases. For example, when $\beta_1 = \beta_2 = 1.0$ and the tuning probability $\mu^{c_1} = 0.5$, the RM estimate of σ^2 appears to be $\hat{\sigma}^2_{RM} = 0.398$ with small standard error 0.092 for the true $\sigma^2 = 0.5$, whereas $\hat{\sigma}^2_{RM} = 0.635$ with standard error 0.108 for the true $\sigma^2 = 0.75$. This results with respect to biases are not unexpected as it is usually difficult to estimate this parameter even if there is no outliers in the data. For example, in the absence of outliers, Breslow and Lin (1995, p. 90) were able to obtain unbiased PQL estimate only when true $\sigma^2 \leq 0.25$. See also Sutradhar and Mukerjee (2005) for the improved results produced by the simulated maximum likelihood approach.

The results of Table 3.6 based on the design D_3 are quite similar to those of Table 3.5 based on the design D_2. To save space, they are not discussed here any further.

3.5 Concluding Remarks

The GQL approach [Sutradhar (2004)] is known to be consistent and more efficient than the simulated MM approaches [Jiang (1998) and Jiang and Zhang (2001)] in estimating the parameters of the mixed models for the clustered count and binary data. This result is, however, valid only when the clustered data do not contain any outliers. In the independent set up, Cantoni and Ronchetti (2001) proposed the MQL approach in estimating the parameters of the generalized linear models for the count and binomial data in the presence of outliers. In this chapter, we have provided a RGQL approach for the estimation of the regression effects and the RM approach for the estimation of the variance component of the random effects for the clustered count and binary data in the presence of outliers. While the RM approach is completely new, the RGQL approach may be treated as a generalization of the MQL approach of Cantoni and Ronchetti (2001) for the estimation of the regression parameters. The proposed fully standardized RGQL and RM approaches appear to produce consistent estimates for the regression effects and variance component, respectively.

We remark here that while we have used the existing definition of outliers for the count data, we have, however, provided a much more clear definition of outliers than the existing definitions for the binary data. Note that the new definition is proposed along the lines of the definition of outliers for the count and continuous such as Gaussian data. This new definition should clarify the concept of outliers to a major extent under the binary mixed models.

We further remark that following the GQL approach for the estimation of the variance component in the absence of outliers [Sutradhar (2004)], one could also develop the RGQL approach to estimate this variance component parameter with smaller bias than that of the RM approach. This new approach could, however, be naturally more complicated as it will involve higher order moments calculations in the presence of outliers.

Table 3.1: Simulated mean (SM), standard errors (SSE), and mean squared errors (MSE) of the **GQL** estimates for the regression parameters and the **MM** estimates of the variance component of the random effects of the **Poisson mixed model** under design D_1 when data contain $m = 4$ outliers, for the selected values of σ^2; $K = 100$; $n_i = 4$ $(i = 1, \ldots, K)$; 500 simulations.

Regression effects (β_1, β_2)	Variance component(σ^2)	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
(1.0, 1.0)	0.25	SM	0.794	-0.043	1.195
		SSE	1.050	0.711	0.523
		MSE	1.144	1.594	1.167
	0.50	SM	0.656	0.051	1.371
		SSE	1.062	0.721	0.594
		MSE	1.247	1.421	1.111
	0.75	SM	0.714	-0.030	1.456
		SSE	1.029	0.698	0.691
		MSE	1.140	1.493	0.976
(1.5, 0.75)	0.25	SM	0.998	-0.205	1.282
		SSE	0.869	0.632	0.697
		MSE	1.007	1.311	1.552
	0.50	SM	1.022	-0.231	1.395
		SSE	0.860	0.613	0.761
		MSE	0.969	1.338	1.381
	0.75	SM	1.007	-0.236	1.488
		SSE	0.787	0.572	0.804
		MSE	0.861	1.299	1.191

Table 3.2: Simulated mean (SM), standard errors (SSE), and mean squared errors (MSE) of the **GQL** estimates for the regression parameters and the **MM** estimates of the variance component of the random effects of the **binary mixed model** under design D_2 when data contain $m = 4$ outliers, for the selected values of σ^2; $K = 100$; $n_i = 4$ $(i = 1, \ldots, K)$; 1000 simulations.

Regression effects (β_1, β_2)	Variance component(σ^2)	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
(1.0, 1.0)	0.25	SM	1.486	-0.097	0.060
		SSE	0.216	0.310	0.074
		MSE	0.283	1.299	0.041
	0.50	SM	1.448	-0.099	0.097
		SSE	0.212	0.307	0.124
		MSE	0.245	1.301	0.178
	0.75	SM	1.432	-0.098	0.194
		SSE	0.209	0.301	0.235
		MSE	0.230	1.296	0.365
(1.5, 0.5)	0.25	SM	2.000	-0.535	0.076
		SSE	0.259	0.336	0.108
		MSE	0.316	1.183	0.042
	0.50	SM	1.952	-0.528	0.136
		SSE	0.257	0.334	0.203
		MSE	0.270	1.168	0.174
	0.75	SM	1.930	-0.529	0.232
		SSE	0.264	0.335	0.299
		MSE	0.254	1.172	0.358

Table 3.3: Simulated mean (SM), standard errors (SSE), and mean squared errors (MSE) of the **GQL** estimates for the regression parameters and the **MM** estimates of the variance component of the random effects of the **binary mixed model** under design D_3 when data contain $m = 4$ outliers, for the selected values of σ^2; $K = 100$; $n_i = 4$ $(i = 1, \ldots, K)$; 1000 simulations.

Regression effects (β_1, β_2)	Variance component(σ^2)	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
(-1.0, -1.0)	0.25	SM	-1.526	-0.529	0.072
		SSE	0.378	0.230	0.090
		MSE	0.419	0.275	0.040
	0.50	SM	-1.451	-0.509	0.184
		SSE	0.376	0.227	0.241
		MSE	0.344	0.292	0.158
	0.75	SM	-1.407	-0.501	0.321
		SSE	0.361	0.217	0.322
		MSE	0.296	0.297	0.288
(-0.5, -1.5)	0.25	SM	-1.158	-0.852	0.062
		SSE	0.400	0.257	0.064
		MSE	0.593	0.485	0.039
	0.50	SM	-1.119	-0.818	0.105
		SSE	0.375	0.239	0.147
		MSE	0.524	0.522	0.178
	0.75	SM	-1.084	-0.798	0.182
		SSE	0.365	0.229	0.230
		MSE	0.474	0.546	0.375

Table 3.4: Simulated mean (SM), standard errors (SSE), and mean squared errors (MSE) of the **RGQL** estimates for the regression parameters and the **RM** estimates of the variance component of the random effects of the **Poisson mixed model** under design D_1 when data contain $m = 4$ outliers, for the selected values of σ^2; tuning constant $c = 1.2$; $K = 100$; $n_i = 4$ $(i = 1, \ldots, K)$; 500 simulations.

Regression effects (β_1, β_2)	Variance component(σ^2)	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
(1.0, 1.0)	0.25	SM	1.023	1.025	0.180
		SSE	0.085	0.097	0.051
		MSE	0.008	0.010	0.008
	0.50	SM	0.961	0.983	0.379
		SSE	0.040	0.043	0.119
		MSE	0.003	0.002	0.029
	0.75	SM	0.973	1.014	0.676
		SSE	0.070	0.057	0.143
		MSE	0.006	0.004	0.026
(1.5, 0.75)	0.25	SM	1.490	0.695	0.234
		SSE	0.082	0.068	0.056
		MSE	0.007	0.008	0.003
	0.50	SM	1.465	0.692	0.408
		SSE	0.065	0.058	0.074
		MSE	0.006	0.007	0.014
	0.75	SM	1.530	0.665	0.687
		SSE	0.067	0.074	0.153
		MSE	0.005	0.013	0.027

Table 3.5: Simulated mean (SM), standard errors (SSE), and mean squared errors (MSE) of the **RGQL** estimates for the regression parameters and the **RM** estimates of the variance component of the random effects of the **binary mixed model** under design D_2 when data contain $m = 4$ outliers, for the selected values of σ^2 and tuning probability (μ^{c_1}); $K = 100$; $n_i = 4$ ($i = 1, \ldots, K$); 1000 simulations.

Regression effects (β_1, β_2)	Variance component(σ^2)	Tuning probability (μ^{c_1})	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
(1.0, 1.0)	0.25	0.5	SM	1.035	0.959	0.190
			SSE	0.269	0.549	0.072
			MSE	0.074	0.303	0.009
		0.6	SM	1.031	0.962	0.187
			SSE	0.264	0.543	0.071
			MSE	0.070	0.296	0.009
		0.9	SM	1.040	0.900	0.069
			SSE	0.282	0.562	0.040
			MSE	0.081	0.326	0.035
	0.50	0.5	SM	1.061	0.976	0.398
			SSE	0.251	0.499	0.092
			MSE	0.066	0.250	0.019
		0.6	SM	1.054	0.977	0.384
			SSE	0.250	0.500	0.089
			MSE	0.065	0.251	0.021
		0.9	SM	1.006	0.952	0.202
			SSE	0.239	0.488	0.072
			MSE	0.057	0.240	0.094
	0.75	0.5	SM	1.092	0.953	0.635
			SSE	0.246	0.491	0.108
			MSE	0.069	0.243	0.025
		0.6	SM	1.088	0.948	0.621
			SSE	0.245	0.489	0.104
			MSE	0.068	0.242	0.028
		0.9	SM	1.052	0.935	0.443
			SSE	0.240	0.482	0.096
			MSE	0.060	0.237	0.104

Cont...........Table 3.5

Regression effects (β_1, β_2)	Variance component(σ^2)	Tuning probability (μ^{c_1})	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma^2}$
(1.5, 0.5)	0.25	0.5	SM	1.529	0.458	0.183
			SSE	0.321	0.611	0.086
			MSE	0.104	0.376	0.012
		0.6	SM	1.517	0.460	0.175
			SSE	0.310	0.589	0.082
			MSE	0.096	0.349	0.012
		0.9	SM	1.524	0.414	0.062
			SSE	0.330	0.627	0.035
			MSE	0.109	0.400	0.037
	0.50	0.5	SM	1.542	0.487	0.395
			SSE	0.289	0.555	0.101
			MSE	0.085	0.308	0.021
		0.6	SM	1.535	0.480	0.384
			SSE	0.288	0.555	0.097
			MSE	0.084	0.309	0.023
		0.9	SM	1.489	0.467	0.195
			SSE	0.277	0.542	0.089
			MSE	0.077	0.295	0.101
	0.75	0.5	SM	1.536	0.528	0.623
			SSE	0.255	0.508	0.107
			MSE	0.066	0.259	0.027
		0.6	SM	1.528	0.528	0.610
			SSE	0.255	0.512	0.103
			MSE	0.066	0.263	0.030
		0.9	SM	1.481	0.512	0.424
			SSE	0.259	0.514	0.090
			MSE	0.067	0.264	0.115

Table 3.6: Simulated mean (SM), standard errors (SSE), and mean squared errors (MSE) of the **RGQL** estimates for the regression parameters and the **RM** estimates of the variance component of the random effects of the **binary mixed model** under design D_3 when data contain $m = 4$ outliers, for the selected values of σ^2 and tuning probability (μ^{c_1}); $K = 100$; $n_i = 4$ ($i = 1, \ldots, K$); 1000 simulations.

Regression	Variance	Tuning		Estimates		
effects (β_1, β_2)	component(σ^2)	probability (μ^{c_1})	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
(-1.0,-1.0)	0.25	0.5	SM	-0.961	-0.964	0.197
			SSE	0.359	0.263	0.200
			MSE	0.131	0.070	0.043
		0.6	SM	-0.955	-0.959	0.175
			SSE	0.354	0.259	0.184
			MSE	0.127	0.069	0.039
		0.9	SM	-0.910	-0.931	0.063
			SSE	0.334	0.246	0.051
			MSE	0.120	0.065	0.038
	0.50	0.5	SM	-0.962	-0.931	0.435
			SSE	0.352	0.249	0.216
			MSE	0.125	0.067	0.051
		0.6	SM	-0.955	-0.926	0.416
			SSE	0.353	0.248	0.207
			MSE	0.126	0.067	0.050
		0.9	SM	-0.900	-0.877	0.284
			SSE	0.337	0.236	0.148
			MSE	0.124	0.071	0.069
	0.75	0.5	SM	-0.960	-0.922	0.650
			SSE	0.300	0.211	0.250
			MSE	0.092	0.050	0.073
		0.6	SM	-0.954	-0.915	0.618
			SSE	0.299	0.209	0.242
			MSE	0.092	0.051	0.076
		0.9	SM	-0.894	-0.866	0.416
			SSE	0.271	0.182	0.173
			MSE	0.084	0.051	0.142

Cont...........Table 3.6

Regression effects (β_1, β_2)	Variance component(σ^2)	Tuning probability (μ^{c_1})	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
(-0.5, -1.5)	0.25	0.5	SM	-0.472	-1.439	0.173
			SSE	0.365	0.290	0.173
			MSE	0.134	0.088	0.036
		0.6	SM	-0.466	-1.438	0.153
			SSE	0.362	0.287	0.158
			MSE	0.132	0.086	0.034
		0.9	SM	-0.440	-1.396	0.061
			SSE	0.342	0.274	0.046
			MSE	0.121	0.086	0.038
	0.50	0.5	SM	-0.490	-1.397	0.418
			SSE	0.351	0.268	0.230
			MSE	0.123	0.082	0.060
		0.6	SM	-0.482	-1.392	0.403
			SSE	0.346	0.264	0.219
			MSE	0.120	0.081	0.057
		0.9	SM	-0.440	-1.332	0.251
			SSE	0.317	0.229	0.146
			MSE	0.104	0.081	0.083
	0.75	0.5	SM	-0.482	-1.393	0.661
			SSE	0.340	0.259	0.226
			MSE	0.116	0.078	0.059
		0.6	SM	-0.480	-1.383	0.634
			SSE	0.330	0.251	0.218
			MSE	0.109	0.077	0.061
		0.9	SM	-0.468	-1.296	0.475
			SSE	0.302	0.219	0.146
			MSE	0.093	0.089	0.097

Chapter 4

Robust Estimation for Longitudinal Count and Binary Data

In the last chapter, we have dealt with the consistent estimation of the parameters of the mixed models for the count and binary data, in the presence of outliers. In practice, as opposed to the familial set up, one may encounter repeated count and/or binary data collected from a large number of independent individuals. The data of this type are referred to as the longitudinal data. It is also possible that a few observations in such a longitudinal set up may be outliers, which may negatively influence the inferences about the true parameters of the model.

Note that in the longitudinal set up, the repeated observations of an individual are likely to be correlated. This correlation mainly occurs because of the stochastic effects of times on the repeated responses. This type of longitudinal correlation structure is different than the familial correlation structure. Recall from (3.1) that in the familial set up, the jth and kth, $(j \neq k;\ j,k = 1,\ldots,n_i)$ members of the ith $(i = 1,\ldots,K)$ family are correlated because they share a common random effect $\gamma_i^* \overset{i.i.d.}{\sim} N(0,\sigma^2)$. This causes familial correlation among the responses of the members of the ith family. For example, in the absence of outliers, when

$$y_{ij|\gamma_i} \sim Poisson(\mu_{ij}^*) \text{ with } \mu_{ij}^* = exp(x_{ij}'\beta - \frac{\sigma^2}{2} + \sigma\gamma_i),$$

one may show that the unconditional correlation between Y_{ij} and Y_{ik} is given by

$$corr(Y_{ij}, Y_{ik}|x_{ij}, x_{ik}) = \frac{c\mu_{ij}\mu_{ik}}{\sqrt{(\mu_{ij} + c\mu_{ij}^2)(\mu_{ik} + c\mu_{ik}^2)}}, \tag{4.1}$$

where [see (3.5) when data contain outliers] $c = exp(\sigma^2) - 1$ and $\mu_{ij} = exp(x_{ij}'\beta)$ for all $i = 1,\ldots,K$, $j = 1,\ldots,n_i$, n_i being the family size. In the longitudinal set up, the correlations among the repeated responses usually arise from a dynamic relationship of the repeated responses. For example, in the absence of outliers, let $y_{i1} \sim Poisson(\mu_{i1})$ with $\mu_{i1} = E(Y_{i1}|x_{i1}) = exp(x_{i1}'\beta)$, and for $t = 2,\ldots,T$, y_{it} be the tth count response collected at time t from the ith individual and $y_{i,t-1}$ be related to y_{it} through the dynamic model

$$y_{it} = \rho * y_{i,t-1} + d_{it},\ t = 2,\ldots,T, \tag{4.2}$$

[see, McKenzie (1988) and Sutradhar (2003)], where $y_{i,t-1} \sim Poisson(\mu_{i,t-1})$ and $d_{it} \sim Poisson(\mu_{it} - \rho\mu_{i,t-1})$, with $\mu_{it} = E(Y_{it}|x_{it},\ldots,x_{i1}) = exp(x_{it}'\beta)$, where x_{it} is a $p-$ dimensional vector of covariates recorded at time t for the ith individual. Here, d_{it} and $y_{i,t-1}$ are assumed to be independent. In (4.2), 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 $P[b_j(\rho) = 0] = 1-\rho$. This operation in (4.2), i.e., $\rho * y_{i,t-1}$ is known as the so-called binomial thinning operation. It can be shown that this dynamic relation (4.2) causes y_{iu} and y_{it} $(u < t)$ to be correlated as

$$corr(Y_{iu}, Y_{it}|x_{it}, \ldots, x_{i1}) = \rho^{t-u}\sqrt{\frac{\mu_{iu}}{\mu_{it}}}, \tag{4.3}$$

where ρ satisfies the range restriction

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

Some authors such as Sinha (2006) and Mills et al. (2002) studied the inferences for the regression effects in the longitudinal set up in the presence of outliers. These authors have, however, used a familial model similar to (4.1) to generate the correlation structure for the repeated responses, which does not appear to accommodate the stochastic time effects. As opposed to Sinha (2006) and Mills et al. (2002), Preisser and Qaqish (1999) and Cantoni (2004) have considered the robust inferences where it is assumed that there exists a stochastic time based correlation structure, but this correlation structure is unspecified. To accommodate the correlation effects, these authors have used a 'working' correlation approach, which in view of Sutradhar and Das (1999) [see also, Crowder (1995)] may produce inconsistent and inefficient estimates.

In this chapter, we consider AR(1) type correlation model (4.2) to generate the count data with a few outliers, and estimate the parameters β and ρ by solving appropriate robust estimating equations. Note that in the present longitudinal set up, we use similar definitions for the outliers as in the familial set up. See, section 3.1.1 for the definition of outliers for the count data, and section 3.1.2 for the outliers in the binary case. We then use the RGQL estimating equation similar to that under the familial models to obtain consistent estimate of the regression effects. The only difference between the RGQL estimating equations in the familial and longitudinal set up is the use of correlation matrices corresponding to these models. To estimate the correlation parameter, we use a robust moment (RM) approach. The RGQL and RM estimating equations are provided in section 4.3.

The performance of the RGQL and RM approaches under the count and binary longitudinal data in the presence of outliers is examined in section 4.4 through a simulation study. In the following sections, we, however, conduct a basic simulation study to examine the effects of outliers on the so-called unmodified GQL inferences for the regression and correlation parameters, for both count and binary data.

4.1 Effects of Outliers on Unmodified GQL Estimation for Longitudinal Count Data

It is well known that in the absence of outliers in the longitudinal clustered data, the GQL approach [Sutradhar (2003)] produces consistent and highly efficient estimates for the regression effects and correlation parameters under a longitudinal model. The main purpose of this section is to examine the effects of the presence of one or more outliers on the unmodified GQL inferences for the count data through a simulation study. We, however, first review the GQL estimation approach as follows in the absence of outliers.

4.1.1 GQL approach in the absence of outlier

It follows from the model (4.2) that

$$E(Y_{i1}|x_{i1}) = \mu_{i1}$$

$$\begin{aligned}
E(Y_{it}|x_{it},\ldots,x_{i1}) &= E_{Y_{i,t-1}} E\left[\sum_{j=1}^{y_{i,t-1}} b_j(\rho) + d_{it}\right] \\
&= E_{Y_{i,t-1}}\left[\rho Y_{i,t-1}|x_{i,t-1},\ldots,x_{i1} + \mu_{it} - \rho\mu_{i,t-1}\right] \\
&= \rho\mu_{i,t-1} + \mu_{it} - \rho\mu_{i,t-1} \\
&= \mu_{it}, \text{ for } t = 2,\ldots,T. \qquad (4.4)
\end{aligned}$$

Similarly, it can be shown that

$$var(Y_{i1}|x_{i1}) = \mu_{i1}$$

$$\begin{aligned}
var(Y_{it}|x_{it},\ldots,x_{i1}) &= E_{Y_{i,t-1}} var(Y_{it}|Y_{i,t-1},x_{it},\ldots,x_{i1}) \\
&\quad + var_{Y_{i,t-1}} E(Y_{it}|Y_{i,t-1},x_{it},\ldots,x_{i1}) \\
&= \mu_{it} = \sigma_{itt}, \text{ for } t = 2,\ldots,T, \qquad (4.5)
\end{aligned}$$

which is the same as the expectation in (4.4). Next, by computing $E(Y_{i,t-1}Y_{it}|x_{it},\ldots,x_{i1})$ as

$$\begin{aligned}
E(Y_{i,t-1}Y_{it}|x_{it},\ldots,x_{i1}) &= E_{Y_{i,t-1}} Y_{i,t-1}|x_{i,t-1},\ldots,x_{i1} E\left[Y_{it}|Y_{i,t-1},x_{it},\ldots,x_{i1}\right] \\
&= E_{Y_{i,t-1}} Y_{i,t-1}|x_{i,t-1},\ldots,x_{i1} E\left[\sum_{j=1}^{y_{i,t-1}} b_j(\rho) + d_{it}\right] \\
&= E_{Y_{i,t-1}}\left[\rho Y_{i,t-1}^2|x_{i,t-1},\ldots,x_{i1} + Y_{i,t-1}|x_{i,t-1},\ldots,x_{i1}(\mu_{it} - \rho\mu_{i,t-1})\right] \\
&= \rho\left[var(Y_{i,t-1}|x_{i,t-1},\ldots,x_{i1}) + \mu_{i,t-1}^2\right] + \mu_{i,t-1}(\mu_{it} - \rho\mu_{i,t-1}) \\
&= \rho\mu_{i,t-1} + \mu_{i,t-1}\mu_{it}, \qquad (4.6)
\end{aligned}$$

one may obtain covariance between $Y_{i,t-1}$ and Y_{it} given by

$$cov(Y_{i,t-1}, Y_{it}|x_{it}, \ldots, x_{i1}) = \sigma_{it,t-1} = \rho\mu_{i,t-1}. \tag{4.7}$$

By similar calculations, it can be shown that the lag ℓ correlation between $Y_{i,t-\ell}$ and Y_{it} is

$$corr(Y_{i,t-\ell}, Y_{it}|x_{it}, \ldots, x_{i1}) = \rho^{\ell}\sqrt{\frac{\mu_{i,t-\ell}}{\mu_{it}}}, \tag{4.8}$$

leading to the correlation structure (4.3).

Let $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$, where y_{it} is a outlier free count response generated by the model (4.2). Further let $\mu_i = (\mu_{i1}, \ldots, \mu_{it}, \ldots, \mu_{iT})'$ and $\Sigma_i = (\sigma_{ut})$, where $\mu_{it} = E(Y_{it}|x_{it}, \ldots, x_{i1}) = exp(x'_{it}\beta)$ by (4.4), $\sigma_{itt} = \mu_{it}$ by (4.5), and $\sigma_{iut} = \rho^{t-u}\mu_{iu}$ for $u < t$ by (4.7). Now, by following Sutradhar (2003), one may write the GQL estimating equation for β as

$$\sum_{i=1}^{K}\left[\frac{\partial\mu_i'}{\partial\beta}\Sigma_i^{-1}(\rho)(y_i - \mu_i)\right] = 0, \tag{4.9}$$

where $\Sigma_i(\rho) = (\sigma_{iut})$, $u, t = 1, \ldots, T$, and $\frac{\partial\mu_i'}{\partial\beta}$ is the $p \times T$ derivative matrix of μ_i with respect to β. It is known that $\hat{\beta}_{GQL}$ obtained from (4.9) is consistent and highly efficient for β.

Note that in estimating β by (4.9), it was assumed that ρ is known. To estimate this parameter, one may use the moment method (MM) and compute $\hat{\rho}_M$ by

$$\hat{\rho}_M = \frac{\frac{\sum_{i=1}^{K}\sum_{t=2}^{T} r^*_{i,t-1} r^*_{it}}{K(T-1)}}{\left[\frac{\sum_{i}^{K}\sum_{i=1}^{T} {r^*_{it}}^2}{KT}\right]\left[\frac{\sum_{i=1}^{K}\sum_{t=2}^{T}\sqrt{\frac{\mu_{i,t-1}}{\mu_{it}}}}{K(T-1)}\right]}, \tag{4.10}$$

where r^*_{it} is the standardized residuals defined as $r^*_{it} = \frac{y_{it}-\mu_{it}}{\sqrt{\sigma_{itt}}}$, with μ_{it} and σ_{itt} as given in (4.4) and (4.5), respectively.

4.1.2 Generating longitudinal count data with outliers

Suppose that there exists m outliers out of KT count responses. For the purpose, we first generate T 'good' correlated observations for each of the K individuals following (4.2) with $\mu_{it} = exp(x'_{it}\beta)$.

We consider $K = 100$ and $T = 4$ in the simulation study. As far as the correlation ρ is concerned, we choose $\rho = 0.25,\ 0.5,$ and 0.8. For the selection of the covariates x_{it}, we use $p\ (= 2)-$ dimensional covariates as

$$x_{it1} = \begin{cases} -1.0 & \text{for } i = 1, \ldots, K/4;\ t = 1, \ldots, T \\ 0.0 & \text{for } i = K/4+1, \ldots, 3K/4;\ t = 1, \ldots, T \\ 1.0 & \text{for } i = 3K/4+1, \ldots, K\ t = 1, \ldots, T, \end{cases},$$

and

$$x_{it2} \sim N(0.5, 1.0),$$

for all $i = 1, \ldots, K,\ t = 1, \ldots, T$. As the effects of these covariates, we consider two versions of the regression parameters, namely $\beta = (0.5, 0.5)'$ and $\beta = (1.0, 1.0)'$. By using the above covariates and the selected parameter values, we now generate KT 'good' responses following the longitudinal AR(1) count model (4.2).

We then create m outliers by choosing m responses and shifting their corresponding covariates by an amount δ, where δ is a vector. To be specific, in creating m outliers, we do not change the values of m responses [i.e., we keep them as they were generated (as 'good' observations)], but their corresponding covariates, say $x_{i't'}$ are replaced with $\tilde{x}_{i't'} = x_{i't'} + \delta$. In this way, the $t'th$ 'good' observation of the $i'th$ individual, that is, $y_{i't'}$ is now converted to an outlier. In general, we now (as before) denote the observed covariates as

$$\tilde{x}_{it} = \begin{cases} x_{it} & \text{for } (i,t) \not\equiv (i',t') \\ x_{it} + \delta & \text{for } (i,t) \equiv (i',t') \end{cases}. \tag{4.11}$$

As far as the values of δ is concerned, we consider $\delta = (\delta_1, \delta_2)' = (3.0, 3.0)'$. Now, only for $m = 4$ responses, we use the observed covariates as $\tilde{x}_{i't'} = x_{i't'} + \delta$, i. e., $\tilde{x}_{i't'1} = x_{i't'1} + \delta_1$ and $\tilde{x}_{i't'2} = x_{i't'2} + \delta_2$. It then follows that these 4 responses are now outliers among the 400 responses, and they may be denoted, for convenience, by $y_{i't'}$.

4.1.3 Unmodified GQL estimation

For the estimation of the regression effects β, we apply the GQL estimating equation (4.9) to the longitudinal count data with $m = 4$ outliers that we have generated in the last subsection. Note that following (4.11), the component μ_{it} in the mean vector μ_i and covariance matrix $\Sigma_i(\rho)$ in (4.9) is now written as

$$\tilde{\mu}_{it} = exp(\tilde{x}'_{it}\beta). \tag{4.12}$$

Thus, when outliers are generated, but the unmodified GQL estimating equation is used, this new equation would simply be derived by replacing x_{it} with $\tilde{x}_{it}$. Note that y_i however remains the same as in (4.9) in such an unmodified GQL estimating equation. For convenience, this unmodified GQL estimating equation is written as

$$\sum_{i=1}^{K} \left[\frac{\partial \tilde{\mu}'_i}{\partial \beta} \tilde{\Sigma}_i^{-1}(\rho)(y_i - \tilde{\mu}_i) \right] = 0, \tag{4.13}$$

where $\tilde{\mu}_i = \mu_i|_{x_{it}=\tilde{x}_{it}}$ and $\tilde{\Sigma}_i(\rho) = \Sigma_i(\rho)|_{x_{it}=\tilde{x}_{it}}$, with μ_i and $\Sigma_i(\rho)$ being given as in (4.9).

Next, to estimate the correlation parameter ρ, we use the formula given in (4.10), but for the computation of this formula, we simply replace x_{it} with $\tilde{x}_{it}$. For convenience, we denote this estimate as

$$\hat{\tilde{\rho}}_M = \hat{\rho}_M|_{x_{it}=\tilde{x}_{it}}, \tag{4.14}$$

where $\hat{\rho}_M$ is given in (4.10).

4.1.4 Simulation results

For a selected value of ρ and a true regression vector β, under each simulation, we now estimate β by solving the unmodified GQL estimating equation (4.13) and estimate ρ by using the moment estimator from (4.14). We consider 1000 simulations. The simulated mean (SM), simulated standard error (SSE), and mean squared error (MSE) computed from these 1000 values for $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\rho}$ are reported in Table 4.1.

It is clear from this table that the GQL estimate of β is adversely affected by the presence of outliers. The performance of the GQL approach, however, appears to be worse under the model with regression effects $\beta = (1.0, 1.0)'$ as compared to that of the model with $\beta = (0.5, 0.5)'$. For example, when $\beta_1 = \beta_2 = 1.0$ and $\rho = 0.5$, the estimates for $\beta_1 = 1.0$ and $\beta_2 = 1.0$ are found to be -0.554 and 0.725, respectively, with corresponding standard errors 0.094 and 0.034; whereas when $\beta_1 = \beta_2 = 0.5$ and $\rho = 0.5$, these estimates are found to be -0.023 and 0.343, respectively, with corresponding standard error 0.069 and 0.043. Note that these estimates are completely unacceptable in both cases, the case with $\beta = (0.5, 0.5)'$ being slightly better. When the regression estimates are compared due to the change in ρ, it is apparent that, in general, the unmodified GQL approach performs worse when the value of the true correlation parameter gets larger.

With regard to the performance of the unmodified MM approach for the estimation of ρ in the presence of outliers, it is found that the MM estimate of ρ also gets biased because of the outliers. The biasness, however, appears to decrease when the true ρ gets larger. For example, when $\beta_1 = \beta_2 = 1.0$, the MM estimate of ρ is 0.847 with standard error 0.060 for true $\rho = 0.25$; whereas for true $\rho = 0.8$, this estimate is found to be 0.964 with standard error 0.016. Similar results hold when $\beta_1 = \beta_2 = 0.5$. Note however that when MM estimates of ρ are compared due to change in β_1 and β_2, it is observed that the estimate of ρ is worse when $\beta_1 = \beta_2 = 1.0$ as compared to that for $\beta_1 = \beta_2 = 0.5$. For example, when $\beta_1 = \beta_2 = 1.0$ and $\rho = 0.25$, the estimate of ρ is 0.847 with standard error 0.060; whereas when $\beta_1 = \beta_2 = 0.5$ and $\rho = 0.25$, this estimate is 0.352 with standard error 0.071.

4.2 Effects of Outliers on Unmodified GQL Estimation for Longitudinal Binary Data

Recall that in the absence of outliers, longitudinal count data were generated following the model (4.2). In the binary case, we use a similar but different model given as follows:

$$y_{i1} \sim bin(\mu_{i1}) \text{ and } y_{it}|y_{i,t-1} \sim bin[\mu_{it} + \rho(y_{i,t-1} - \mu_{i,t-1})], \tag{4.15}$$

[Zeger et al. (1985) and Qaqish (2003)] with $\mu_{it} = \frac{exp(x'_{it}\beta)}{1+exp(x'_{it}\beta)}$. Marginally, y_{it} follows the binary distribution with probability of success μ_{it}.

In the following section 4.2.1, we provide a GQL estimating equation approach for the estimation of β and ρ parameters in the absence of outliers. In section 4.2.2, we demonstrate how to generate a few outliers in the binary longitudinal set up.

In section 4.2.3, we indicate how the GQL approach of section 4.2.1 can be used for the longitudinal binary data with outliers. Some simulation results are given in section 4.2.4 to examine the effects of outliers on the unmodified GQL approach to be discussed in section 4.2.3.

4.2.1 GQL approach in the absence of outlier

Note that we have used the GQL estimating equation (4.9) for the estimation of the regression effects β and the MM estimating equation (4.10) for the estimation of ρ parameter, for the longitudinal count data in the absence of outliers. We may still use these equations with slight change in the present binary case. For convenience, we re-write the Poisson model based GQL estimating equation (4.9) as

$$\sum_{i=1}^{K}\left[\frac{\partial\mu_i'}{\partial\beta}\Sigma_i^{-1}(\rho)(y_i-\mu_i)\right]=0, \tag{4.16}$$

but μ_i and $\Sigma_i(\rho)$ in (4.16) would be different than those in (4.9), under the present binary case. More specifically, for $\mu_i=(\mu_{i1},\ldots,\mu_{it},\ldots,\mu_{iT})'$, the component μ_{it} has the formula given by

$$\mu_{it}=\frac{exp(x_{it}'\beta)}{1+exp(x_{it}'\beta)}, \tag{4.17}$$

for all $t=1,\ldots,T$; where $x_{it}=(x_{it1},\ldots,x_{itu},\ldots,x_{itp})'$ and β is the $p-$ dimensional vector for the regression parameters. This is because

$$E(Y_{i1}|x_{i1}) \;=\; \mu_{i1}$$

$$\begin{aligned}
E(Y_{it}|x_{it},\ldots,x_{i1}) &= E_{Y_{i1}}E_{Y_{i2}}\ldots E_{Y_{i,t-2}}E_{Y_{i,t-1}}[\mu_{it}+\rho(Y_{i,t-1}|x_{i,t-1},\ldots,x_{i1}-\mu_{i,t-1})]\\
&= E_{Y_{i1}}E_{Y_{i2}}\ldots E_{Y_{i,t-2}}[\mu_{it}+\rho^2(Y_{i,t-2}|x_{i,t-2},\ldots,x_{i1}-\mu_{i,t-2})]
\end{aligned}$$

$$
\begin{aligned}
&\vdots \qquad \vdots \qquad \vdots \qquad \vdots \\
&= E_{Y_{i1}}[\mu_{it} + \rho^{t-1}(Y_{i1}|x_{i1} - \mu_{i1})] \\
&= \mu_{it},\ t = 2, \ldots, T.
\end{aligned} \tag{4.18}
$$

With regard to the computation of the elements of $\Sigma_i(\rho)$ matrix, we first note that the variance of the binary variable Y_{it} is given by

$$
\begin{aligned}
var(Y_{it}|x_{it}, \ldots, x_{i1}) &= E(Y_{it}^2|x_{it}, \ldots, x_{i1}) - [E(Y_{it}|x_{it}, \ldots, x_{i1})]^2 \\
&= \mu_{it}(1 - \mu_{it}) = \sigma_{itt}.
\end{aligned} \tag{4.19}
$$

Next, for the computation of the non-diagonal elements, we compute the covariance between Y_{iu} and Y_{it}, $u < t$, as

$$
\begin{aligned}
&cov(Y_{iu}, Y_{it}|x_{it}, \ldots, x_{i1}) = E[(Y_{iu}|x_{iu}, \ldots, x_{i1} - \mu_{iu})(Y_{it}|x_{it}, \ldots, x_{i1} - \mu_{it})] \\
&= E_{Y_{iu}} E_{Y_{i,u+1}} \cdots E_{Y_{i,t-1}}[\rho(Y_{iu}|x_{iu}, \ldots, x_{i1} - \mu_{iu})(Y_{i,t-1}|x_{i,t-1}, \ldots, x_{i1} - \mu_{i,t-1})] \\
&= E_{Y_{iu}} E_{Y_{i,u+1}} \cdots E_{Y_{i,t-2}}[\rho^2(Y_{iu}|x_{iu}, \ldots, x_{i1} - \mu_{iu})(Y_{i,t-2}|x_{i,t-2}, \ldots, x_{i1} - \mu_{i,t-2})] \\
&\vdots \qquad \vdots \qquad \vdots \qquad \vdots \\
&= E_{Y_{iu}}[\rho^{t-u}(Y_{iu}|x_{iu}, \ldots, x_{i1} - \mu_{iu})^2] = \rho^{t-u}\sigma_{iuu}.
\end{aligned} \tag{4.20}
$$

This implies that lag $t-u$ correlation between Y_{iu} and Y_{it} can be written as

$$
corr(Y_{iu}, Y_{it}|x_{it}, \ldots, x_{i1}) = \rho^{t-u}\sqrt{\frac{\sigma_{iuu}}{\sigma_{itt}}}. \tag{4.21}
$$

Note that to compute the derivative matrix $\frac{\partial \mu_i'}{\partial \beta}$, it sufficient to calculate $\frac{\partial \mu_{it}}{\partial \beta}$. This derivative for the present binary case is given by

$$
\frac{\partial \mu_{it}}{\partial \beta} = \mu_{it}(1 - \mu_{it})x_{it}, \tag{4.22}
$$

where μ_{it} is given in (4.17).

With regard to the estimation of the correlation parameter ρ, we first observe by (4.8) that $corr(Y_{iu}, Y_{it}|x_{it}, \ldots, x_{i1}) = \rho^{t-u}\sqrt{\frac{\mu_{iu}}{\mu_{it}}}$, for $u < t$, under the Poisson longitudinal model, whereas under the present binary model, the formula for this correlation is given by (4.21). Consequently, by making a slight change in (4.10), i.e., by replacing $\frac{\mu_{iu}}{\mu_{it}}$ with $\frac{\sigma_{iuu}}{\sigma_{itt}}$, one may write the MM estimate of ρ as

$$\hat{\rho}_M = \frac{\frac{\sum_{i=1}^K \sum_{t=2}^T r^*_{i,t-1} r^*_{it}}{K(T-1)}}{\left[\frac{\sum_i^K \sum_{i=1}^T {r^*_{it}}^2}{KT}\right]\left[\frac{\sum_{i=1}^K \sum_{t=2}^T \sqrt{\frac{\sigma_{i,t-1t-1}}{\sigma_{itt}}}}{K(T-1)}\right]}, \tag{4.23}$$

where $r^*_{it} = \frac{y_{it}-\mu_{it}}{\sqrt{\sigma_{itt}}}$, with μ_{it} and σ_{itt} as given in (4.17) and (4.19), respectively.

4.2.2 Generating longitudinal binary data with outliers

In practice, similar to that of the familial model for the binary data, one may deal with two types of binary outliers, namely one and two sided outliers under a longitudinal set up. These one and two sided longitudinal binary outliers can therefore be defined in the manner similar to that of the familial model described in section 3.1.2. For convenience, we now consider only one sided binary outliers where bulk of the 'good' observations are generated with small success probabilities.

To generate KT longitudinal binary observation with m outliers, we first generate KT 'good' responses following the model (4.12) with $\mu_{it} = \frac{exp(x'_{it}\beta)}{1+exp(x'_{it}\beta)}$ as given in (4.17).

In the simulation study, we consider $K = 100$ and $T = 4$. As the true values of correlation index parameter ρ, we choose the same values as in the longitudinal count case, namely $\rho = 0.25,\ 0.5,$ and 0.8. As far as the covariates are concerned to generate $KT = 400$ 'good' observations, we consider $p = 2$ with

$$x_{it1} = \begin{cases} -1.0 & \text{for } i = 1, \ldots, K/2;\ t = 1, \ldots, T \\ -0.25 & \text{for } i = K/2+1, \ldots, 3K/4;\ t = 1, \ldots, T \\ -0.5 & \text{for } i = 3K/4+1, \ldots, K\ t = 1, \ldots, T, \end{cases},$$

and

$$x_{it2} \sim N(-0.5, 0.25),$$

for all $i = 1, \ldots, K$, $t = 1, \ldots, T$. We consider two sets of values for the regression parameters, namely $\beta = (1.0, 1.0)'$ and $\beta = (1.5, 0.5)'$, leading to produce small success probabilities for the 400 longitudinal binary responses. Following model (4.15), we now generate KT 'good' binary responses by using the above covariates and the selected parameter values of β and ρ.

To create $m = 4$ outliers out of $KT = 400$ longitudinal binary responses, we choose m responses and change their corresponding covariates by an amount δ, δ being a real valued vector. Note that in creating m outliers, we do not change the values of these m responses (i.e., keep them as they were generated as 'good' observations); but we change their corresponding covariates, say $x_{i't'}$ by adding δ with it, that is, $\tilde{x}_{i't'} = x_{i't'} + \delta$. Similar to the longitudinal count case, we denote the observed covariates by $\tilde{x}_{it}$, where

$$\tilde{x}_{it} = \begin{cases} x_{it} & \text{for } (i,t) \not\equiv (i',t') \\ x_{ij} + \delta & \text{for } (i,t) \equiv (i',t') \end{cases}. \tag{4.24}$$

As the value of δ, we choose $\delta = (\delta_1, \delta_2)' = (3.0, 1.0)'$. Therefore, the covariates corresponding to the outlying responses will be $\tilde{x}_{i't'1} = x_{i't'1} + \delta_1$ and $\tilde{x}_{i't'2} = x_{i't'2} + \delta_2$, which yields larger success probabilities for the outlying responses. Thus, these $m = 4$ responses are now outliers, which are denoted, for convenience, by $y_{i't'}$.

4.2.3 Unmodified GQL estimation

To estimate the GQL estimate of the regression parameter β in the presence of binary outliers, we simply solve the unmodified GQL estimating equation given by

$$\sum_{i=1}^{K}\left[\frac{\partial\tilde{\mu}_i'}{\partial\beta}\tilde{\Sigma}_i^{-1}(\rho)(y_i - \tilde{\mu}_i)\right] = 0, \tag{4.25}$$

where $\tilde{\mu}_i = \mu_i|_{x_{it}=\tilde{x}_{it}}$ and $\tilde{\Sigma}_i(\rho) = \Sigma_i(\rho)|_{x_{it}=\tilde{x}_{it}}$, with μ_i and $\Sigma_i(\rho)$ as given in (4.16) for the binary case in the absence of outliers, and $\tilde{x}_{it}$ as in (4.24). To be specific, the $\tilde{\mu}_{it}$ component of $\tilde{\mu}_i$, for example, is given by

$$\tilde{\mu}_{it} = \frac{exp(\tilde{x}_{it}'\beta)}{1 + exp(\tilde{x}_{it}'\beta)}. \tag{4.26}$$

Note that $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ in (4.25) is generated following (4.15) in the absence of outliers.

Next, one may compute the unmodified MM estimate of ρ by replacing the covariate x_{it} with $\tilde{x}_{it}$ in the MM estimate of ρ obtained in the absence of outliers. That is, this estimate is obtained as

$$\hat{\tilde{\rho}}_M = \hat{\rho}_M|_{x_{it}=\tilde{x}_{it}}, \tag{4.27}$$

where $\hat{\rho}_M$ is given in (4.23).

4.2.4 Simulation results

To examine the effects of outliers on the unmodified GQL and MM estimating equations (4.25) and (4.27), respectively, we exploit these equations by using the longitudinal binary data with outliers generated in section 4.2.2. The unmodified GQL estimate of β and MM estimate ρ are repeatedly obtained based on 1000 simulations. We then compute the SM, SSE, and MSE for each of these three estimates, which are reported in Table 4.2.

It is clear from this table that similar to the Poisson case, the unmodified GQL estimates of β_1 and β_2 are highly biased. The unmodified GQL approach performs worst for the model with $\beta_1 = 1.5$ and $\beta_2 = 0.5$ as compared to that for the model with $\beta_1 = \beta_2 = 1.0$. As in the Poisson case, the biasness in the estimates of β_1 and β_2, however, increases with the increase in the true correlation parameter ρ.

As far as the estimation of the correlation parameter ρ is concerned, it is apparent from Table 4.2 that the unmodified MM approach performs well in estimating ρ. These estimates appear to be almost unbiased under the longitudinal binary models whether $\beta_1 = \beta_2 = 1.0$ or $\beta_1 = 1.5$ and $\beta_2 = 0.5$ is used. The standard errors of these estimates are also found to be small.

In summary, the presence of outliers adversely affects the estimate of both β and ρ parameters under the longitudinal model for the count data. In the binary case, the outliers do not appear to affect the estimation of ρ parameter, but they adversely affect the estimation of the main parameter β.

4.3 Robust GQL Estimation

In the last section it was demonstrated that the GQL estimation for the regression effect β both under the count and binary longitudinal models were adversely affected by the presence of outliers. Also, outliers were found to negatively influence the estimation of the correlation parameter ρ under the longitudinal count model, whereas ρ estimation was interestingly not found to be affected under the longitudinal binary model. This raises a concern to estimate the regression effect β in a robust way by downweighting the outlying responses so that the estimates can be consistent. This is needed to be done for both count and binary cases. As far as the consistent estimation of the ρ parameter is concerned, we use a Pearsonian type correlation formula by exploiting the downweighting responses. Note that in the β estimation, one can, however, avoid the direct estimate of ρ.

4.3.1 Robust GQL estimating equation for β

For both count and binary data, we may use the common RGQL estimating equation (3.19) that we have used under the familial model. Recall that Ω_i weight matrix in (3.19) has the form

$$\Omega_i = A_{i\xi}^{\frac{1}{2}} C_{i\xi} A_{i\xi}^{\frac{1}{2}}, \tag{4.28}$$

where $A_{i\xi} = diag[var(\psi_c(r_{i1})), \ldots, var(\psi_c(r_{it})), \ldots, var(\psi_c(r_{iT}))]$ and $C_{i\xi} = (c_{i,ut,\xi})$, with $c_{i,ut,\xi} = corr[\psi_c(r_{iu}), \psi_c(r_{it})]$ for $u, t = 1, \ldots, T$ and $\psi_c(r_{it})$ being the downweighting function. Note that the elements of Ω_i matrix were, however, computed directly in the appendix under the familial count and binary models. The direct computation for the elements of $C_{i\xi}$ matrix does not appear to be easy under the longitudinal models, especially under the longitudinal count data model. For convenience, we make an assumption that $C_{i\xi}$ is a constant matrix for all $i = 1, \ldots, K$, i.e., $C_{i\xi} \equiv C_\xi^*$, and estimate this matrix as

$$\hat{C}_\xi^* = (\hat{c}_{ut,\xi}), \tag{4.29}$$

where

$$\hat{c}_{ut,\xi} = \frac{\sum_{i=1}^K [\psi_c(r_{iu}) - \bar{\xi}_u][\psi_c(r_{it}) - \bar{\xi}_t]}{\sqrt{\sum_{i=1}^K [\psi_c(r_{iu}) - \bar{\xi}_u]^2 \sum_{i=1}^K [\psi_c(r_{it}) - \bar{\xi}_t]^2}}, \tag{4.30}$$

where

$$\bar{\xi}_t = \frac{1}{K}\sum_{i=1}^{K}\psi_c(r_{it}).$$

Now, by computing the estimate of the correlation matrix C_ξ^* as given in (4.29), we write the RGQL estimating equation (3.19) for β under the longitudinal models as

$$\sum_{i=1}^{K}\left[W_i\frac{\partial}{\partial\beta}\left\{\xi_i - K^{-1}\sum_{i=1}^{K}\lambda_i\right\}'\left[A_{i\xi}^{\frac{1}{2}}\hat{C}_\xi^* A_{i\xi}^{\frac{1}{2}}\right]^{-1}\left\{\xi_i - K^{-1}\sum_{i=1}^{K}\lambda_i\right\}\right] = 0. \qquad (4.31)$$

Note that in (4.31), the formula for $\xi_i = [\psi_c(r_{i1}), \ldots, \psi_c(r_{it}), \ldots, \psi_c(r_{iT})]'$, i.e., $\psi_c(r_{it})$ remains the same as in the familial models. More specifically, under the longitudinal count model, the formula for the downweighting function $\psi_c(r_{it})$ will be the same as in (3.15) given for the familial count model. Similarly, under the longitudinal binary model, the formulas of $\psi_c(r_{it})$ for the one sided upper and lower, and two sided outliers are given in (3.16), (3.17), and (3.18), respectively. Note however that $\tilde{\mu}_{it}$ ($\tilde{\mu}_{ij}$ in the familial model) and $\tilde{\sigma}_{itt}$ ($\tilde{\sigma}_{ijj}$ in the familial model) have the formulas

$$\tilde{\mu}_{it} = \tilde{\sigma}_{itt} = exp(\tilde{x}_{it}'\beta), \qquad (4.32)$$

under the longitudinal count model, and

$$\tilde{\mu}_{it} = \frac{exp(\tilde{x}_{it}'\beta)}{1 + exp(\tilde{x}_{it}'\beta)} \text{ and } \tilde{\sigma}_{itt} = \tilde{\mu}_{it}(1 - \tilde{\mu}_{it}), \qquad (4.33)$$

respectively, under the longitudinal binary model.

Note that in (4.31), $\lambda_i = E(\xi_i) = (\lambda_{i1}, \ldots, \lambda_{it}, \ldots, \lambda_{iT})'$, which may be computed from following (2.8) for the longitudinal count data. Also, for the two sided binary outliers case, λ_{it} for all $t = 1, \ldots, T$, can be computed by (2.15). Similarly, one can compute λ_i for the upper or lower sided outliers.

Next, for $t = 1, \ldots, T$, $var[\psi_c(r_{it})]$ under the count model may be computed by (2.9) and under the two sided binary outliers case, this may be computed by (2.16).

Furthermore, the derivatives of $\psi_c(r_{it})$ and λ_{it} with respect to β may be computed following section 2.1.1 for the count data case and section 2.1.2 for the two sided binary outliers case.

4.3.2 RM estimation of ρ parameter

The direct estimation of ρ parameter under the longitudinal model with outliers is different. Since it was found that outliers did not affect the estimate of ρ parameter in the longitudinal binary case, we have chosen to estimate this AR(1) type correlation parameter by using lag 1 correlations constructed based on the downweighting responses. Thus, we estimate ρ by

$$\hat{\rho} = \frac{\frac{\sum_{i=1}^{K}\sum_{u=1}^{T-1}[\psi_c(r_{iu})w_{iu}-\bar{\xi}_{u,w}][\psi_c(r_{i,u+1})w_{i,u+1}-\bar{\xi}_{u+1,w}]}{K(T-1)}}{\frac{\sum_{i=1}^{K}\sum_{u=1}^{T}[\psi_c(r_{iu})w_{iu}-\bar{\xi}_{u,w}]^2}{KT}}, \tag{4.34}$$

where

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

We refer to (4.34) as the RM estimator of the ρ parameter. Note that we attempt to use the same formula for the count data case.

In the following subsection, we conduct a simulation study to examine the performance of the estimates of β and ρ obtained from (4.31) and (4.34), respectively.

4.4 Simulation Study

Recall from sections 4.1.4 and 4.2.4 that in general, the unmodified GQL and MM approaches for the count and binary data produce inconsistent estimates for the regression parameter β and the correlation parameter ρ in the presence of one or more outliers. We now conduct a simulation study to examine the performance of the RGQL estimating equation (4.31) to estimate β and the RM estimator (4.34) for ρ. We conduct this simulation study both for the longitudinal count and binary data in the presence of outliers.

4.4.1 Count case

In this case, we first generate the longitudinal count data with outliers following section 4.1.2. As far as the value of the tuning constant is concerned, we choose $c = 1.55$. Next, we exploit the RGQL estimating equation (4.31) to estimate $\beta = (\beta_1, \beta_2)'$ and the RM estimator (4.34) to estimate ρ, under each of the 1000 simulations. The SM, SSE, and MSE for these three estimates are computed from 1000 values of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\rho}$, which are reported in Table 4.3.

It is clear that the RGQL estimation approach produce almost unbiased estimates for the regression parameter β. It is also found that this approach appears to perform well under the model with $\beta_1 = \beta_2 = 1.0$ as compared to that of the model with $\beta_1 = \beta_2 = 0.5$. For example, for $\rho = 0.5$, the estimates of β_1 and β_2 are found to be 1.021 and 0.988, respectively, with corresponding standard errors 0.127 and 0.051 when the true $\beta_1 = \beta_2 = 1.0$; whereas these estimates are found to be 0.535 and 0.444, respectively, with standard errors 0.153 and 0.121 for true $\beta_1 = \beta_2 = 0.5$.

As far as the estimation of the correlation parameter ρ is concerned, for the both longitudinal count models, i.e., models with $\beta_1 = \beta_2 = 0.5$ and $\beta_1 = \beta_2 = 1.0$, the RM approach appears to produce unbiased estimate for ρ with small standard error. For example, when $\beta_1 = \beta_2 = 0.5$ and $\rho = 0.8$, the estimate of ρ is 0.786 with standard error 0.041; and this estimate is 0.788 with standard error 0.041 for the model with $\beta_1 = \beta_2 = 1.0$ and $\rho = 0.8$.

4.4.2 Binary case

To generate the longitudinal binary data with outliers, we follow the same procedure that was considered in section 4.2.2. We also consider the same covariates and parameter values as in section 4.2.2. As far as the value of the tuning constant related probability is concerned, we choose $\mu^{c_1} = 0.5,\ 0.6,$ and 0.9. We then use the RGQL estimating equation (4.31) to estimate the regression parameters β_1 and β_2, and the RM estimator (4.34) to estimate ρ parameter. This we do for 1000 simulations. The SM, SSE, and MSE computed from 1000 values of these three estimates are reported in Table 4.4.

It appears from this table that the RGQL approach produces almost unbiased estimates for the regression effects β, when the tuning constant related probability is small such as $\mu^{c_1} = 0.5$. Note that as expected, the biasness in the estimates of β increases as the value of μ^{c_1} increases. In practice, it is however not recommended to use μ^{c_1} far away from 0.5. The standard errors of these estimates are found to be relatively larger as compared to those under the longitudinal count model.

With regard to the estimation of the ρ parameter, the RM approach is found to yield almost unbiased estimate of ρ irrespective of the values of μ^{c_1}. Similar to the count data case, the standard error of the estimate of ρ is also small.

Table 4.1: Simulated mean (SM), standard errors (SSE), and mean squared errors (MSE) of the **GQL** estimates for the regression parameters and the **MM** estimates of the correlation parameter of the **Poisson longitudinal model** when data contain $m = 4$ outliers, for the selected values of ρ; $K = 100$; $T = 4$; 1000 simulations.

Regression effects (β_1, β_2)	Correlation parameter(ρ)	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$
(0.5, 0.5)	0.25	SM	0.084	0.369	0.352
		SSE	0.055	0.034	0.071
		MSE	0.176	0.018	0.016
	0.50	SM	-0.023	0.343	0.598
		SSE	0.069	0.043	0.064
		MSE	0.278	0.027	0.014
	0.80	SM	-0.184	0.312	0.857
		SSE	0.075	0.053	0.035
		MSE	0.474	0.038	0.004
(1.0, 1.0)	0.25	SM	-0.509	0.723	0.847
		SSE	0.857	0.274	0.060
		MSE	3.011	0.152	0.360
	0.50	SM	-0.554	0.725	0.903
		SSE	0.094	0.034	0.034
		MSE	2.424	0.163	0.163
	0.80	SM	-0.640	0.700	0.964
		SSE	0.512	0.573	0.016
		MSE	2.951	0.418	0.027

Table 4.2: Simulated mean (SM), standard errors (SSE), and mean squared errors (MSE) of the **GQL** estimates for the regression parameters and the **MM** estimates of the correlation parameter of the **binary longitudinal model** when data contain $m = 4$ outliers, for the selected values of ρ; $K = 100$; $T = 4$; 1000 simulations.

Regression effects (β_1, β_2)	Correlation parameter(ρ)	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$
(1.0, 1.0)	0.25	SM	0.583	1.264	0.244
		SSE	0.294	0.405	0.070
		MSE	0.260	0.234	0.005
	0.50	SM	0.258	1.462	0.500
		SSE	0.297	0.453	0.070
		MSE	0.638	0.419	0.005
	0.80	SM	-0.346	1.869	0.816
		SSE	0.239	0.503	0.042
		MSE	1.868	1.008	0.002
(1.5, 0.5)	0.25	SM	0.966	0.828	0.238
		SSE	0.317	0.404	0.075
		MSE	0.386	0.271	0.006
	0.50	SM	0.555	1.076	0.499
		SSE	0.352	0.465	0.081
		MSE	1.017	0.548	0.007
	0.80	SM	-0.215	1.592	0.824
		SSE	0.249	0.496	0.043
		MSE	3.003	1.439	0.002

Table 4.3: Simulated mean (SM), standard errors (SSE), and mean squared errors (MSE) of the **RGQL** estimates for the regression parameters and the **RM** estimates of the correlation parameter of the **Poisson longitudinal model** when data contain $m = 4$ outliers, for the selected values of ρ; tuning constant $c = 1.55$; $K = 100$; $T = 4$; 1000 simulations.

Regression effects (β_1, β_2)	Correlation parameter(ρ)	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$
(0.5, 0.5)	0.25	SM	0.528	0.462	0.238
		SSE	0.101	0.064	0.062
		MSE	0.011	0.006	0.004
	0.50	SM	0.535	0.444	0.486
		SSE	0.153	0.121	0.059
		MSE	0.025	0.018	0.004
	0.80	SM	0.533	0.483	0.786
		SSE	0.268	0.168	0.041
		MSE	0.073	0.029	0.002
(1.0, 1.0)	0.25	SM	1.024	0.999	0.240
		SSE	0.080	0.034	0.061
		MSE	0.007	0.001	0.004
	0.50	SM	1.021	0.988	0.484
		SSE	0.127	0.051	0.060
		MSE	0.017	0.003	0.004
	0.80	SM	1.006	0.988	0.788
		SSE	0.202	0.075	0.041
		MSE	0.041	0.006	0.002

Table 4.4: Simulated mean (SM), standard errors (SSE), and mean squared errors (MSE) of the **RGQL** estimates for the regression parameters and the **RM** estimates of the correlation parameter of the **binary longitudinal model** when data contain $m = 4$ outliers, for the selected values of ρ and tuning probability (μ^{c_1}); $K = 100$; $T = 4$; 1000 simulations.

Regression effects (β_1, β_2)	Correlation parameter (ρ)	Tuning probability (μ^{c_1})	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$
(1.0, 1.0)	0.25	0.5	SM	0.996	0.987	0.241
			SSE	0.307	0.407	0.068
			MSE	0.094	0.166	0.005
		0.6	SM	0.988	0.982	0.241
			SSE	0.305	0.404	0.068
			MSE	0.093	0.164	0.005
		0.9	SM	0.933	0.953	0.241
			SSE	0.302	0.399	0.068
			MSE	0.096	0.161	0.005
	0.5	0.5	SM	0.963	0.999	0.486
			SSE	0.332	0.455	0.066
			MSE	0.111	0.207	0.005
		0.6	SM	0.943	0.995	0.486
			SSE	0.330	0.453	0.065
			MSE	0.112	0.205	0.005
		0.9	SM	0.827	0.974	0.486
			SSE	0.320	0.439	0.065
			MSE	0.132	0.194	0.005
	0.8	0.5	SM	0.940	1.009	0.784
			SSE	0.358	0.492	0.046
			MSE	0.132	0.242	0.002
		0.6	SM	0.933	0.993	0.783
			SSE	0.349	0.480	0.045
			MSE	0.126	0.230	0.002
		0.9	SM	0.755	0.902	0.780
			SSE	0.311	0.436	0.046
			MSE	0.157	0.199	0.003

Cont...........Table 4.4

Regression effects (β_1, β_2)	Correlation parameter (ρ)	Tuning probability (μ^{c_1})	Statistic	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$
(1.5, 0.5)	0.25	0.5	SM	1.500	0.486	0.240
			SSE	0.329	0.403	0.071
			MSE	0.108	0.162	0.005
		0.6	SM	1.490	0.482	0.240
			SSE	0.328	0.401	0.071
			MSE	0.108	0.161	0.005
		0.9	SM	1.427	0.459	0.241
			SSE	0.320	0.394	0.071
			MSE	0.107	0.157	0.005
	0.5	0.5	SM	1.477	0.491	0.484
			SSE	0.370	0.467	0.071
			MSE	0.138	0.219	0.005
		0.6	SM	1.454	0.489	0.484
			SSE	0.368	0.465	0.071
			MSE	0.138	0.217	0.005
		0.9	SM	1.336	0.466	0.483
			SSE	0.350	0.445	0.069
			MSE	0.149	0.199	0.005
	0.8	0.5	SM	1.452	0.522	0.781
			SSE	0.395	0.498	0.049
			MSE	0.158	0.249	0.003
		0.6	SM	1.428	0.501	0.779
			SSE	0.381	0.483	0.048
			MSE	0.150	0.234	0.003
		0.9	SM	1.267	0.399	0.773
			SSE	0.326	0.435	0.048
			MSE	0.161	0.199	0.003

Chapter 5

Conclusion

To reduce the influence of possible outliers, in this thesis, we have considered robust inferences in the GLMs for the independent; GLMMs for the familially correlated; and GLMs for the longitudinally correlated, count and binary data.

In the independence set up, we have proposed a FSMQL approach, which produces regression effects with smaller biases as compared to the existing MQL approach considered by Cantoni and Ronchetti (2001). It has been demonstrated that the FSMQL approach produces uniformly better estimates than the MQL approach, both for the count and binary data with outliers, in the independence set up. Note that for the binary case in particular, we have given a new definition for outliers, which accommodates one sided (upper or lower) as well as two sided outliers, whereas in the literature, only two sided outliers are defined.

For the case, when the count or binary data including one or more outliers occur in a clustered form, correlations of the responses in a cluster may arise either because (1) the cluster contains responses from the members of a family causing familial correlations or (2) the cluster contains repeated responses from the same individual causing longitudinal correlations. Some of the existing studies such as Mills et al. (2002) and Sinha (2006) have used the familial, that is, random effects approach to model the longitudinal correlations and developed the robust inferences based on such familial correlation structures for the longitudinal data. In the present thesis, as opposed to the these studies, we have clearly demonstrated how one can develop a proper correlation structure under both familial and longitudinal set up. These structures are then exploited to develop the RGQL approach both in the familial and longitudinal set up, for the consistent estimation of the parameters of the underlying model.

The simulation studies show that the proposed RGQL approach works well in estimating the parameters of both familial and longitudinal models for both count and binary data. These estimation procedures should be useful to the practitioners dealing with inferences for the count or binary data in the presence of outliers.

Remark that in some situations, one may encounter the repeated count or binary data collected from the members of a large number of independent families. It may also happen that this type of familial-longitudinal data contains one or more outliers. One may attempt to combine our RGQL approaches proposed for the familial and longitudinal models to analyze the familial-longitudinal data in the presence of outliers. This type of robust analysis for the combined data is however beyond the scope of the present thesis.

Appendix A

Formulas for λ_i, $\frac{\partial}{\partial\beta}\xi_i'$, $\frac{\partial}{\partial\beta}\lambda_i'$, and Ω_i to construct (3.19) for the count data

Derivation for $\lambda_i = E(\xi_i)$

Note that in (3.19) $\xi_i = [\psi_c(r_{i1}), \ldots, \psi_c(r_{ij}), \ldots, \psi_c(r_{in_i})]'$, where $\psi_c(r_{ij})$ is defined as in (3.15) with $r_{ij} = \frac{y_{ij}-\tilde{\mu}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}}$. To derive its expectation, i.e., $\lambda_i = E(\xi_i) = E[\lambda_{i1}, \ldots, \lambda_{ij}, \ldots, \lambda_{in_i}]'$, one simply needs to compute λ_{ij} as

$$\begin{aligned}\lambda_{ij} &= E[\psi_c(r_{ij})] \\ &= \sum_{y_{ij}=0}^{\infty} \psi_c(r_{ij}) f(y_{ij}), \end{aligned} \tag{A.1}$$

where

$$f(y_{ij}) = \int_{-\infty}^{\infty} P(y_{ij}|\gamma_i)\phi(\gamma_i)\partial\gamma_i,$$

with $P(y_{ij}|\gamma_i)$ being the Poisson probability mass function with parameter $\tilde{\mu}^*_{ij} = exp(\tilde{x}'_{ij}\beta - \frac{\sigma^2}{2} + \sigma\gamma_i)$ and $\phi(\cdot)$ being the standard normal density function. Now, by using the simulated integration approach [Fahrmeir and Tutz (1994)], (A.1) may be re-expressed as

$$\lambda_{ij} = \frac{1}{M}\sum_{\ell=1}^{M} a^*_{ij,\ell}, \tag{A.2}$$

where

$$
\begin{aligned}
a^*_{ij,\ell} &= \sum_{y_{ij}=0}^{\infty} \psi_c(r_{ij}) P(y_{ij}|\gamma_{i\ell}) \\
&= c\left[1 - F_{Y_{ij}|\gamma_{i\ell}}(I^U_{ij}) - F_{Y_{ij}|\gamma_{i\ell}}(I^L_{ij})\right] \\
&+ \frac{\tilde{\mu}^*_{ij,\ell}}{\sqrt{\tilde{\sigma}_{ijj}}}\left[F_{Y_{ij}|\gamma_{i\ell}}(I^U_{ij}-1) - F_{Y_{ij}|\gamma_{i\ell}}(I^L_{ij}-1)\right] \\
&- \frac{\tilde{\mu}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}}\left[F_{Y_{ij}|\gamma_{i\ell}}(I^U_{ij}) - F_{Y_{ij}|\gamma_{i\ell}}(I^L_{ij})\right],
\end{aligned} \tag{A.3}
$$

[Cantoni and Ronchetti (2001)]. In (A.3), $\tilde{\mu}^*_{ij,\ell} = exp(\tilde{x}'_{ij}\beta - \frac{\sigma^2}{2} + \sigma\gamma_{i\ell})$ is obtained from $\tilde{\mu}^*_{ij}$ by replacing γ_i with $\gamma_{i\ell}$, $\ell = 1, \ldots, M$ and $F_{Y_{ij}|\gamma_{i\ell}}(z_{ij})$ is the cumulative probability function Y_{ij} conditional on $\gamma_{i\ell}$ given by

$$F_{Y_{ij}|\gamma_{i\ell}}(z_{ij}) = \sum_{y_{ij}=0}^{z_{ij}} P(y_{ij}|\gamma_{i\ell})$$

Also, in (A.3), I^U_{ij} and I^L_{ij} are the nearest integer values of $\tilde{\mu}_{ij}+c\sqrt{\tilde{\sigma}_{ijj}}$ and $\tilde{\mu}_{ij}-c\sqrt{\tilde{\sigma}_{ijj}}$, respectively with $\tilde{\mu}_{ij}$ and $\tilde{\sigma}_{ijj}$ being defined as in (3.4) and (3.5), respectively.

Derivation for $\frac{\partial}{\partial\beta}\xi'_i$

To compute the $p \times n_i$ derivative matrix $\frac{\partial}{\partial\beta}\xi'_i$, it is sufficient to calculate $\frac{\partial}{\partial\beta}\psi_c(r_{ij})$. For the Poison data case, the gradient function of the downweighting function $\psi_c(r_{ij})$ with respect to β may be obtained from (3.15) as

$$\frac{\partial}{\partial\beta}\psi_c(r_{ij}) = \begin{cases} -\frac{1}{\sqrt{\tilde{\sigma}_{ijj}}}\tilde{\mu}_{ij}\tilde{x}_{ij}, & |r_{ij}| \le c, \\ 0, & |r_{ij}| > c. \end{cases} \tag{A.4}$$

Derivation for $\frac{\partial}{\partial\beta}\lambda'_i$

For the computation of the derivative matrix $\frac{\partial}{\partial\beta}\lambda_i'$, one may need to calculate $\frac{\partial}{\partial\beta}\lambda_{ij}$. The gradient function of λ_{ij} may be obtained from (A.2), which is given by

$$\frac{\partial}{\partial\beta}\lambda_{ij} = \frac{1}{M}\sum_{\ell=1}^{M}\frac{\partial}{\partial\beta}a^*_{ij,\ell}, \tag{A.5}$$

where

$$\begin{aligned}\frac{\partial}{\partial\beta}a^*_{ij,\ell} &= -c\left[\frac{\partial}{\partial\beta}F_{Y_{ij}|\gamma_{i\ell}}(I^U_{ij}) + \frac{\partial}{\partial\beta}F_{Y_{ij}|\gamma_{i\ell}}(I^L_{ij})\right] \\ &+ \frac{\tilde{\mu}^*_{ij,\ell}\tilde{x}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}}\left[F_{Y_{ij}|\gamma_{i\ell}}(I^U_{ij}-1) - F_{Y_{ij}|\gamma_{i\ell}}(I^L_{ij}-1)\right] \\ &+ \frac{\tilde{\mu}^*_{ij,\ell}}{\sqrt{\tilde{\sigma}_{ijj}}}\left[\frac{\partial}{\partial\beta}F_{Y_{ij}|\gamma_{i\ell}}(I^U_{ij}-1) - \frac{\partial}{\partial\beta}F_{Y_{ij}|\gamma_{i\ell}}(I^L_{ij}-1)\right] \\ &- \frac{\tilde{\mu}_{ij}\tilde{x}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}}\left[F_{Y_{ij}|\gamma_{i\ell}}(I^U_{ij}) - F_{Y_{ij}|\gamma_{i\ell}}(I^L_{ij})\right] \\ &- \frac{\tilde{\mu}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}}\left[\frac{\partial}{\partial\beta}F_{Y_{ij}|\gamma_{i\ell}}(I^U_{ij}) - \frac{\partial}{\partial\beta}F_{Y_{ij}|\gamma_{i\ell}}(I^L_{ij})\right],\end{aligned} \tag{A.6}$$

with, for example,

$$\frac{\partial}{\partial\beta}F_{Y_{ij}|\gamma_{i\ell}}(I^U_{ij}) = \sum_{k=1}^{I^U_{ij}}\frac{\partial}{\partial\beta}P(Y_{ij}=k|\gamma_{i\ell}),$$

and

$$\frac{\partial}{\partial\beta}P(Y_{ij}=k|\gamma_{i\ell}) = P(Y_{ij}=k|\gamma_{i\ell})(k-\tilde{\mu}^*_{ij,\ell})\tilde{x}_{ij}.$$

Derivation for Ω_i

To compute the $n_i \times n_i$ covariance matrix $\Omega_i = cov(\xi_i)$, it is sufficient to show how to compute $var[\psi_c(r_{ij})]$, $i = 1, \ldots, K$, $j = 1, \ldots, n_i$, and $cov[\psi_c(r_{ij}), \psi_c(r_{ik})]$, $j \neq k$, which can be computed by using the formulas

$$var[\psi_c(r_{ij})] = E[\psi_c^2(r_{ij})] - [E\{\psi_c(r_{ij})\}]^2, \tag{A.7}$$

and

$$cov[\psi_c(r_{ij}), \psi_c(r_{ik})] = E[\psi_c(r_{ij})\psi_c(r_{ik})] - E[\psi_c(r_{ij})]E[\psi_c(r_{ik})], \tag{A.8}$$

respectively. Now, following equation (A.2), the expectation of $\psi_c^2(r_{ij})$ can be written as

$$E[\psi_c^2(r_{ij})] = \frac{1}{M}\sum_{\ell=1}^{M} b^*_{ij,\ell}, \tag{A.9}$$

where $b^*_{ij,\ell}$, after some algebras, reduces to the form

$$\begin{aligned} b^*_{ij,\ell} &= \sum_{y_{ij}=0}^{\infty} \psi_c^2(r_{ij})P(y_{ij}|\gamma_{i\ell}) \\ &= c^2\left[1 - F_{Y_{ij}|\gamma_{i\ell}}(I_{ij}^U) + F_{Y_{ij}|\gamma_{i\ell}}(I_{ij}^L)\right] \\ &+ \frac{\tilde{\mu}_{ij}^2}{\tilde{\sigma}_{ijj}}\left[F_{Y_{ij}|\gamma_{i\ell}}(I_{ij}^U) - F_{Y_{ij}|\gamma_{i\ell}}(I_{ij}^L)\right] \\ &- \frac{2\tilde{\mu}_{ij}\tilde{\mu}^*_{ij,\ell}}{\tilde{\sigma}_{ijj}}\left[F_{Y_{ij}|\gamma_{i\ell}}(I_{ij}^U - 1) - F_{Y_{ij}|\gamma_{i\ell}}(I_{ij}^L - 1)\right] \\ &+ \frac{\tilde{\mu}^*_{ij,\ell}}{\tilde{\sigma}_{ijj}}\left[F_{Y_{ij}|\gamma_{i\ell}}(I_{ij}^U - 1) - F_{Y_{ij}|\gamma_{i\ell}}(I_{ij}^L - 1)\right] \\ &+ \frac{\tilde{\mu}^*_{ij,\ell}\tilde{\mu}^*_{ij,\ell}}{\tilde{\sigma}_{ijj}}\left[F_{Y_{ij}|\gamma_{i\ell}}(I_{ij}^U - 2) - F_{Y_{ij}|\gamma_{i\ell}}(I_{ij}^L - 2)\right], \end{aligned} \tag{A.10}$$

[Cantoni and Ronchetti (2001)]. Next, for any pair, the uncondition expectation of product of $\psi_c(r_{ij})$ and $\psi_c(r_{ij})$ may be derived as

$$E[\psi_c(r_{ij})\psi_c(r_{ik})] = \sum_{y_{ij}=0}^{\infty}\sum_{y_{ik}=0}^{\infty} \psi_c(r_{ij})\psi_c(r_{ik})f(y_{ij}, y_{ik}), \tag{A.11}$$

where $f(y_{ij}, y_{ik})$ is the joint probability function of Y_{ij} and Y_{ik}. Since for given γ_i, count responses y_{ij} and y_{ik}, $(j \neq k)$, are independent, this joint probability function can be computed as

$$f(y_{ij}, y_{ik}) = \int_{-\infty}^{\infty} P(y_{ij}|\gamma_i)P(y_{ik}|\gamma_i)\phi(\gamma_i)\partial\gamma_i. \tag{A.12}$$

Consequently, equation (A.11) can be re-expressed as

$$E[\psi_c(r_{ij})\psi_c(r_{ik})] = \frac{1}{M}\sum_{\ell=1}^{M} a^*_{ij,\ell}a^*_{ik,\ell}, \tag{A.13}$$

where $a^*_{ij,\ell}$ is given as in (A.3).

Formulas for λ_i, $\frac{\partial}{\partial\beta}\xi'_i$, $\frac{\partial}{\partial\beta}\lambda'_i$, and Ω_i to construct (3.19) for the binary data

Here, the formulas for $E(\xi_i)$, $\frac{\partial}{\partial\beta}\xi'_i$, $\frac{\partial}{\partial\beta}E(\xi'_i)$, and Ω_i are given only for the binary data with a two sided outlier. These formulas in the one sided outlier cases may be obtained as the special cases of the two sided outlier case. Now, for the two sided outlier case, we consider the downweighting function $\psi_c(r_{ij})$, $i = 1, \ldots, K$, $j = 1, \ldots, n_i$, given as in (3.18). It is assumed that conditional on the random effect γ_i, the response y_{ij} follows a binary distribution with the observed success probability $\tilde{\mu}^*_{ij} = \frac{exp(\tilde{x}'_{ij}\beta+\gamma_i\sigma)}{1+exp(\tilde{x}'_{ij}\beta+\gamma_i\sigma)}$.

<u>Derivation for $\lambda_i = E(\xi_i)$</u>

Similar to the count data case, $\lambda_{ij} = E[\psi_c(r_{ij})]$ by (3.18) can be obtained as

$$\lambda_{ij} = \sum_{y_{ij}=0}^{1}\left[P_1\frac{y_{ij}-\mu^{(c_1)}_{ij}}{\sqrt{\sigma^{(c_1)}_{ijj}}} + P_2\frac{y_{ij}-\tilde{\mu}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}} + P_3\frac{y_{ij}-\mu^{(c_2)}_{ij}}{\sqrt{\sigma^{(c_2)}_{ijj}}}\right] f(y_{ij}), \tag{A.14}$$

where $\tilde{\mu}_{ij}$ and $\tilde{\sigma}_{ijj}$ are defined as in (3.6) and (3.7), respectively. Note that in (A.14), P_1, P_2, and P_3 are the probabilities for a binary observation to satisfy the conditions $P(Y_{ij} = 1) > p_{sb}$, $p_{lb} \leq P(Y_{ij} = 1) \leq p_{sb}$, and $P(Y_{ij} = 1) < p_{lb}$, respectively, where, for example, one may compute P_1 as

$$P_1 = \frac{\text{Number of observations satisfying} P(Y_{ij} = 1|\tilde{x}_{ij}) > p_{sb}}{\text{Total observation}(\sum_{i=1}^{K} n_i)}.$$

Since $f(y_{ij}) = \int_{-\infty}^{\infty} P(y_{ij}|\gamma_i)\phi(\gamma_i)\partial\gamma_i$ with $P(y_{ij}|\gamma_i)$ being the binary probability mass function with parameter $\tilde{\mu}_{ij}^*$ conditional on γ_i, equation (A.14) can be re-expressed as

$$\begin{aligned}\lambda_{ij} &= \frac{1}{M}\sum_{\ell=1}^{M} a_{ij,\ell}^* \\ &= P_1\frac{\tilde{\mu}_{ij} - \mu_{ij}^{(c_1)}}{\sqrt{\sigma_{ijj}^{(c_1)}}} + P_3\frac{\tilde{\mu}_{ij} - \mu_{ij}^{(c_2)}}{\sqrt{\sigma_{ijj}^{(c_2)}}},\end{aligned} \tag{A.15}$$

where $a_{ij,\ell}^*$ can be written as

$$\begin{aligned}a_{ij,\ell}^* &= \sum_{y_{ij}=0}^{1}\left[P_1\frac{y_{ij} - \mu_{ij}^{(c_1)}}{\sqrt{\sigma_{ijj}^{(c_1)}}} + P_2\frac{y_{ij} - \tilde{\mu}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}} + P_3\frac{y_{ij} - \mu_{ij}^{(c_2)}}{\sqrt{\sigma_{ijj}^{(c_2)}}}\right] P(y_{ij}|\gamma_{i\ell}) \\ &= P_1\frac{\tilde{\mu}_{ij,\ell}^* - \mu_{ij}^{(c_1)}}{\sqrt{\sigma_{ijj}^{(c_1)}}} + P_2\frac{\tilde{\mu}_{ij,\ell}^* - \tilde{\mu}_{ij}}{\sqrt{\tilde{\sigma}_{ijj}}} + P_3\frac{\tilde{\mu}_{ij,\ell}^* - \mu_{ij}^{(c_2)}}{\sqrt{\sigma_{ijj}^{(c_2)}}}.\end{aligned} \tag{A.16}$$

Derivation for $\frac{\partial}{\partial\beta}\xi_i'$

For the binary case, by (3.18), one may obtain the gradient function of the down-weighting function $\psi_c(r_{ij})$ with respect to β as

$$\frac{\partial}{\partial\beta}\psi_c(r_{ij}) = \begin{cases} 0, & P(Y_{ij}=1|\tilde{x}_{ij}) > p_{sb}, (i,j) \equiv (i',j'), \\ -\frac{1}{\sqrt{\tilde{\sigma}_{ijj}}}\frac{\partial}{\partial\beta}\tilde{\mu}_{ij}, & p_{lb} \le P(Y_{ij}=1|\tilde{x}_{ij}) \le p_{sb}, (i,j) \not\equiv (i',j'), \\ 0, & P(Y_{ij}=1|\tilde{x}_{ij}) < p_{lb}, (i,j) \equiv (i',j'), \end{cases} \tag{A.17}$$

where

$$\frac{\partial}{\partial\beta}\tilde{\mu}_{ij} = \frac{1}{M}\sum_{\ell=1}^{M}\frac{\partial}{\partial\beta}\tilde{\mu}_{ij,\ell}^*, \tag{A.18}$$

with

$$\frac{\partial}{\partial\beta}\tilde{\mu}_{ij,\ell}^* = \tilde{\mu}_{ij,\ell}^*(1-\tilde{\mu}_{ij,\ell}^*)\tilde{x}_{ij}.$$

Derivation for $\frac{\partial}{\partial\beta}\lambda_i'$

Next, to compute the gradient function of the expectation of $\psi_c(r_{ij})$ with respect to β, we exploit (A.15) and write

$$\frac{\partial}{\partial\beta}\lambda_{ij} = \left[\frac{P_1}{\sqrt{\sigma_{ijj}^{(c_1)}}} + \frac{P_3}{\sqrt{\sigma_{ijj}^{(c_2)}}}\right]\frac{\partial}{\partial\beta}\tilde{\mu}_{ij}, \tag{A.19}$$

where $\frac{\partial}{\partial\beta}\tilde{\mu}_{ij}$ is given as in (A.18).

Derivation for Ω_i

Similar to the count data case, we compute the variance-covariance matrix $\Omega_i = cov(\xi_i)$ for the binary data by computing $E[\psi_c^2(r_{ij})]$ and $E[\psi_c(r_{ij})\psi_c(r_{ik})]$, $(j \neq k)$. The expectation of $\psi_c^2(r_{ij})$ can be written as

$$\begin{aligned} E[\psi_c^2(r_{ij})] &= \sum_{y_{ij}=0}^{1}\left[P_1\frac{(y_{ij}-\mu_{ij}^{(c_1)})^2}{\sigma_{ijj}^{(c_1)}} + P_2\frac{(y_{ij}-\tilde{\mu}_{ij})^2}{\tilde{\sigma}_{ijj}} + P_3\frac{(y_{ij}-\mu_{ij}^{(c_2)})^2}{\sigma_{ijj}^{(c_2)}}\right]f(y_{ij}) \\ &= \frac{1}{M}\sum_{\ell=1}^{M} b_{ij,\ell}^* \\ &= P_1\frac{\tilde{\mu}_{ij}(1-2\mu_{ij}^{(c_1)}) + \mu_{ij}^{(c_1)2}}{\sigma_{ijj}^{(c_1)}} + P_2 + P_3\frac{\tilde{\mu}_{ij}(1-2\mu_{ij}^{(c_2)}) + \mu_{ij}^{(c_2)2}}{\sigma_{ijj}^{(c_2)}}, \end{aligned} \tag{A.20}$$

where

$$\begin{aligned} b_{ij,\ell}^* &= P_1\frac{\tilde{\mu}_{ij,\ell}^* - 2\tilde{\mu}_{ij,\ell}^*\mu_{ij}^{(c_1)} + \mu_{ij}^{(c_1)2}}{\sigma_{ijj}^{(c_1)}} + P_2\frac{\tilde{\mu}_{ij,\ell}^* - 2\tilde{\mu}_{ij,\ell}^*\tilde{\mu}_{ij} + \tilde{\mu}_{ij}^2}{\tilde{\sigma}_{ijj}} \\ &+ P_3\frac{\tilde{\mu}_{ij,\ell}^* - 2\tilde{\mu}_{ij,\ell}^*\mu_{ij}^{(c_2)} + \mu_{ij}^{(c_2)2}}{\sigma_{ijj}^{(c_2)}} \end{aligned} \tag{A.21}$$

Again, since conditional on γ_i, responses y_{ij} and y_{ik} are independent, the unconditional expectation of the product of $\psi_c(r_{ij})$ and $\psi_c(r_{ik})$ can be obtained as

$$\begin{aligned}
E[\psi_c(r_{ij})\psi_c(r_{ik})] &= \frac{1}{M}\sum_{\ell=1}^{M} a^*_{ij,\ell}a^*_{ik,\ell} \\
&= (\tilde{\mu}_{ijk} - \tilde{\mu}_{ij}\tilde{\mu}_{ik})\left[\frac{P_1P_2}{\sqrt{\tilde{\sigma}_{ijj}\sigma^{(c_1)}_{ikk}}} + \frac{P_1P_2}{\sqrt{\sigma^{(c_1)}_{ijj}\tilde{\sigma}_{ikk}}} + \frac{P_2^2}{\sqrt{\tilde{\sigma}_{ijj}\tilde{\sigma}_{ikk}}}\right. \\
&+ \left.\frac{P_2P_3}{\sqrt{\sigma^{(c_2)}_{ijj}\tilde{\sigma}_{ikk}}} + \frac{P_2P_3}{\sqrt{\tilde{\sigma}_{ijj}\sigma^{(c_2)}_{ikk}}}\right] \\
&+ P_1^2\frac{\tilde{\mu}_{ijk} - \tilde{\mu}_{ik}\mu^{(c_1)}_{ij} - \tilde{\mu}_{ij}\mu^{(c_1)}_{ik} + \mu^{(c_1)}_{ij}\mu^{(c_1)}_{ik}}{\sqrt{\sigma^{(c_1)}_{ijj}\sigma^{(c_1)}_{ikk}}} \\
&+ P_1P_3\frac{\tilde{\mu}_{ijk} - \tilde{\mu}_{ik}\mu^{(c_2)}_{ij} - \tilde{\mu}_{ij}\mu^{(c_1)}_{ik} + \mu^{(c_2)}_{ij}\mu^{(c_1)}_{ik}}{\sqrt{\sigma^{(c_2)}_{ijj}\sigma^{(c_1)}_{ikk}}} \\
&+ P_1P_3\frac{\tilde{\mu}_{ijk} - \tilde{\mu}_{ik}\mu^{(c_1)}_{ij} - \tilde{\mu}_{ij}\mu^{(c_2)}_{ik} + \mu^{(c_1)}_{ij}\mu^{(c_2)}_{ik}}{\sqrt{\sigma^{(c_1)}_{ijj}\sigma^{(c_2)}_{ikk}}} \\
&+ P_3^2\frac{\tilde{\mu}_{ijk} - \tilde{\mu}_{ik}\mu^{(c_2)}_{ij} - \tilde{\mu}_{ij}\mu^{(c_2)}_{ik} + \mu^{(c_2)}_{ij}\mu^{(c_2)}_{ik}}{\sqrt{\sigma^{(c_2)}_{ijj}\sigma^{(c_2)}_{ikk}}},
\end{aligned} \tag{A.22}$$

where $a^*_{ij,\ell}$ is given as in (A.16) and $\tilde{\mu}_{ijk} = \frac{1}{M}\sum_{\ell=1}^{M}\tilde{\mu}^*_{ij,\ell}\tilde{\mu}^*_{ik,\ell}$ is given as in (3.8). Now, one can easily compute $var[\psi_c(r_{ij})]$ and $cov[\psi_c(r_{ij}), \psi_c(r_{ik})]$ by using the formulas given in (A.7) and (A.8), respectively.

We remak here that for the binary data with two sided outliers, the formulas for λ_{ij}, $\frac{\partial}{\partial\beta}\psi_c(r_{ij})$, $\frac{\partial}{\partial\beta}\lambda_{ij}$, $E[\psi_c^2(r_{ij})]$, and $E[\psi_c(r_{ij})\psi_c(r_{ik})]$, $(j \neq k)$ are provided in (A.15), (A.17), (A.19), (A.20), and (A.22), respectively. The computations for these quantities under the upper or lower sided outliers case are now immediate from the formulas for the two sided outliers case. For example, to compute the expectation of $\psi_c^2(r_{ij})$ under the upper sided outliers case, we first modify the limits in (3.18) by replacing p_{lb} with 0 (which provides (3.16)). We then , by similar calculations as in

(A.20), obtain

$$E[\psi_c^2(r_{ij})] = P_1 \frac{\tilde{\mu}_{ij}(1 - 2\mu_{ij}^{(c_1)}) + \mu_{ij}^{(c_1)^2}}{\sigma_{ijj}^{(c_1)}} + P_2,$$

under the upper sided outliers case. In the same fashion, the formulas of λ_{ij}, $\frac{\partial}{\partial\beta}\psi_c(r_{ij})$, $\frac{\partial}{\partial\beta}\lambda_{ij}$, $E[\psi_c^2(r_{ij})]$, and $E[\psi_c(r_{ij})\psi_c(r_{ik})]$ for the binary lower sided outlying observations can be computed. To do this, we first change the limits in (3.18) with $p_{sb} = 1$ (which provides (3.17)). Next, based on the changed limits, the formulas follow from (A.15), (A.17), (A.19), (A.20), and (A.22), respectively.

Appendix B

Formulas for $g(\sigma^2)$ and $\frac{\partial}{\partial\sigma^2}g(\sigma^2)$ for the count data

Derivation for $g(\sigma^2)$

Recall from (3.22) that $g(\sigma^2) = \varphi - E(\varphi)$, where φ as a function of $\psi_c(r_{ij})$ is defined in (3.21). Note that the downweighting function $\psi_c(r_{ij})$ for the count data is given in (3.15). Now, to compute $E(\varphi)$, we follow (3.23) and simply use the appropriate formulas for $a^*_{ij,\ell}$ and $b^*_{ij,\ell}$. These formulas for $a^*_{ij,\ell}$ and $b^*_{ij,\ell}$ for the count data are given in (A.3) and (A.10), respectively.

Derivation for $\frac{\partial}{\partial\sigma^2}g(\sigma^2)$

The derivative of $g(\sigma^2)$ with respect to σ^2 can be obtained as

$$\begin{aligned}
\frac{\partial}{\partial\sigma^2}g(\sigma^2) &= -\frac{\partial}{\partial\sigma^2}E(\varphi) \\
&= -\frac{1}{M}\left[\sum_{\ell=1}^{M}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\partial}{\partial\sigma^2}a^*_{ij,\ell} + \sum_{\ell=1}^{M}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\partial}{\partial\sigma^2}b^*_{ij,\ell}\right. \\
&+ \left.\sum_{\ell=1}^{M}\sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\left\{a^*_{ij,\ell}\frac{\partial}{\partial\sigma^2}a^*_{ik,\ell} + a^*_{ik,\ell}\frac{\partial}{\partial\sigma^2}a^*_{ij,\ell}\right\}\right], \qquad \text{(B.1)}
\end{aligned}$$

where

$$\frac{\partial}{\partial \sigma^2} a^*_{ij,\ell} = \frac{1}{2\sqrt{\tilde{\sigma}_{ijj}}} \tilde{\mu}^*_{ij,\ell} \left(\frac{\gamma_{i\ell}}{\sigma} - 1 \right) \left[F_{ij|\gamma_{i\ell}}(I^U_{ij} - 1) - F_{ij|\gamma_{i\ell}}(I^L_{ij} - 1) \right], \tag{B.2}$$

and

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} b^*_{ij,\ell} &= -\frac{\tilde{\mu}_{ij}}{\tilde{\sigma}_{ijj}} \tilde{\mu}^*_{ij,\ell} \left(\frac{\gamma_{i\ell}}{\sigma} - 1 \right) \left[F_{ij|\gamma_{i\ell}}(I^U_{ij} - 1) - F_{ij|\gamma_{i\ell}}(I^L_{ij} - 1) \right] \\ &+ \frac{1}{2\tilde{\sigma}_{ijj}} \tilde{\mu}^*_{ij,\ell} \left(\frac{\gamma_{i\ell}}{\sigma} - 1 \right) \left[F_{ij|\gamma_{i\ell}}(I^U_{ij} - 1) - F_{ij|\gamma_{i\ell}}(I^L_{ij} - 1) \right] \\ &+ \frac{1}{\tilde{\sigma}_{ijj}} \tilde{\mu}^*_{ij,\ell} \tilde{\mu}^*_{ij,\ell} \left(\frac{\gamma_{i\ell}}{\sigma} - 1 \right) \left[F_{ij|\gamma_{i\ell}}(I^U_{ij} - 2) - F_{ij|\gamma_{i\ell}}(I^L_{ij} - 2) \right], \end{aligned} \tag{B.3}$$

with $\tilde{\mu}^*_{ij,\ell} = exp(\tilde{x}'_{ij}\beta - \frac{\sigma^2}{2} + \sigma\gamma_{i,\ell})$ and $F_{ij|\gamma_{i\ell}}(\cdot)$ as the cumulative distribution function conditional on $\gamma_{i\ell}$, where for $\ell = 1, \ldots, M$, $\gamma_{i\ell}$ is the ℓth realized quantity for γ_i, generated following $\gamma_i \overset{i.i.d.}{\sim} N(0, 1)$.

Formulas for $g(\sigma^2)$ and $\frac{\partial}{\partial \sigma^2} g(\sigma^2)$ for the binary data

Derivation for $g(\sigma^2)$

For convenience, we first explain how to compute $g(\sigma^2) = \varphi - E(\varphi)$ for the binary data with two sided outliers. For the purpose, we follow (3.21) and simply use the formula for $\psi_c(r_{ij})$, $i = 1, \ldots, K$, $j = 1, \ldots, n_i$, given in (3.18). Similar to the count data case, $E(\varphi)$ now may be computed by using (A.16), and (A.21) in (3.23).

Next, for the upper and lower sided outliers cases, we first define φ by modifying the limits in (3.18) appropriately. For example, for the upper sided outliers case, we replace p_{lb} with 0 in (3.18). Now, to compute $E(\varphi)$, the formulas for $a^*_{ij,\ell}$ and $b^*_{ij,\ell}$ for the upper sided outliers case may be computed from (A.16), and (A.21), respectively, by reflecting the new limits imposed in (3.18).

The formulas for φ and $E(\varphi)$ for the lower sided outliers case can be derived by following the upper sided outliers case, but by considering $p_{sb} = 1$ in the limits in (3.18).

Derivation for $\frac{\partial}{\partial\sigma^2}g(\sigma^2)$

Since the formulas for $\frac{\partial}{\partial\sigma^2}g(\sigma^2)$ in the one sided outliers case can be derived from those obtained for the two sided outliers case as a special case, here we only show how to compute $\frac{\partial}{\partial\sigma^2}g(\sigma^2)$ in the two sided outlier case. For the purpose, we consider the downweighting function $\psi_c(r_{ij})$ as given in (3.18). Now, the derivative of $g(\sigma^2)$ with respect to σ^2 can be obtained as

$$\frac{\partial}{\partial\sigma^2}g(\sigma^2) = \frac{\partial}{\partial\sigma^2}\varphi - \frac{\partial}{\partial\sigma^2}E(\varphi). \tag{B.4}$$

In (B.4),

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}\varphi &= \sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\partial}{\partial\sigma^2}\psi_c(r_{ij}) + 2\sum_{i=1}^{K}\sum_{j=1}^{n_i}\psi_c(r_{ij})\frac{\partial}{\partial\sigma^2}\psi_c(r_{ij}) \\ &+ \sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\left[\psi_c(r_{ij})\frac{\partial}{\partial\sigma^2}\psi_c(r_{ik}) + \psi_c(r_{ik})\frac{\partial}{\partial\sigma^2}\psi_c(r_{ij})\right],\end{aligned} \tag{B.5}$$

where

$$\frac{\partial}{\partial\sigma^2}\psi_c(r_{ij}) = \begin{cases} 0, & P(Y_{ij}=1|\tilde{x}_{ij}) > p_{sb}, (i,j) \equiv (i',j'), \\ -\frac{1}{\sqrt{\tilde{\sigma}_{ijj}}}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij}, & p_{lb} \leq P(Y_{ij}=1|\tilde{x}_{ij}) \leq p_{sb}, (i,j) \not\equiv (i',j'), \\ 0, & P(Y_{ij}=1|\tilde{x}_{ij}) < p_{lb}, (i,j) \equiv (i',j'), \end{cases}$$

with

$$\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij} = \frac{1}{M}\sum_{\ell=1}^{M}\frac{\partial}{\partial\sigma^2}\tilde{\mu}^*_{ij,\ell}$$

$$= \frac{1}{2M\sigma}\sum_{\ell=1}^{M}\tilde{\mu}^*_{ij,\ell}(1-\tilde{\mu}^*_{ij,\ell})\gamma_{i\ell}. \tag{B.6}$$

Also in (B.4),

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}E(\varphi) &= \sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\partial}{\partial\sigma^2}E[\psi_c(r_{ij})]+\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\partial}{\partial\sigma^2}E[\psi_c^2(r_{ij})]\\ &+ \sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i}\frac{\partial}{\partial\sigma^2}E[\psi_c(r_{ij})\psi_c(r_{ik})], \end{aligned}\tag{B.7}$$

where $E[\psi_c(r_{ij})]$, $E[\psi_c^2(r_{ij})]$, and $E[\psi_c(r_{ij})\psi_c(r_{ik})]$ are given as in (A.15), (A.20), and (A.22), respectively, so that

$$\frac{\partial}{\partial\sigma^2}E[\psi_c(r_{ij})] = \frac{P_1}{\sqrt{\sigma^{(c_1)}_{ijj}}}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij}+\frac{P_3}{\sqrt{\sigma^{(c_2)}_{ijj}}}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij},$$

$$\frac{\partial}{\partial\sigma^2}E[\psi_c^2(r_{ij})] = \frac{P_1(1-2\mu^{(c_1)}_{ij})}{\sigma^{(c_1)}_{ijj}}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij}+\frac{P_3(1-2\mu^{(c_2)}_{ij})}{\sigma^{(c_2)}_{ijj}}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij},$$

and

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}E[\psi_c(r_{ij})\psi_c(r_{ik})] &= \left[\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ijk}-\left\{\tilde{\mu}_{ij}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ik}+\tilde{\mu}_{ik}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij}\right\}\right]\\ &\times\left[\frac{P_1P_2}{\sqrt{\tilde{\sigma}_{ijj}\sigma^{(c_1)}_{ikk}}}+\frac{P_1P_2}{\sqrt{\sigma^{(c_1)}_{ijj}\tilde{\sigma}_{ikk}}}+\frac{P_2^2}{\sqrt{\tilde{\sigma}_{ijj}\tilde{\sigma}_{ikk}}}+\frac{P_2P_3}{\sqrt{\sigma^{(c_2)}_{ijj}\tilde{\sigma}_{ikk}}}+\frac{P_2P_3}{\sqrt{\tilde{\sigma}_{ijj}\sigma^{(c_2)}_{ikk}}}\right]\\ &+P_1^2\frac{\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ijk}-\mu^{(c_1)}_{ij}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ik}-\mu^{(c_1)}_{ik}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij}}{\sqrt{\sigma^{(c_1)}_{ijj}\sigma^{(c_1)}_{ikk}}}+P_1P_3\frac{\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ijk}-\mu^{(c_2)}_{ij}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ik}-\mu^{(c_1)}_{ik}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij}}{\sqrt{\sigma^{(c_2)}_{ijj}\sigma^{(c_1)}_{ikk}}}\\ &+P_1P_3\frac{\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ijk}-\mu^{(c_1)}_{ij}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ik}-\mu^{(c_2)}_{ik}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij}}{\sqrt{\sigma^{(c_1)}_{ijj}\sigma^{(c_2)}_{ikk}}}+P_3^2\frac{\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ijk}-\mu^{(c_2)}_{ij}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ik}-\mu^{(c_2)}_{ik}\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij}}{\sqrt{\sigma^{(c_2)}_{ijj}\sigma^{(c_2)}_{ikk}}},\end{aligned}$$

with

$$\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ijk} = \frac{1}{2M\sigma}\sum_{\ell=1}^{M}\gamma_{i\ell}\tilde{\mu}^*_{ij,\ell}\tilde{\mu}^*_{ik,\ell}(2-\tilde{\mu}^*_{ij,\ell}-\tilde{\mu}^*_{ik,\ell}),$$

and $\frac{\partial}{\partial\sigma^2}\tilde{\mu}_{ij}$ as in (B.6).

Bibliography

[1] Amemiya, T. (1985) *Advanced Econometrics.* Harvard University Press, Cambridge, Massachusetts.

[2] Bockenholt, U. (1999) Analyzing multiple emotions over time by autoregressive negative multinomial regression models. *Journal of the American Statistical Association*, **94**, 757-765.

[3] Breslow, N. E., and Clayton, D. G. (1993) Approximate Inference in Generalized Linear Mixed Models. *Journal of American Statistical Association,* **88**, 9-25.

[4] Breslow, N. E., and Lin, X. (1995) Bias correction in generalized linear models with a single component of dispersion. *Biometrika*, **82**, 81-91.

[5] Cantoni, E. (2004) A robust approach to longitudinal data analysis. *The Canadian Journal of Statistics*, **32**, 169-180.

[6] Cantoni, E. and Ronchetti, E. (2001) Robust inference for generalized linear models. *Journal of the American Statistical Association*, **96**, 1022-1030.

[7] Carroll, R. J. and Pederson, S. (1993) On robustness in the logistic regression model. *Journal of the Royal Statistical Society, B*, **55**, 693-706.

[8] Copas, J. B. (1988) Binary regression models for contaminated data (with discussion). *Journal of the Royal Statistical Society, B*, **50**, 225-265.

[9] Crowder, M. (1995) On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika*, **82**, 407-410.

[10] Davis, R. A., Dunsmuir, W. T. M., and Wang, Y. (2000) On autocorrelation in a Poisson regression model. *Biometrika*, **87**, 491-506.

[11] Fahrmeir, L. and Tutz, G. T. (1994) *Multivariate statistical modelling based on generalized linear models.* New York, Springer-Verlag.

[12] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986) *Robust Statistics: The Approach Based on Influence Functions.* John Wiley and Sons, New York.

[13] He, X., Fung, W. K., and Zhu, Z. (2005) Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association,* **100**, 1176-1184.

[14] Heyde, C. C. (1997) *Quasi-Likelihood and its Application.* New York: Springer-Verlag.

[15] Huber, P. J. (2004) *Robust Statistics.* John Wiley and Sons, New York.

[16] Jiang, J. (1998) Consistent Estimators in Generalized Linear Mixed Models. *Journal of the American Statistical Association,* **93**, 720-729.

[17] Jiang, J. and Zhang, W. (2001) Robust estimation in generalized linear mixed models. *Biometrika,* **88**, 753-765.

[18] Jowaheer,V. and Sutradhar, B. C. (2002) Analyzing longitudinal count data with over dispersion. *Biometrika,* **89**, 389-399.

[19] Kuk, A. Y. C. (1995) Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society, B,* **57**, 395-407.

[20] Künsch, H. R., Stefanski, L. A., and Carroll, R. J. (1989) Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models. *Journal of the American Statistical Association,* **84**, 460-466.

[21] Lee, Y. and Nelder, J. A. (1996) Hierarchical generalized linear models. *Journal of the Royal Statistical Society, B,* **58**, 619-678.

[22] Liang, K. Y. and Zeger, S. L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika,* **73**, 13-22.

[23] Lin, X. and Breslow, N. E. (1996) Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association,* **91**, 1007-1016.

[24] Mallows, C. L. (1975) On some topics in robustness. *Technical Memorandum,* Bell Telephone Laboratories, Murray Hill.

[25] McCullagh, P. (1983) Quasilikelihood functions. *The Annals of Statistics*, **11**, 59-67.

[26] McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

[27] McCulloch, C. E. (1994) Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association*, **89**, 330-335.

[28] McCulloch, C. E. (1997) Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, **92**, 162-170.

[29] McKenzie, E. (1988) Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, **20**, 822-835.

[30] Mills, J. E., Field, C. A., and Dupuis, D. J. (2002) Marginally specified generalized linear mixed models: A robust approach. *Biometrics*, **58**, 727-734.

[31] Morgenthaler, S. (1992) Least-absolutely-deviations fits for generalized linear models. *Biometrika*, **79**, 747-754.

[32] Pregibon, D. (1982) Resistant fits for some commonly used logistic models with medical applications. *Biometrics*, **38**, 485-498.

[33] Preisser, J. S. and Qaqish, B. F. (1999) Robust regression for clustered data with applications to binary regression. *Biometrics*, **55**, 574-579.

[34] Qaqish, B. F. (2003) A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika*, **90**, 455-463.

[35] Rousseeuw, P. J. and Leroy, A. M. (1987) *Robust Regression and Outlier Detection.* John Wiley and Sons, New York.

[36] Ruckstuhl, A. F. and Welsh, A. H. (1999) Robust fitting of the binomial models. CMA, Australian National University, manuscript.

[37] Schall, R. (1991) Estimation in generalized linear models with random effects. *Biometrika*, **78**, 719-727.

[38] Sinha, S. K. (2004) Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association*, **99**, 451-460.

[39] Sinha, S. K. (2006) Robust inference in generalized linear model for longitudinal data. *The Canadian Journal of Statistics*, **34**, 1-18.

[40] Stefanski, L. A., Carroll, R. J., and Ruppert, D. (1986) Optimally bounded score functions for generalized linear models with applications to logistic regression. *Biometrika*, **73**, 413-424.

[41] Sutradhar, B. C. (2004) On Exact Quasilikelihood Inference in Generalized Linear Mixed Models. *Sankhya : The Indian Journal of Statistics,* **66**, 261-289.

[42] Sutradhar, B. C. (2003) A Review on Regression Models for Discrete Longitudinal Responses. *Statistical Science,* **18**, 377-393.

[43] Sutradhar, B. C. and Das, K. (1999) On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika*, **86**, 459-65.

[44] Sutradhar, B.C. and Mukerjee, R. (2005) On Likelihood Inference in Binary Mixed Model with an Application to COPD Data. *Computational Statistics and Data Analysis,* **48**, 345-361.

[45] Sutradhar, B.C. and Qu, Z. (1998) On approximate likelihood inference in a Poisson mixed model. *The Canadian Journal of Statistics*, **26**, 169-186.

[46] Sutradhar, B.C. and Rao, R. P. (2001) On marginal quasi-likelihood inference in generalized linear mixed models. *Journal of Multivariate Analysis*, **76**, 1-34.

[47] Sutradhar, B.C. and Rao, R. P. (2003) On Quasilikelihood Inference in Generalized Linear Mixed Models With Two Components of Dispersion. *The Canadian Journal of Statistics,* **31**, 415-435.

[48] Thall, P. F. and Vail, S. C. (1990) Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657-671.

[49] Wedderburn, R.W.M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439-447.

[50] Zeger, S. L. and Karim, M. R. (1991) Generalized linear models with random effects: A gibbs sampling approach. *Journal of the American Statistical Association*, **86**, 79-86.

[51] Zeger, S.L., Liang, K. Y., and Self, S. G. (1985) The analysis of binary longitudinal data with time independent covariates. *Biometrika*, 72, 31-38.