The generalized linear model and extensions: a review and some biological and environmental applications

Sudhir Paul[1] and Krishna K. Saha[2*,†]

[1]*Department of Mathematics and Statistics, University of Windsor, 401 Sunset, Windsor, Ont. N9B 3P4, Canada*
[2]*Department of Mathematical Sciences, Central Connecticut State University,*
1615 Stanley Street, New Britain, CT 06050, USA

SUMMARY

The generalized linear model (GLM) encompasses many discrete and continuous models and it is particularly useful for analyzing discrete data. However, in many real life applications, the full distributional assumption of the GLM cannot be justified. Further, the GLM cannot accommodate over-dispersion in the data. Since the inception of the GLM by Nelder and Wedderburn (1972) a number of its extensions have been proposed in the literature for robust analysis of discrete data. The purpose of this paper is to critically review these extensions. Applications to over-dispersed Poisson and binomial models are shown. Some simulations are conducted to compare, in terms of bias and efficiency, the estimates of mean and the dispersion parameters by different methods. Applications to some biological and environmental data are given. Copyright © 2007 John Wiley & Sons, Ltd.

1. INTRODUCTION

Nelder and Wedderburn (1972) extend the general linear models to generalized linear models (GLMs) by including the exponential family of error distributions along with the normal. The GLM requires full distributional assumption. However, in some situations, specially in discrete data problems, sometimes a full distributional assumption becomes restrictive. To avoid this full distributional assumption Wedderburn (1974) introduces the quasi-likelihood (QL) based on the assumption of only the first two moments of the response variable. The QL methodology is useful for estimating only mean or the regression parameters. By introducing a normalizing factor to the QL Nelder and Pregibon (1987) and Goodambe and Thompson (1989) proposed the extended quasi-likelihood (EQL). The EQL resembles a likelihood involving not only the mean (regression) parameters but also the variance parameter of the response variable. In situations where a full distributional assumption for the response variable is not available, the EQL methodology can be used to jointly estimate the mean and the variance parameters. The EQL methodology is then useful for the joint estimation of the mean and the dispersion

---
*Correspondence to: K. K. Saha, Department of Mathematical Sciences, Central Connecticut State University, 1615 Stanley Street, New Britain, CT 06050, USA.
†E-mail: saha.krishna@gmail.com

parameters in over-dispersed discrete models, such as the over-dispersed binomial and over-dispersed Poisson models. For the joint estimation of the mean and the dispersion parameters in over-dispersed models or in variance components models Lee and Nelder (1996) developed hierarchical-likelihood (h-likelihood) methodology. In a h-likelihood methodology we first form a (log) likelihood, say, l_0 by assuming that the random variable Y given some random effect u follows a GLM. Further, a conjugate GLM is produced from this likelihood for the random effect u given some mean parameter and some dispersion parameter from which a further (log) likelihood, say, l_1 is obtained. Then $l = l_0 + l_1$ is the hierarchical (log) likelihood. Inference for the mean and the dispersion parameter is then made based on the hierarchical (log) likelihood l. The h-likelihood (HL), again, is based on full distributional assumptions. In order to avoid full distributional assumption at both stages of the hierarchy, Lee and Nelder (2001) introduce double extended quasi-likelihood (DEQL) for the joint estimation of the mean and the dispersion parameters. The DEQL methodology requires an EQL for Y_i given some random effect u and an EQL for the random effect parameter u. The DEQL then is obtained by combining these two EQLs.

For the analysis of discrete data with over-dispersion the generalized linear mixed models (GLMMs) (See, for example, Schall, 1991 and Breslow and Clayton, 1993) are also useful. However, in this paper we are concerned with the robust procedures requiring minimum assumptions, such as the assumption of only the first two moments of the response variable. Our main purpose then is to review the QL, EQL, and DEQL along with the GLM and the HL procedures and to show applications to some discrete data problems.

A brief review of GLM, QL, EQL, HL, and DEQL methodologies are given in Section 2 and applications to Poisson and binomial data are shown in Section 3. Some simulations pertaining to the estimation of the over-dispersion parameters in binomial and Poisson data are conducted in Section 4. Applications to some biological and environmental data are given in Section 5.

2. THE GENERALIZED LINEAR MODEL AND EXTENSIONS

2.1. A generalized linear model (GLM)

The random variable Y has a distribution of the GLM form if

$$f(y; \theta) = \exp[\{y\theta(\mu) - b(\theta(\mu))\}/\lambda + c(y, \lambda)]$$

where $E(Y) = \mu$, $\text{var}(y) = \lambda V(\mu)$, and $\theta(\mu)$ is GLM canonical parameter. In GLMs, $V(\mu)$ is a variance function which characterizes a particular GLM family of distributions. The discrete models, namely, the binomial model and the Poisson model belong to this family. A set of covariates $x_1, x_2, \ldots, x_k$ is related to the mean μ by $\theta(\mu) = X\beta$. The log-likelihood for n observations $y_1, \ldots, y_n$ then is

$$l(y, \mu) = \sum_{i=1}^{n} \left[\{y_i\theta(\mu) - b(\theta(\mu))\}/\lambda + c(y_i, \lambda) \right] \tag{2.1}$$

Usually, λ is a known constant and inference procedures regarding the mean μ or the regression parameters $\beta_1, \ldots, \beta_k$ are made using $l(y, \mu)$.

Table 1. Number of runs discarded for 10 000 acceptable samples for over-dispersed Poisson case, $m = 1, 3, 5$ and $c = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

			c								
Data distribution	n	m	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
NB	10	1	8953	7641	6652	5917	5418	5025	4693	4435	4216
		3	1714	1074	699	509	369	268	203	173	155
		5	1634	1072	679	467	327	219	285	123	138
	50	1	1120	771	550	372	284	214	169	117	82
		3	2776	871	244	81	34	3	3	2	2
		5	2419	763	289	67	24	11	6	2	1

2.2. The quasi-likelihood

In many real life data analysis the full distributional assumption of the GLM cannot always be justified. To avoid the full distributional assumption, Wedderburn (1974) proposed the QL which is based on the knowledge of the form of first two moments of the random variable Y. The quasi-log-likelihood with the above mean and variance is given by

$$Q(y, \mu) = \sum_{i=1}^{n} \left[\int_{y_i}^{\mu} \frac{(y_i - t)}{\lambda V(t)} dt \right] \tag{2.2}$$

With some specification of the variance function $Q(y, \mu)$ replaces $l(y, \mu)$ for the estimation of the regression parameters. Many authors refer the quasi-log-likelihood simply as QL. Therefore, henceforth, we will refer this as QL.

2.3. The extended quasi-likelihood

In certain situations data show evidence of over-dispersion. For example, some biological and environmental data in the form of counts or proportions show over-dispersion. Table 3 contains count data for four treatments. The means of the data in the four treatment groups are 4.033, 3.167, 1.483, and 1.508. The corresponding variances are 16.056, 7.7703, 3.1929, and 3.6302. There is strong evidence of over-dispersion in all the four groups. Table 6 contains data in the form of proportions for six dose groups. The variances of the proportions (y/n, y: number of combined endpoints, n: number of implants) for the dose groups 0, 30, 45, 60, 75, and 90 are 0.005718, 0.023296, 0.080641, 0.152547, 0.109430, and 0.048894, respectively. The corresponding estimated variances using a binomial variance formula $\hat{p}(1 - \hat{p})/N$ are 0.006204, 0.010353, 0.018283, 0.023549, 0.016079, and 0.004430. Note, here we have used $\hat{p} = \sum y / \sum n$ and $N = \sum n / \#$ of litters. Clearly, in some groups, there is evidence of over-dispersion in the data. Under-dispersion is seen in some groups.

Parametric models, such as the negative binomial model for count data and the beta-binomial model for the data in the form of proportions, are available for taking the over-dispersion into account. However, a full distributional assumption may not be justified in real life situations. One can incorporate the over-dispersion in the variance function of the QL in which case it takes the same form as Equation (2.2), where, now, $V(\mu)$ is a function of the mean parameter μ and some over-dispersion parameter θ. Since this is not a likelihood it can not be used to estimate the over-dispersion parameter. In this situation, for

the joint estimation of the mean and dispersion parameters, Nelder and Pregibon (1987) and Godambe and Thompson (1989) suggest to use an EQL. The EQL with some known mean $E(Y) = \mu$ and variance $\text{var}(y) = \lambda V(\mu)$ is

$$Q^+(y, \mu, \lambda, \theta) = \sum_{i=1}^{n} \left[-\frac{1}{2} \ln[2\pi\lambda V(y)] + Q(y, \mu, \theta) \right] \qquad (2.3)$$

Note that the second term on the right hand side of the above equation is the QL for μ and θ and the first term is the normalizing factor; thus making $\exp(Q^+)$ resemble a likelihood. The EQL Q^+ then can be used for the joint estimation of the mean parameter (or the regression parameters) and the over-dispersion parameter θ. The advantage of using Q^+ for the estimation of the over-dispersion parameter θ is that it can be robust compared to the ML (maximum likelihood) estimates as the full distributional assumptions are not required, yet Q^+ behaves like a log-likelihood.

2.4. Hierarchical generalized linear models

In a full distributional model framework the GLM can be used only for the estimation of the mean (regression) parameters and not for the dispersion parameter. Lee and Nelder (1996) extend the GLM for the joint estimation of the mean and the dispersion parameters or the variance component parameters using full distributional assumptions. This extended model is the hierarchical generalized linear model (HGLM).

Let y be the response and u be some unobserved random variable with $E(y|u) = \mu_0$ and $\text{var}(y|u) = \lambda V_0(\mu_0)$, where λ is some dispersion parameter and $V_0(\mu_0)$ is some known variance function. Then the HGLM is defined according to the following two stages.

(i) the response variable y, given the unobserved random effect u, has the GLM, that is,

$Y_i|u_i \sim$ GLM with $E(Y_i|u_i) = \mu_{0i}$ and $\text{var}(Y_i|u_i) = \lambda V_0(\mu_{0i})$. The log-likelihood for stage 1 of the hierarchy can be written as

$$l_o(\theta(\mu_0), \lambda; y|u) = \sum_{i=1}^{n} \left[\{y_i\theta(\mu_0) - b(\theta(\mu_0))\}/\lambda + k(y_i, \lambda) \right]$$

where $\theta(\mu_0)$ is the canonical parameter. The linear predictor η_0 takes the form

$$\eta_0 = g(\mu_0) = X\beta + v$$

where $g(\cdot)$ is the link function and $v = g_1(u)$ is some strictly monotonic function of the random effect u. The scale change $v = g_1(u)$ is made such that the random effect u occurs linearly in the linear predictor. For example, for Poisson data $E(y|u) = \mu_0$, $\eta_0 = \log(\mu_0) = X\beta + v$, where $v = \log(u)$.

(ii) The random effect u, given some dispersion parameter φ, has the GLM, that is, $u_i \sim$ GLM with $E(u_i) = \omega$ and $\text{var}(u_i) = \varphi V_1(\omega)$. Then the kernel of the log-likelihood of this model can be written as

$$l_1(v; u) = \sum_i \left[\omega\theta(u) - b(\theta(u)) \right]/v$$

where ν is a function of φ. Thus, this can be viewed as the likelihood $l_1(u; \theta(\omega))$ of quasi-data ω with quasi-fixed parameters u and hence $\theta(u)$ has the purely formal relationship

$$E(\omega) = u = b'(\theta(u)) \quad \text{and} \quad \text{var}(\omega) = \nu V_1(u)$$

where $V_1(u) = b''(\theta(u))$. For details of application see Section 3.

For inferences for the mean (regression) and the dispersion parameters in the HGLMs Lee and Nelder (1996) propose to use the hierarchical (log)-likelihood (h-likelihood). The h-likelihood is the logarithm of the joint density function of y and u which is defined by

$$h = l_0(\theta(\mu_0), \lambda; y|u) + l_1(\nu; u)$$

where $l_1(\nu; u)$ is the log-likelihood for u with parameter ν. Using the scale change $\nu = g_1(u)$, suggested above, of the random effect u the h-likelihood becomes

$$h = l_0(\theta(\mu_0), \lambda; y|\nu) + l_1(\nu; \nu)$$

where $l_0(\theta(\mu_0), \lambda; y|\nu) = l_0(\theta(\mu_0), \lambda; y|g_1(u))$ is the log-likelihood for $y|\nu$ and $l_1(\nu; \nu) = l_1(\nu; u) + \sum \ln(|\partial u_i / \partial v_i|)$ is the log-likelihood for ν, with $|\partial u_i / \partial v_i|$ being the Jacobian.

2.5. Double extended quasi-likelihood

As the GLM, the HGLM also suffers from the drawback that full distributional assumptions are required at both stages of the hierarchy. In order to avoid to have to assume full distributional assumptions Lee and Nelder (2001) propose the use of the DEQL for the joint estimation of the mean and the dispersion parameters. The DEQL methodology replaces $l_0(\theta(\mu_0), \lambda; y|u)$ by an EQL for $Y|u$ and $l_1(\nu; u)$, by an EQL for u. The DEQL then is obtained by combining these two EQLs. For details of application see Section 3.

3. APPLICATIONS TO OVER-DISPERSED POISSON AND BINOMIAL MODELS

In what follows we show applications of the methodologies discussed in Section 2 to Poisson and binomial data.

3.1. Over-dispersed Poisson data

For the Poisson distributed data, $\theta(m) = \ln m$, $b(\theta(m)) = m$, $\lambda = 1$, $V(m) = m$ and $c(y, \lambda) = -\log(y!)$. Then, for a sample $y_1, \ldots, y_n$, apart from a constant, we have $l(y, m) = Q(y, m) = \sum_{i=1}^{n}[y_i \ln m - m]$. Thus, for Poisson data $l(y, m)$ and $Q(y, m)$ are identical.

Neither $l(y, m)$ nor $Q(y, m)$ is suitable for the estimation of over-dispersion. We can use a specific over-dispersion model for this. For example a convenient over-dispersion model for count Y is the

negative binomial (m, c) distribution

$$f(y|m, c) = Pr(Y = y|m, c) = \frac{\Gamma(y + c^{-1})}{y!\Gamma(c^{-1})} \left(\frac{cm}{1 + cm}\right)^y \left(\frac{1}{1 + cm}\right)^{c^{-1}} \tag{3.1}$$

for $m > 0$. Now, $\text{var}(Y) = m(1 + mc)$ and $c > -1/m$. Since c can take a positive as well as a negative value it is called a dispersion parameter, rather than an over-dispersion parameter, and with this range of c, $f(y|m, c)$ is a valid probability function. Obviously, when $c \to 0$, the $NB(m, c)$ becomes Poisson(m).

The log-likelihood, apart from a constant, for n observations $y_1, \ldots, y_n$, is

$$l(y, m, c) = \sum_{i=1}^{n} \left[y_i \ln(m) - \left(y_i + \frac{1}{c}\right) \ln(1 + cm) + \sum_{j=0}^{y_i - 1} \ln(1 + cj) \right] \tag{3.2}$$

Inference for both m and c can then be made using $l(y, m, c)$. However, as mentioned in Subsection 2.3, where full distributional assumptions cannot be made for the response variable, an EQL might be used for the estimation of the regression parameters and the over-dispersion parameter simultaneously. As indicated in Section 2.3, an EQL requires the specification of the variance function $V(Y)$. Now, for over-dispersed Poisson data $\lambda = 1$ and $V(Y) = y(1 + yc)$. With this, the QL for n observations $y_1, \ldots, y_n$ can be written as

$$Q(y, m, c) = \sum_{i=1}^{n} \left[\left(y_i + \frac{1}{c}\right) \ln \left(\frac{1 + cy_i}{1 + cm}\right) - y_i \ln \frac{y_i}{m} \right]$$

and the EQL is

$$Q^+(y, m, c) = \sum_{i=1}^{n} \left[-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln[y_i(1 + cy_i)] + (y_i + \frac{1}{c}) \ln \left(\frac{1 + cy_i}{1 + cm}\right) - y_i \ln \frac{y_i}{m} \right] \tag{3.3}$$

The EQL Q^+ with the variance function $V(y) = y(1 + cy)$ is an approximation of the log-likelihood of the negative binomial model when the factorials are replaced by their usual Stirling's approximations

$$z! \simeq (2\pi z)^{1/2} z^z \mathrm{e}^{-z}$$

As we can see, for $y = 0$, Q^+ becomes infinity. To avoid this, Nelder and Pregibon (1987) suggest to use the modified variance function $V(y) = (y + 1/6)(1 + cy)^2(1 + c/6)/(1 + cy + c/6)$ which is obtained by using the modified Stirling's approximation

$$z! \simeq \left\{ 2\pi \left(z + \frac{1}{6}\right) \right\}^{1/2} z^z \mathrm{e}^{-z}$$

Using this variance function the modified EQL becomes

$$Q^{+*} = \sum_{i=1}^{n} \left[-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \left\{ \frac{(y_i + 1/6)(1 + cy_i)^2 (1 + c/6)}{1 + cy_i + c/6} \right\} \right.$$
$$\left. + \left(y_i + \frac{1}{c} \right) \ln \left(\frac{1 + cy_i}{1 + cm} \right) - y_i \ln \frac{y_i}{m} \right]$$

Then, inference for the parameters m and c can be made using Q^{+*}.

We now obtain the hierarchical log-likelihood for the over-dispersed Poisson data. For the Poisson-gamma GLM we first assume that $Y|u \sim \text{Poisson}(u)$. This can equivalently be written as a GLM

$$f(y|u, \mu_0, \lambda) = \exp[\{y\theta(\mu_0) - b(\theta(\mu_0))\}/\lambda + k(y, \lambda)] \tag{3.4}$$

with $\mu_0 = E(Y_i|u)$, $\theta(\mu_0) = \log(\mu_0)$, $b(\theta(\mu_0)) = \exp(\theta(\mu_0))$, $\lambda = 1$ and $k(y, \lambda) = -\ln(y!)$. The log-likelihood of model (3.4) for $Y_i|u_i$, $i = 1, \ldots, n$ is given by

$$l_0[\mu_0, \lambda; y|u] = \sum_{i=1}^{n} [\{y_i\theta(\mu_{0i}) - b(\theta(\mu_{0i}))\} - \ln(y!)] \tag{3.5}$$

We assume further that the random effect parameter u_i is distributed as a gamma distribution with $E(u_i) = v\alpha = m$ and $\text{var}(u_i) = v\alpha^2 = \alpha m$ with $\alpha = cm$. This model can again be written as a GLM of which the log-likelihood is

$$l_1[\alpha; u] = \sum_{i=1}^{n} [\{m\theta(u_i) - b(\theta(u_i))\}/\alpha + g_1(u_i, m, \alpha)] \tag{3.6}$$

where $\theta(u_i) = \ln(u_i)$, $b(\theta(u_i)) = \exp(\theta(u_i))$, and $g_1(u_i, m, \alpha) = -\ln(u_i) - \ln \Gamma(m/\alpha) - \frac{m}{\alpha} \ln(\alpha)$. For this Poisson-gamma hierarchical model the h-likelihood is given by

$$h(m, \alpha, v_i) = l_0[\mu_0, \lambda; y|u] + l_1[\alpha; u] \tag{3.7}$$

With the change of scale $v(u) = \theta(u) = \ln(u)$ the h-likelihood can be written as

$$h(m, \alpha, v_i) = l_0[\mu_0, \lambda; y|u] + l_1[\alpha; v]$$
$$= \sum_{i=1}^{n} \left[\left(y_i + \frac{m}{\alpha} \right) v_i - \left(1 + \frac{1}{\alpha} \right) \exp(v_i) + g_2(m, \alpha) + k(y_i, \lambda) \right] \tag{3.8}$$

where $l_1[\alpha; v] = l_1[\alpha; u] + \sum_{i=1}^{n} \ln\{|\partial u_i/\partial v_i|\}$, $|\partial u_i/\partial v_i| = u_i$ is the Jacobian, and $g_2(m, \alpha) = -\ln \Gamma(m/\alpha) - \frac{m}{\alpha} \ln(\alpha)$. To estimate m and α we must integrate the v_i's out from $h(m, \alpha, v_i)$. In the Poisson case the resultant integrant, however, is identical with the negative binomial log-likelihood,

that is, (see Lee and Nelder, 1996; p. 622)

$$l(y, m, \alpha) = \ln\left[\int \exp(h)\mathrm{d}v\right]$$

$$= \ln\left[\int \exp\left\{\sum_{i=1}^{n}\left[\left(y_i + \frac{m}{\alpha}\right)v_i - \left(1 + \frac{1}{\alpha}\right)\exp(v_i) + k(y_i, \lambda) + g_2(m, \alpha)\right]\right\}\mathrm{d}v\right]$$

As we know $v_i = \ln(u_i)$, where $u_i \sim \mathrm{Gamma}(\frac{m}{\alpha}, \alpha)$ so v_i has the log-gamma distribution. However, for simplicity, we integrate out the above expressions in terms of u_i instead of v_i, which, after simplification, is

$$l(y, m, \alpha) = \ln\left[\prod_{i=1}^{n}\frac{1}{y_i!\Gamma(m/\alpha)\alpha^{m/\alpha}}\int_{0}^{\infty}u_i^{y_i + \frac{m}{\alpha} - 1}\exp\left[-\left(1 + \frac{1}{\alpha}\right)u_i\right]\mathrm{d}u_i\right]$$

and the resultant integrant then becomes

$$l(y, m, \alpha) = \sum_{i=1}^{n}\left[\ln\left\{\frac{\Gamma\left(y_i + \frac{m}{\alpha}\right)}{\Gamma\left(\frac{m}{\alpha}\right)}\right\} - \frac{m}{\alpha}\ln(\alpha) - \left(y_i + \frac{m}{\alpha}\right)\ln\left(1 + \frac{1}{\alpha}\right) - \ln(y_i!)\right]$$

According to Lawless (1987), the 1st term of the above expression can be written as

$$\ln\left\{\frac{\Gamma\left(y_i + \frac{m}{\alpha}\right)}{\Gamma\left(\frac{m}{\alpha}\right)}\right\} = \sum_{j=1}^{y_i - 1}\ln\left(\frac{m}{\alpha} + j\right)$$

Then, after simplification, we finally obtain

$$l(y, m, \alpha) = \sum_{i=1}^{n}\left[\sum_{j=0}^{y_i - 1}\ln\{1 + j(\alpha/m)\} + y_i\ln(m) - \left(y_i + \frac{m}{\alpha}\right)\ln(1 + \alpha) - \ln(y_i!)\right]$$

$$= \sum_{i=1}^{n}\left[\sum_{j=0}^{y_i - 1}\ln(1 + cj) + y_i\ln(m) - \left(y_i + \frac{1}{c}\right)\ln(1 + cm) - \ln(y_i!)\right]$$

As indicated earlier, to avoid full distributional assumptions, inference for m and α can be based on the DEQL. Now, to construct the DEQL we need to replace the log-likelihoods l_0 and l_1 on the right hand side of Equation (3.8) by the corresponding EQL functions. As described earlier, the negative binomial model in (3.1) can be written as a hierarchical Poisson-gamma GLM with

 (i) $Y_i|u_i \sim \mathrm{GLM}$ with $E(Y_i|u_i) = u_i = \mu_{0i}$ and $\mathrm{var}(Y_i|u_i) = \lambda V(\mu_{0i})$, where $\lambda = 1$ and $V(\mu_{0i}) = \mu_{0i}$,
(ii) $u_i \sim \mathrm{GLM}$ with $\theta(m) = \ln(m)$, $b(\theta(m)) = \exp(\theta(m))$, $E(u_i) = b'(\theta(m)) = m$ and $\mathrm{var}(u_i) = \alpha V(m)$, where $\alpha = cm$ and $V(m) = b''(\theta(m)) = m$.

To construct the DEQL for the above HGLM, we first need to define the EQLs of Nelder and Pregibon (1987) for both $Y_i|u_i$ and u_i distributions. Now, the EQL for $Y_i|u_i$ in (i) with mean and variance specified

above is

$$Q_0[\theta(\mu_0), \lambda; y|u] = -\frac{1}{2} \sum_{i=1}^{n} \left[d_{0i}/\lambda + \ln\{2\pi\lambda V(y_i)\} \right]$$

where $d_{0i} = 2\int_{\mu_{0i}}^{y_i} (y_i - t)/V(t)\mathrm{d}t = 2[y_i \ln(y_i/\mu_{0i}) - (y_i - \mu_{0i})]$ is the deviance components of $y|u$. Note that the log-likelihood $l_1(\alpha; v)$ can be viewed as the log-likelihood of quasi-data m with quasi-fixed parameters u and variance function $V(m) = m$. Thus, the EQL for the quasi-data m is

$$Q_1[u; m] = -\frac{1}{2} \sum_{i=1}^{n} [d_{1i}/\alpha + \ln(2\pi\alpha V(m)]$$

where $d_{1i} = 2\int_{u_i}^{m} (m - t)/V(t)\mathrm{d}t = 2[m \ln(m/u_i) - (m - u_i)]$ is the deviance components of u. With the scale change $v(m) = \theta(m) = \ln(m)$,

$$Q_1[u; \theta(m)] = Q_1[u; m] + \sum_{i=1}^{n} \ln\{|\partial m/\partial\theta(m)|\}$$

where $|\partial m/\partial\theta(m)| = V(m) = m$ is the Jacobian term.

Finally, combining the EQL for the data $Y_i|u_i$ and that for the quasi-data $\theta(m)$, the DEQL of Lee and Nelder (2001) is defined as

$$\mathrm{DEQ} = \sum_{i=1}^{n} \left[\left(y_i + \frac{1}{c} \right) v_i - \left(1 + \frac{1}{cm} \right) \exp(v_i) + \hbar_1(y_i, c) - \hbar_2(m, c) \right] \qquad (3.9)$$

where $c = \frac{\alpha}{m}$, $v_i = \theta(u_i) = \ln(u_i)$, $\hbar_1(y_i, c) = (y_i + \frac{1}{c}) - (y_i + \frac{1}{2})\ln(y_i) - \frac{1}{2}\ln(2\pi)$, and $\hbar_2(m, \alpha) = (\frac{1}{c} - \frac{1}{2})\ln(m) + \frac{1}{2}\ln(cm) + \frac{1}{2}\ln(2\pi)$. For estimating the mean m and the dispersion parameters c we need to integrate the unobserved random variables v_i. In order to avoid this multiple integration Lee and Nelder (1996, 2001) proposed to use the profile DEQL, which, in this case, is

$$p_v(\mathrm{DEQ}) = \left[\mathrm{DEQ} - \frac{1}{2} \ln \left| \det \left\{ \frac{1}{2\pi} D(\mathrm{DEQ}, v) \right\} \right| \right]_{v=\hat{v}}$$

where $D(\mathrm{DEQ}, v) = -\partial^2\mathrm{DEQ}/\partial v^2$, $\partial^2\mathrm{DEQ}/\partial v_i^2 = -(1 + \frac{1}{\alpha})u_i$, $\partial^2\mathrm{DEQ}/\partial v_i\partial v_j = 0$, for $i \neq j$ and $\hat{v}_i = \ln(\hat{u}_i)$ is the solution of $\partial\mathrm{DEQ}/\partial v = 0$ with $\hat{u}_i = (y_i + m/\alpha)/(1 + 1/\alpha)$. Let $\hat{\mathrm{DEQ}} = \mathrm{DEQ}|_{v=\hat{v}}$. Then, the profile DEQL can be written as

$$p_v(\mathrm{DEQ}) = \hat{\mathrm{DEQ}} - \frac{1}{2} \sum_{i=1}^{n} \ln \left\{ \frac{1}{2\pi} \left(1 + \frac{1}{cm} \right) \exp(\hat{v}_i) \right\}$$

$$= \sum_{i=1}^{n} \left[\vartheta_s(y_i, c) - \eta_s(c) - k_s(y_i, \lambda) - y_i \ln \left(1 + \frac{1}{cm} \right) - \frac{1}{c} \ln(1 + cm) \right]$$

where $\vartheta_s(y_i, c) = (y_i + \frac{1}{c} - \frac{1}{2})\ln(y_i + \frac{1}{c}) - (y_i + \frac{1}{c}) + \frac{1}{2}\ln(2\pi)$, $\eta_s(c) = (\frac{1}{c} - \frac{1}{2})\ln(\frac{1}{c}) - \frac{1}{c} + \frac{1}{2}\ln(2\pi)$, and $k_s(y_i, \lambda) = (y_i + \frac{1}{2})\ln(y_i) - y_i + \frac{1}{2}\ln(2\pi)$. It is interesting to note that $\vartheta_s(y_i, c)$, $\eta_s(c)$, and $k_s(y_i, \lambda)$ are the Stirling approximations of $\ln\Gamma(y_i + \frac{1}{c})$, $\ln\Gamma(\frac{1}{c})$, and $\ln(y_i!)$, respectively in the negative binomial model (3.1). The Stirling's approximation may not be good for small z, so they recommend to use the modified Stirling's approximation

$$\ln\Gamma(z) \simeq \left(z - \frac{1}{2}\right)\ln(z) + \frac{1}{2}\ln(2\pi) - z + \frac{1}{12z} \tag{3.10}$$

The profile DEQL with the modified Stirling approximation then is

$$p_v^*(\text{DEQ}) = \sum_{i=1}^{n}\left[y_i\ln(m) + \left(y + \frac{1}{c}\right)\ln\left(\frac{1+cy_i}{1+cm}\right) - \frac{1}{2}\ln(1+cy_i) - \left(y + \frac{1}{2}\right)\ln(y_i)\right.$$
$$\left. + \frac{c}{12(1+cy_i)} - \frac{c}{12} - \frac{1}{12y_i} - \frac{1}{2}\ln(2\pi)\right]$$

Inference for m and c can then be made based on $p_v^*(\text{DEQ})$.

3.2. Over-dispersed binomial data

For the binomial distributed data, $\theta(\mu) = \ln(\mu)/\{r_i - \mu\}$, $b(\theta(\mu)) = r_i\ln(1 + \exp(\theta(\mu))$, $\lambda = 1$, $V(\mu) = \mu(r_i - \mu)/r_i$, and $c(y, \lambda) = \ln[r_i!/\{y_i!(r_i - y_i)\}]$. The log-likelihood for the binomial regression model is the same as that given in Subsection 2.1. Also, for binomial data $l(y, \mu)$ and $Q(y, \mu)$ are identical.

Neither $l(y, \mu)$ nor $Q(y, \mu)$ is suitable for the estimation of over-dispersion. We can use a specific over-dispersion model for this. For example a convenient over-dispersion model for binomial data Y is the beta-binomial (ϕ, ρ) distribution

$$f(y_i|\phi, \rho) = Pr(y_i; \phi, \rho|r_i)$$
$$= \binom{r_i}{y_i}\frac{\prod_{j=0}^{y_i-1}[(1-\rho)\phi + j\rho]\prod_{j=0}^{r_i-y_i-1}[(1-\phi)(1-\rho) + j\rho]}{\prod_{j=0}^{r_i-1}[(1-\rho) + j\rho]} \tag{3.11}$$

for $y_i = 0, 1, 2, \ldots, 0 \le \phi \le 1$, and $\rho > 0$. Prentice (1986) extends the above distribution by allowing $max(\frac{-1}{r_i-1}) < \rho < 1$. The parameter ρ can also be interpreted as an intraclass correlation parameter. Since ρ can take a positive as well as a negative value it is called a dispersion parameter, rather than an over-dispersion parameter, and with this range of ρ, $f(y|\phi, \rho)$ is a valid probability function. Obviously, when $\rho \to 0$, the $BB(\phi, \rho)$ becomes Binomial(ϕ). Note that $E(Y_i) = r_i\phi$ and $var(Y_i) = r_i\phi(1-\phi)\{1 + (r_i - 1)\rho\}$.

The log-likelihood, apart from a constant, can be written as

$$l(y, \phi, \rho) = \sum_{i=1}^{n}\left[\sum_{j=0}^{y_i-1}\ln\{(1-\rho)\phi + j\rho\} + \sum_{j=0}^{r_i-y_i-1}\ln\{(1-\phi)(1-\rho) + j\rho\} - \sum_{j=0}^{r_i-1}\{(1-\rho) + j\rho\}\right]$$
$$\tag{3.12}$$

Inference for both ϕ and ρ can then be made using $l(y, \phi, \rho)$. However, it requires a full distributional assumption for the response variable to make an inference about ϕ and ρ. In absence of this assumption, an EQL might be useful for the joint estimation of both ϕ and ρ, which requires the specification of the variance function $V(Y)$, in this case, $V(Y) = n_i y_i (1 - y_i)\{1 + (r_i - 1)\rho\}$. The EQL for an observation $z_i = y_i / r_i$, apart from a constant, with this variance function is

$$Q^+(z_i, \phi, \rho) = -\frac{1}{2} \sum_{i=1}^{n} \left[\ln\{1 + (r_i - 1)\rho\} + \frac{2}{1 + (r_i - 1)\rho} \left\{ y_i \ln(\phi/z_i) + (r_i - y_i) \ln\left(\frac{1 - \phi}{1 - z_i}\right) \right\} \right]$$

(3.13)

Inference for the parameters ϕ and ρ can be made using $Q^+(z_i, \phi, \rho)$ (see Paul and Islam, 1995).

In a manner described in Subsection 3.1 a hierarchical likelihood can be obtained for the binomial data. Because, in general, the distributional assumptions are hard to justify, a DEQL is more sensible. In what follows we develop the DEQL for over-dispersed binomial data (see Lee, 2004).

We assume that $Y_i | p_i \sim$ binomial (r_i, p_i), for $i = 1, \ldots, n$ which can equivalently be written as a GLM with $\pi_{0i} = E(Y_i | p_i)$

$$f(y_i | p_i; \pi_{0i}, \lambda) = \exp[\{y_i \theta(\pi_{0i}) - b(\theta(\pi_{0i}))\}/\lambda + k(y_i, \lambda)]$$

(3.14)

where $\theta(\pi_{0i}) = \log(\pi_{0i}/\{r_i - \pi_{0i}\})$, $b(\theta(\pi_{0i})) = n_i \ln(1 + \exp(\theta(\pi_{0i})))$, $\lambda = 1$, and $k(y_i, \lambda) = \ln[r_i!/\{y_i!(r_i - y_i)\}]$. We assume further that the binomial probability p_i is distributed as a beta distribution with $E(p_i) = \phi$ and $\text{var}(p_i) = \phi(1 - \phi)\rho$. The kernel of this model can equivalently be written as a GLM

$$f(p_i; \phi, \delta) = \exp[\phi\theta(p_i) - b(\theta(p_i))]/\delta$$

(3.15)

where $\theta(p_i) = \ln(p_i/\{1 - p_i\})$, $b(\theta(p_i)) = \ln\{1 + exp(\theta(p_i))\}$, and $\delta = \rho/(1 - \rho)$. Thus, $f(p_i; \phi, \delta)$ can be viewed as the distribution of quasi-data ϕ with quasi-fixed parameters p and hence $\theta(p)$, satisfying the purely formal relationships $E(\phi) = p = b'(\theta(p))$ and $\text{var}(\phi) = \delta V(p)$, where $V(p) = b''(\theta(p))$, while the random effects p actually satisfy $E(p_i) = \phi$ and $\text{var}(p_i) = \rho V(\phi)$ for some $\rho = \delta/(1 - \delta)$. Then, the beta-binomial model in Equation (3.11) can be written as a hierarchical binomial-beta GLM with

(i) $Y_i | p_i \sim$ GLM with $E(Y_i | p_i) = n_i p_i = \pi_{0i}$ and $\text{var}(Y_i | p_i) = \lambda V(\pi_{0i})$, where $\lambda = 1$ and $V(\pi_{0i}) = \pi_{0i}(r_i - \pi_{0i})/r_i$,
(ii) $p_i \sim$ GLM with $E(p_i) = \phi$ and $\text{var}(p_i) = \rho V(\phi)$, where $V(\phi) = \phi(1 - \phi)$.

Now, to construct the DEQL for the above HGLM, we first need to define the EQLs of Nelder and Pregibon (1987) for both $Y_i | p_i$ and p_i distributions. So, the EQL for $Y_i | p_i$ distribution in (i) with mean and variance specified above is

$$Q_0[\theta(\pi_{0i}), \lambda; y|p] = -\frac{1}{2} \sum_{i=1}^{n} \left[d_{0i}/\lambda + \ln\{2\pi\lambda V(y_i)\} \right]$$

where $d_{0i} = 2\int_{\pi_{0i}}^{y_i}(y_i - t)/V(t)\mathrm{d}t = 2[y_i \ln(y_i/n_i p_i) + (r_i - y_i)\ln\{(r_i - y_i)/(r_i - n_i p_i)\}]$ is the deviance components of $y|p$. Then the EQL for the quasi-data ϕ is

$$Q_1[p;\phi] = -\frac{1}{2}\sum_{i=1}^{n}[d_{1i}/\delta + \ln\{2\pi\delta V(\phi)\}]$$

where $d_{1i} = 2\int_{p_i}^{\phi}(\phi - t)/V(t)\mathrm{d}t = 2[\phi\ln(\phi/p_i) + (1 - \phi)\ln\{(1 - \phi)/(1 - p_i)\}]$ is the deviance components of p. Hence, the EQL for the quasi-data $\theta(\phi) = \ln\{\phi/(1 - \phi)\}$ is

$$Q_1[p;\theta(\phi)] = Q_1[p;\phi] + \sum_{i=1}^{n}\ln\{|\partial\phi/\partial\theta(\phi)|\}$$

where $|\partial\phi/\partial\theta(\phi)| = V(\phi) = \phi(1 - \phi)$ is the Jacobian term.

Finally, combining the EQLs for the data $Y_i|p_i$ and the quasi-data $\theta(\phi)$ in above, the DEQL of Lee and Nelder (2001) for the above HGLM model becomes

$$Q^+ = \sum_{i=1}^{n}\left[\left(y_i + \frac{\phi}{\delta}\right)v_i - \left(r_i + \frac{1}{\delta}\right)\ln\{1 + \exp(v_i)\} + k_s(y_i, \lambda)\right] - n\eta_s(\phi, \delta) \tag{3.16}$$

where $v_i = \theta(p_i)$, $k_s(y_i, \lambda) = (r_i + \frac{1}{2})\ln(r_i) - (y_i + \frac{1}{2})\ln(y_i) - (r_i - y_i + \frac{1}{2})\ln(r_i - y_i) - \frac{1}{2}\ln(2\pi)$ and $\eta_s(\phi, \delta) = (\frac{\phi}{\delta} - \frac{1}{2})\ln(\phi/\delta) + (\frac{1-\phi}{\delta} - \frac{1}{2})\ln[(1 - \phi)/\delta] - (\frac{1}{\delta} - \frac{1}{2})\ln(1/\delta) + \frac{1}{2}\ln(2\pi)$. For estimating the mean parameter ϕ and the dispersion parameter δ, Lee and Nelder (1996, 2001) proposed to use the profile DEQL which is defined by

$$p_v(Q^+) = \left[Q^+ - \frac{1}{2}\det\left\{\frac{1}{2\pi}D\left(Q^+, v\right)\right\}\right]_{v=\hat{v}}$$

where $D(Q^+, v) = -\partial^2 Q^+/\partial v^2$ and $\hat{v}$ is the solution of $\partial Q^+/\partial v = 0$. In the beta-binomial model, $\partial^2 Q^+/\partial v_i^2 = -(r_i + \frac{1}{\delta})p_i(1 - p_i)$, $\partial^2 Q^+/\partial v_i\partial v_j = 0$, for $i \neq j$ and $\hat{v}_i = \ln\{\hat{p}_i/(1 - \hat{p}_i)\}$, where $\hat{p}_i = (y_i + \phi/\delta)/(r_i + 1/\delta)$. Finally, the resulting profile DEQL becomes

$$p_v(Q^+) = \sum_{i=1}^{n}[\vartheta_s(y_i, \phi, \delta) + k_s(y_i, \lambda)] - n\eta_s(\phi, \delta) \tag{3.17}$$

where $\vartheta_s(y_i, \phi, \delta) = (y_i + \frac{\phi}{\delta} - \frac{1}{2})\ln(y_i + \frac{\phi}{\delta}) + (r_i - y_i + \frac{1-\phi}{\delta} - \frac{1}{2})\ln(r_i - y_i + \frac{1-\phi}{\delta}) - (r_i + \frac{1}{\delta} - \frac{1}{2})\ln(r_i + \frac{1}{\delta}) + \frac{1}{2}\ln(2\pi)$. Note here that $\vartheta_s(y_i, \phi, \delta)$, $\eta_s(\phi, \delta)$, and $k_s(y_i, \lambda)$ are , respectively, the Stirling approximation to $\vartheta(y_i, \phi, \delta) = \ln\Gamma(y_i + \frac{\phi}{\delta}) + \ln\Gamma(r_i - y_i + \frac{1-\phi}{\delta}) - \ln\Gamma(r_i + \frac{1}{\delta})$, $\eta(\phi, \delta) = \ln\Gamma(\frac{\phi}{\delta}) + \ln\Gamma(\frac{1-\phi}{\delta}) - \ln\Gamma(\frac{1}{\delta})$, and $k(y_i, \lambda)$ in beta-binomial log-likelihood, using $\ln\Gamma(z) \simeq (z - 0.5)\ln(z) + \frac{1}{2}\ln(2\pi) - z$. The ordinary Stirling approximation may not be good for small z. So for estimating ϕ and δ, Lee and Nelder (2001) and Lee (2004) recommend to use the modified Stirling approximation given by Equation (3.10).

The profile DEQL using Equation (3.10), apart from a constant, can be written as

$$p_v(Q^+) = \sum_{i=1}^{m} \left[\left(\frac{P_{1i}}{\delta} - \frac{1}{2} \right) \ln \left(\frac{P_{1i}}{\delta} \right) + \left(\frac{P_{2i}}{\delta} - \frac{1}{2} \right) \ln \left(\frac{P_{2i}}{\delta} \right) - \left(\frac{P_{3i}}{\delta} - \frac{1}{2} \right) \ln \left(\frac{P_{3i}}{\delta} \right) \right.$$
$$\left. + \frac{\delta}{12 P_{1i}} + \frac{\delta}{12 P_{2i}} - \frac{\delta}{12 P_{3i}} \right] - m \left(\frac{\phi}{\delta} - \frac{1}{2} \right) \ln \left(\frac{\phi}{\delta} \right) - m \left(\frac{1 - \phi}{\delta} - \frac{1}{2} \right) \ln \left(\frac{1 - \phi}{\delta} \right)$$
$$+ m \left(\frac{1}{\delta} - \frac{1}{2} \right) \ln \left(\frac{1}{\delta} \right) - \frac{m\delta}{12\phi} - \frac{m\delta}{12(1 - \phi)} + \frac{m\delta}{12},$$

where $P_{1i} = \phi + \delta y_i$, $P_{2i} = 1 - \phi + \delta(r_i - y_i)$, and $P_{3i} = 1 + \delta r_i$.

Hence, the DEQL estimators of ϕ and δ can be obtained by solving the profile QL score equations

$$\frac{\partial p_v(Q^+)}{\partial \phi} = \sum_{i=1}^{n} \left[\frac{1}{\delta} \ln \left(\frac{P_{1i}(1 - \phi)}{\phi P_{2i}} \right) + \frac{P_{1i} - P_{2i}}{2 P_{1i} P_{2i}} \left(1 + \frac{\delta P_{3i}}{6 P_{1i} P_{2i}} \right) + \frac{2\phi - 1}{2\phi(1 - \phi)} \left(1 + \frac{\delta}{6\phi(1 - \phi)} \right) \right]$$

and

$$\frac{\partial p_v(Q^+)}{\partial \delta} = \sum_{i=1}^{n} \left[\frac{\phi}{\delta^2} \ln \left(\frac{\phi P_{2i}}{P_{1i}(1 - \phi)} \right) + \frac{1}{\delta} \ln \left(\frac{(1 - \phi) P_{3i}}{P_{2i}} \right) + \frac{\phi}{2 P_{1i}} \left(\frac{1}{\delta} + \frac{1}{6 P_{1i}} \right) \right.$$
$$\left. + \frac{1 - \phi}{2 P_{2i}} \left(\frac{1}{\delta} + \frac{1}{6 P_{2i}} \right) - \frac{1}{2 P_{3i}} \left(\frac{1}{\delta} + \frac{1}{6 P_{3i}} \right) - \frac{1}{2\delta} - \frac{1 - \phi(1 - \phi)}{12\phi(1 - \phi)} \right]$$

simultaneously. The DEQL estimators for ϕ and δ can be obtained by solving the above equations and hence we can obtain DEQL estimator for $\rho = \delta/(1 + \delta)$.

4. SIMULATIONS

4.1. Simulations for over-dispersed poisson case

In this section we conducted a simulation study to compare bias and efficiency properties of ML, EQL, and DEQL estimators of the dispersion parameter for data generated from a over-dispersed Poisson model. We considered the $NB(m, c)$ as the over-dispersed Poisson model. We have generated samples of sizes $n = (10, 50)$ from negative binomial distribution for all combinations of $m = (1, 3, 5)$ and $c = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 0.9)$. In the simulations small means have been considered as this is often relevant to environmental data (see the means in Example 1 in Section 5). Further, for small means the asymptotic properties for the estimator of the dispersion parameter c can be misleading, so in such situations properties for this estimator must be studied through simulations. Estimates of c by all three methods were constrained to be ($\hat{c} > -1/y_{max}$, where y_{max} is the largest y value). We computed the mean bias and the mean squared error (MSE) for each estimator based on 10 000 simulation runs.

Each estimating equation was solved by using the IMSL subroutine NEQNF. In the process of sampling and estimation if an estimate of c by any of the methods was $< -1/y_{max} + 0.00001$ the sample was discarded. Also for some samples convergence of some of the estimating equations (the ML estimating equation or some others) were not achieved. The samples were discarded in this situation
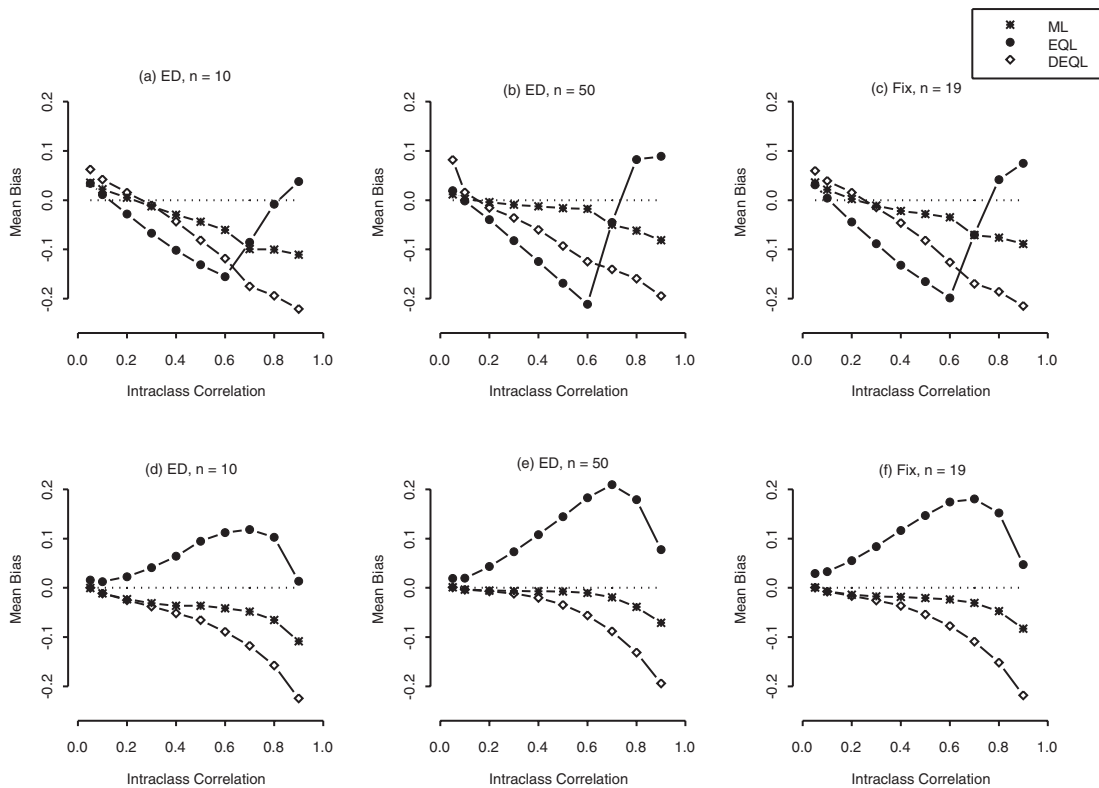
Figure 1.   Mean biases of three estimators of c by the ML, EQL, and DEQL methods when data simulated from negative binomial (over-dispersed Poisson) distribution under varying simulation conditions: two values of $n = 10[(a), (b), (c)]$, $50[(d), (e), (f)]$, and $m = 1[(a), (d)]$, $3[(b), (e)]$, $5[(c), (f)]$

also. It is to be noted that convergence of an estimating equation depended on the initial value required to be provided in the subroutine. So, if the same initial value is used for all samples a large number of samples need to be discarded. In order to minimize number of rejected samples we used repeated initial values $-0.9911/y_{max} + 0.001(i - 1)$, $i = 1, \ldots, 1000$. In Table 1 we provide the number of runs discarded for $10\,000$ acceptable simulated samples. From Table 1 we see that larger number of samples were discarded for smaller values of m and c. This is reasonable because convergence problem is expected to occur more often at the boundary value of the parameter space of c and small values of c are expected for smaller values of m.

Then, relative efficiencies (REs) of the estimators, t, where $t = $ EQL, DEQL, were calculated by $RE(t) = MSE(ML)/MSE(t)$, where, MSE of an estimator t is obtained as $MSE(t) = \sum_{i=1}^{10\,000}(t_i - c)^2/(10\,000)$. The bias results are presented in Figure 1 and the efficiency results are presented in Figure 3.

From Figure 1 we see that bias of the EQL estimate is the lowest and that of the DEQL estimate is the highest except for $m = 1$. For $n = 50$, the EQL estimate has the smallest bias which is near zero. For $m = 1$ and $n = 50$, all the estimates have larger bias, however, the DEQL estimate has the smallest bias. For $m = 1$ and $n = 50$, the ML estimate has the smallest bias. On the average, the EQL estimate has
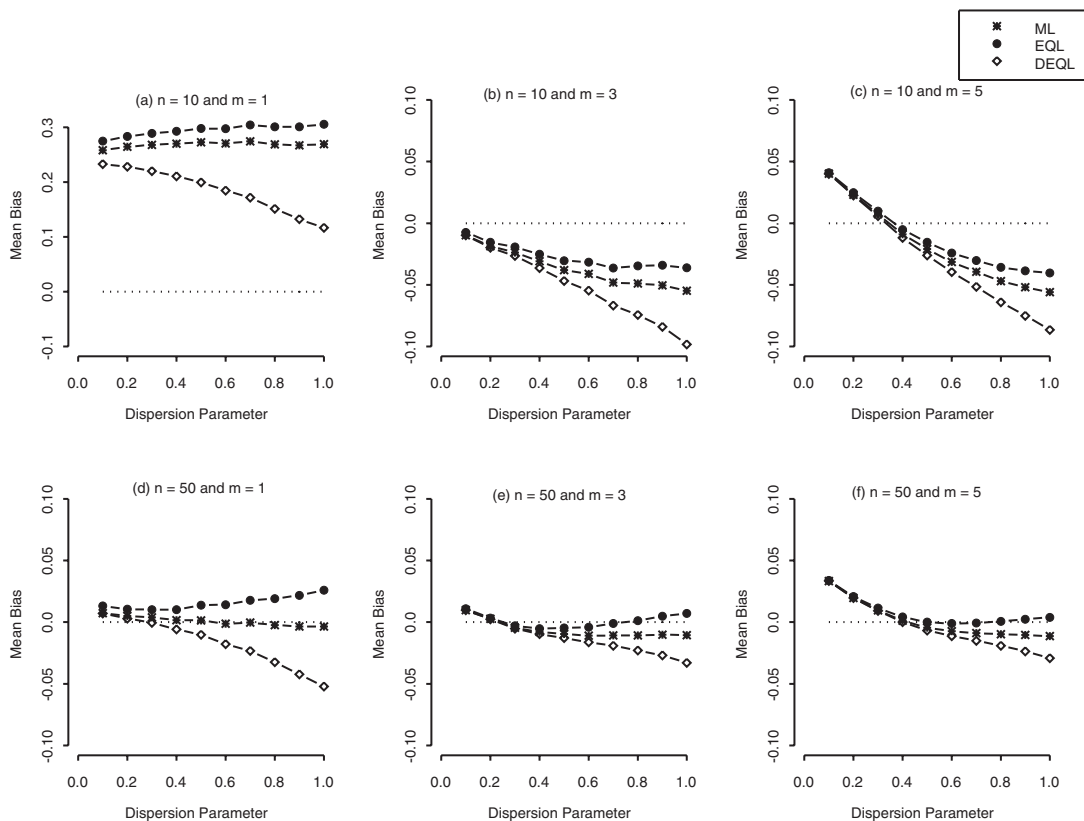
Figure 2. Mean biases of three estimators of ρ by the ML, EQL, and DEQL methods when data simulated from beta-binomial (over-dispersed binomial) distribution under varying simulation conditions: two values of $n = 10$ [(a), (d)], 19 [(c), (f)], 50 [(b), (e)], and $\phi = 0.1$ [(a), (b), (c)], 0.4 [(d), (e), (f)]. The group sizes were generated from an empirical distribution (ED) [(a), (b), (d), (e)] and Fixed group sizes (Fix) were taken from Paul (1982) [(c), (f)]

the smallest bias except for small m. Figure 1 also shows that all the estimates are positively biased and the biases are relatively larger for $m = 1$ and $n = 10$. Whereas, in most other instances, the biases are small and negative. The reason is not immediately apparent, although large biases may result because of the small sample size and small value of m.

From Figure 3 we see that the efficiency of the DEQL estimate is the best over the parameter space examined. The efficiency of the DEQL estimate never drops below that of the ML estimate whereas the efficiency of the EQL estimate drops below that of the ML estimate.

4.2. Simulations for over-dispersed binomial case

In this simulation study, we investigated the performance of ML, EQL, and DEQL estimators for the intraclass correlation parameter ρ in terms of bias and efficiency properties. Simulations were run for two values of ϕ (0.1, 0.4), ten values of ρ (0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) and three values of n (10, 19, 50). Following Paul *et al*. (2003), we considered only the distribution of group

Figure 3. Relative efficiency of two estimators of c, EQL, and DEQL, when data simulated from negative binomial (over-dispersed Poisson) distribution under varying simulation conditions: two values of $n = 10$ [(a), (b), (c)], 50 [(d),(e),(f)], and $m = 1$ [(a), (d)], 3 [(b), (e)], 5 [(c), (f)]

sizes as the empirical distribution (ED) of 523 litter sizes, quoted by Kupper *et al*. (1986) where the litter sizes range was from 1 to 19 with a mean of 12 and standard deviation 2.98. For each combination of (ϕ, ρ) we generated data set from the beta-binomial distribution using the litter sizes given above. Further, simulations were run for the same values of ϕ and ρ, and the same probability distributions above using the fixed litter sizes of the low-dose group of Paul (1982) in which $n = 19$ and litter sizes $r_i = 5, 11, 7, 9, 12, 8, 6, 7, 6, 4, 6, 9, 6, 7, 5, 9, 1, 6, 9$. Using 10 000 acceptable simulation runs based on the same criteria as that of Paul *et al*. (2003) for rejection of data, we computed the mean bias and the MSE for each estimator. In Table 2 we provide the number of runs discarded for 10 000 acceptable simulated samples. We also see here that larger number of runs was discarded for smaller values of ϕ and ρ. Then, RE of each of the EQL and DEQL estimators was calculated using the formula given above. The bias efficiency results are plotted in Figures 2 and 4 and are summarized in what follows.

The biases of the EQL and DEQL estimates are larger than those of the ML estimate. However, the bias of the DEQL is smaller than that of the EQL estimate. Also, for small values of ρ the bias of the DEQL is consistently near to that of the ML estimate except few cases. The behavior of the bias property of the EQL estimate is not consistent.

Table 2. Number of runs discarded for 10 000 acceptable samples for over-dispersed binomial case, $\phi = 0.1, 0.4$ and $\rho = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

Data distribution	Litter size distribution	n	ϕ	ρ									
				0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BB	ED	10	0.1	8970	4808	2891	2735	2719	3258	3769	935	523	691
			0.4	1952	687	142	31	15	20	54	97	283	1393
		50	0.1	5798	1193	188	174	283	365	586	15	2	0
			0.4	1425	154	4	0	0	0	1	0	0	10
	Fix	19	0.1	7821	3663	1420	1090	1081	1369	1670	138	115	465
			0.4	1247	376	42	6	9	33	109	251	362	422

The efficiency properties of the EQL and DEQL estimates are inconsistent for $\phi = 0.1$. For $\phi = 0.1$, the efficiency of the DEQL estimate drops below than that of the ML estimate over the parameter space examined. However, the DEQl estimate has the best efficiency property and its efficiency is above 1 for $\phi = 0.4$ and smaller values of $\rho (\leq 0.5)$.
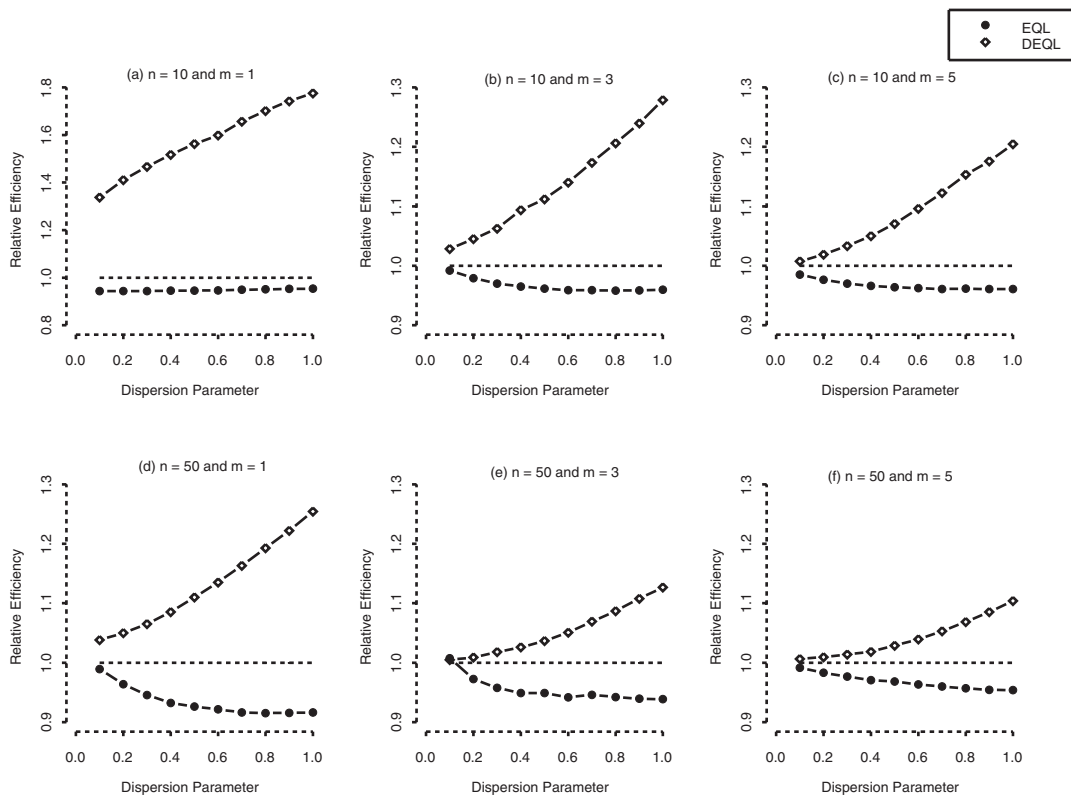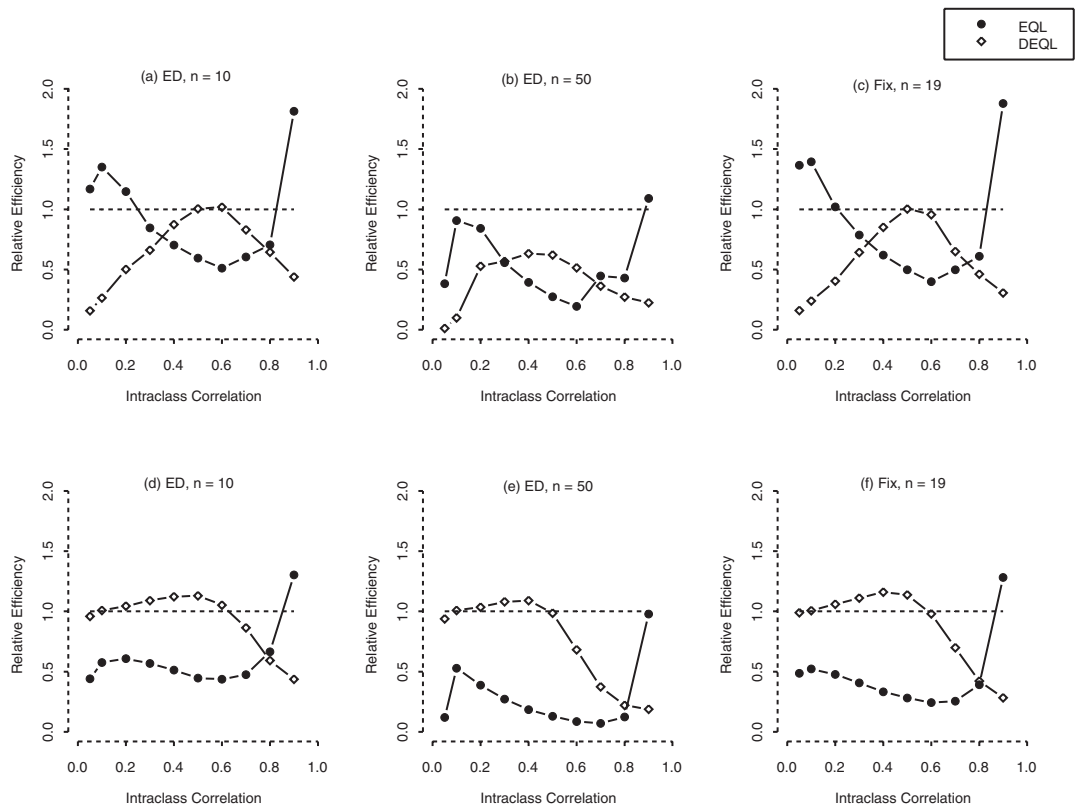

Figure 4. Relative efficiency of two estimators of ρ, EQL, and DEQL, when data simulated from beta-binomial (over-dispersed binomial) distribution under varying simulation conditions: two values of $n = 10$ [(a), (d)], 19 [(c), (f)], and 50 [(b),(e)], and $\phi = 0.1$ [(a), (b), (c)], 0.4 [(d), (e), (f)]. The group sizes were generated from an empirical distribution (ED) [(a), (b), (d), (e)] and Fixed group sizes (Fix) were taken from Paul (1982) [(c), (f)]

Table 3.   Distribution of corn borers in a field experiment arranged in 15 randomized blocks from Bliss and Fisher (1953)

| Borers per hill | Observations for | | | |
	Treatment 1	Treatment 2	Treatment 3	Treatment 4
0	19	24	43	47
1	12	16	35	23
2	18	16	17	27
3	18	18	11	9
4	11	15	5	7
5	12	9	4	3
6	7	6	1	1
7	8	5	2	1
8	4	3	2	
9	4	4		
10	1	3		1
11				1
12	1	1		
13	1			
15	1			
17	1			
19	1			
26	1			

5. EXAMPLES

Example 1 (Biological data).   The data in Table 3, originally given and analyzed by Bliss and Fisher (1953), refer to the total number of borers per plot and as frequency distributions showing for each treatment the number of hills with borers. In a field experiment of insect pests on the corn borer, four treatments were arranged in 15 randomized blocks. At the end of the season, eight hills of corn were selected at random in each plot and the borers recorded from each hill.

Bliss and Fisher (1953) analyzed these data by fitting separate negative binomial distributions for each of the four treatment groups. The estimates of the parameters m and c for the four groups by the ML, EQL, and DEQL methods are given in Table 4 and their estimated standard errors are given in Table 5. The maximum likelihood estimates of m and c and their estimated standard errors are in

Table 4.   Parameter estimates by the three methods for each treatment group for data in Table 3

| | Estimates of m | | | Estimates of c | | |
	ML	EQL	DEQL	ML	EQL	DEQL
Treatment 1	4.0333	4.0333	4.0333	0.6654	0.6734	0.6606
Treatment 2	3.1667	3.1667	3.1667	0.5680	0.5752	0.5644
Treatment 3	1.4833	1.4833	1.4833	0.7501	0.7664	0.7384
Treatment 4	1.5083	1.5083	1.5083	0.8669	0.8877	0.8493

Table 5. Standard errors (SE) of the estimates by the three methods for data in Table 3

	SE (Estimates of m)			SE (Estimates of c)		
	ML	EQL	DEQL	ML	EQL	DEQL
Treatment 1	0.3519	0.3535	0.3510	0.1266	0.1290	0.1245
Treatment 2	0.2718	0.2729	0.2713	0.1299	0.1323	0.1281
Treatment 3	0.1616	0.1625	0.1610	0.2101	0.2160	0.2039
Treatment 4	0.1703	0.1715	0.1694	0.2339	0.2411	0.2248

agreement with those given by Bliss and Fisher (1953). Note that Bliss and Fisher (1953) provided the results with respect to the parameters m and $k = c^{-1}$.

From Table 5, we see that the standard errors of the DEQL estimates of m and c show some improvement over their ML and EQL estimates for all treatment groups.

Example 2 (2,4,5-trichlorophenoxiacetic acid data). The data are given in Table 6 (see also Table 3 in Bowman and George, 1995) from a large-scale developmental toxicology study of the herbicide 2,4,5-trichlorophenoxiacetic acid (2,4,5-T) conducted at the US National Center for Toxicological Research. From day 6 to day 14 of gestation, pregnant mice from several strains were given daily does of 2,4,5-T and recorded here only the number of implantation sites, foetal deaths, resorptions, and cleft palate malformations from the outbreed strain CD-1 for each female mouse. For this outbreed strain CD-1, there were six dose groups corresponding to exposure levels of 0, 30, 45, 60, 75, and 90 mg/kg of 2,4,5-T. A summary of this data set is also given in Table 7. Bowman *et al.* (1995) analyzed these data by fitting a separate beta-binomial distribution for each dose group and used response as a combined end point of death, resorptions, or malformation.

Table 6. Frequency distribution of the number of implants (n) and number of combined endpoints (y) following exposure to 2,4,5-T

Dose		Number of Implants (n)																				
(mg/kg)	y	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	0	1		1		1	1	2	1	5	4	6	7	2	2							
	1					1				2	1	7	6	5	2			1				
	2									1	1	5	3		1			1				
	3													1			1					
	4														1							
30	0		1	1	2		2	2	1	2	4	8	2	3	2	1						
	1							1	1	2	8	4	5	5	2	2						
	2							2	2		1	1	2	1	2	1						
	3										1		3	2		1						
	4											1	1									
	5														1							
	6									1												
	7													1								
	8														1							
	15																		1			

(continued)

Table 6.   *(continued)*

Dose (mg/kg)	y	Number of Implants (n)																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
45	0	1		1	1				1	1	1	8	3		1							
	1			1		1		1	1	3	2	4	3	1	1	1						
	2					1	1			1		5	4	3	2	1						
	3											1	2	3	1	1	1					
	4									1			1		2	1						
	5										1	1	1	3	1	1						
	6										1		2		3	1	1					
	7												1									
	8									1				1								
	9										1											
	10											1										
	11												1									
	12												2									
	13													2	1							
	14														1							
	18																					1
60	0	1								1		1	3			1						
	1				1					2	2	4	2	1	1							
	2							1	1	1	1			4		1						
	3			2						1				3	1							
	4					1				1		2										
	5						2			1			2	1			1					
	6						1			1												
	7							1						1		1						
	8								2	1				2								
	9									1		2	1									
	10											5	1									
	11											2		1								
	12												3									
	13													2								
	14														2							
75	0									1		1										
	1								1													
	2											2	2									
	3									1												
	4										1											
	5					2			1	1												
	7									1					1							
	8								1		1	1										
	9									3	1											
	10										1		1	1								
	11											4	2									
	12												3	4								
	13													3								
	14														2							
	15														1							

(continued)

Table 6. *(continued)*

90	0	1								
	3	1								
	4		1			1				
	6			1						
	7				1					
	10						4	2		1
	11							2		
	12								6	
	13									2
	14									2

The estimates of the parameters ϕ and ρ and their estimated standard errors for the six dose categories by the ML, EQL, and DEQL methods are shown in Tables 8 and 9, respectively. These estimates are obtained by solving the corresponding joint estimating equations of ϕ and ρ iteratively. Standard errors of the EQL and DEQL estimates are obtained by computing the variance–covariance matrices based on the method given by Inagaki (1973). The maximum likelihood estimates are not in agreement with those given by Bowman *et al.* (1995). Note that a summary data given by Bowman *et al.* (1995) does not match with the summary data we found in Table 7. So there could be some typo in the data set in Table 3 given by Bowman and George (1995).

Standard errors of the DEQL estimates of ϕ and ρ show much improvement over their ML and EQL estimates for all dose groups.

Table 7. Summary data of litters for the data set given in Table 6

Dose (mg/kg)	Responses	Inplants	Litters
0	59	802	73
30	124	952[a]	87
45	338	1126[a]	98
60	390	806	76
75	372	482	44
90	242	254[a]	25

[a] *Note*: These values are not same as reported by Bowman *et al.* (1995).

Table 8. Parameter estimates by the three methods for each treatment group for data in Table 6

Dose (mg/kg)	Estimates of ϕ			Estimates of ρ		
	ML	EQL	DEQL	ML	EQL	DEQL
0	0.073724	0.073278	0.073724	−0.004152	0.008620	−0.004160
30	0.127513	0.124082	0.127630	0.110606	0.126375	0.108414
45	0.311507	0.286992	0.312224	0.313689	0.382804	0.305558
60	0.545865	0.505713	0.542144	0.525582	0.731732	0.496890
75	0.759171	0.763098	0.751907	0.508362	0.591027	0.466559
90	0.925876	0.942139	0.927062	0.499571	0.126536	0.337872

Table 9.   Standard errors (SE) of the estimates by the three methods for data in Table 6

Dose (mg/kg)	SE (Estimates of ϕ)			SE (Estimates of ρ)		
	ML	EQL	DEQL	ML	EQL	DEQL
0	0.017857	0.009614	[b]	0.033404	0.021941	[b]
30	0.031886	0.016219	0.006644	0.056257	0.050624	0.004076
45	0.058733	0.030409	0.043792	0.064687	0.060549	0.004316
60	0.075806	0.050060	0.061330	0.06282	0.045272	0.000807
75	0.060472	0.050888	0.048936	0.062987	0.145983	0.000745
90	0.033726	0.021879	0.026570	0.072088	0.146242	0.001006

[b] *Note*: Standard error does not exist.

6. DISCUSSION

The main purpose of our paper is to review the robust procedures that stem from the GLM and the hierarchical generalized linear model (HL) and to show their applications to discrete data analysis with over-dispersion. The EQL is the natural extension of the GLM to situations where full distributional assumptions are not satisfied. Similarly, the DEQL is the natural extension of the HL. The HL requires full distributional assumptions at both stages of the hierarchy, whereas, the DEQL requires the assumption of only the first two moments at both stages. For the analysis of discrete data with over-dispersion the GLMMs can be used. Again, the GLMM is based on full distributional assumptions. So, we did not consider this in our review. We have attempted to critically review GLM and the HL procedures and their robust counterparts EQL and DEQL. We have also reviewed methodology for analyzing over-dispersed binomial and over-dispersed Poisson data based on full distributional assumptions (beta-binomial for over-dispersed binomial data and negative binomial for over-dispersed Poisson data) and those based on the EQL and DEQL. Further, we showed applications to some biological and environmental data.

Our simulations show that even when over-dispersed Poisson data arise from a negative binomial distribution the ML estimate is not a clear winner in terms of bias, and efficiency of the DEQL estimator is larger than 1.00 and that of the EQL, in general, is smaller than 1.00, indicating that the variance of the DEQL estimate is the smallest. Similar property, in general, holds for the DEQL estimator when over-dispersed binomial data are generated from the beta-binomial distribution. The real biological and environmental data analysis in Section 5 supports this finding in that the standard errors of the DEQL estimates are the smallest.

Within the parametric model framework Lee and Nelder (2006) extended their HL model to double HGLMs. This new class of models incorporates many existing models including the HGLM. This extension is also based on full parametric distributional assumptions and therefore may not be robust when data distributions do not satisfy the full distributional assumptions.

REFERENCES

Bliss CI, Fisher RA. 1953. Fitting the negative binomial distribution to biological data. *Biometrics* **9**: 176–200.
Bowman D, George EO. 1995. A saturated model for analyzing exchangeable binary data: applications to clinical and developmental toxicity studies. *Journal of the American Statistical Association* **90**: 871–879.
Bowman D, Chen JJ, George EO. 1995. Estimating variance functions in developmental toxicity studies. *Journal of the American Statistical Association* **90**: 871–879.

Breslow NE, Clayton DG. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**: 9–25.

Godambe VP, Thompson ME. 1989. An extension of quasi-likelihood estimation. *Journal of Statistical Planning and Inference* **22**: 137–152.

Inagaki N. 1973. Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator. *Annals of the Institute of Statistical Mathematics* **25**: 1–26.

Kupper LL, Portier C, Hogan MD, Yamamoto E. 1986. The impact of litter effects on dose-response modeling in teratology. *Biometrics* **42**: 85–98.

Lawless JF. 1987. Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* **15**: 209–225.

Lee Y, Nelder JA. 1996. Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B* **58**: 619–678.

Lee Y, Nelder JA. 2001. Hierarchical generalized linear models: asynthesis of generalized linear models, random effect models and structured dispersions. *Biometrika* **88**: 987–1006.

Lee Y. 2004. Estimating intraclass correlation for binary data using extended quasi-likelihood. *Statistical Modelling* **4**: 113–126.

Lee Y, Nelder JA. 2006. Double hierarchical generalized linear models. *Applied Statistics* **55**(Part 2): 139–185.

Nelder JA, Pregibon D. 1987. An extended quasi-likelihood function. *Biometrika* **74**: 221–232.

Nelder JA, Wedderburn RWM. 1972. Generalized linear models. *Journal of the Royal Statistical Society, Series A* **135**: 370–384.

Paul SR. 1982. Analysis of proportions of affected foetuses in teratological experiments. *Biometrics* **38**: 361–370.

Paul SR, Islam AS. 1995. Analysis of proportions in the presence of over/under-dispersion. *Biometrics* **51**: 1400–1410.

Paul SR, Saha KK, Balasooriya U. 2003. An empirical investigation of different operating characteristics of several estimators of the intraclass correlation in the analysis of binary data. *Journal of Statistical Computation and Simulation* **73**: 507–523.

Prentice RL. 1986. Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association* **81**: 321–327.

Schall R. 1991. Estimation in generalized linear models with random effects. *Biometrika* **40**: 917–927.

Wedderburn RWM. 1974. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61**: 439–447.