Journal of Statistical Research *ISSN 0256 - 422 X*
2010, Vol. 44, No. 1, pp. 147-165
Bangladesh

VARIANCE FUNCTION IN SEMI-PARAMETRIC ANALYSIS OF COUNT DATA

Sudhir R. Paul
Department of Mathematics and Statistics, University of Windsor,
401 Sunset, Windsor, ON, N9B 3P4, Canada
Email: smjp@uwindsor.ca

Tasneem Zaihra
Department of Computer Science and Applied Statistics
University of New Brunswick, Saint John, NB, E2L4L5, Canada
Email: tzaihra@unb.ca

summary

The purpose of this paper is to determine an appropriate variance function (mean-variance relationship) which can be used in the semi-parametric analysis of over-dispersed count data (for example, for analysis of count data by extended quasi-likelihood and double extended quasi-likelihood). We use hypothesis testing approach through a broader class of models and data analytic approach. The models considered are the three parameter negative binomial distribution and the extended quasi-likelihood. Wide analysis involving tests, data analysis and simulations indicate that the three parameter generalized negative binomial distribution does not improve in fit to count data over the simpler negative binomial distribution. Further data analysis and simulations using the extended quasi-likelihood indicate that the negative binomial variance function $\mu + c\mu^2$ is preferable over a simpler variance function $c_3\mu^2$ for data with small mean and small over-dispersion. Otherwise $c_3\mu^2$ is a preferable variance function over the negative binomial variance function.

Keywords and phrases: Dispersion parameter; Extended quasi-likelihood; Extended quasi-likelihood estimator; Negative binomial model; Three parameter negative binomial model; Variance function.

1 Introduction

Count data with over-dispersion arise in many diverse fields, including biostatistics, radioimmunoassay, econometrics, pharmacokinetics modelling, enzyme kinetics, chemical kinetics, quality control among others. For example, Bliss and Fisher (1953) present a set of count data (see Table 3) consisting of the number of European red mites on apple leaves for which the mean and the variance are 1.15 and 2.27 respectively showing that the variance exceeds the mean. The embryonic death counts data set by McCaughran and Arnold (1976), given in Table 1 have similar properties having mean 1.20 and variance 1.733.

These data are often modelled using a negative binomial distribution. See, for example, Anscombe (1949), Bliss and Fisher (1953), Bliss and Owen (1958), McCaughran and Arnold, (1976). Different authors have expressed the negative binomial distribution in different forms; see, for example, Bliss and Fisher (1953), Johnson and Kotz (1969), Bliss and Owen (1958) and Collings and Margolin (1985). The most convenient is that proposed by Bliss and Owen (1958) and used by Collings and Margolin (1985), Barnwal and Paul (1988), Paul and Banerjee (1998) and others in which the random variable Y follows a negative binomial distribution with mean μ and coefficient c, denoted by $Y \sim NB(\mu, c)$, if

$$Pr(Y=y) = \frac{\Gamma(y+c^{-1})}{y!\ \Gamma(c^{-1})}\Big(\frac{1}{1+c\mu}\Big)^{c^{-1}}\Big(\frac{c\mu}{1+c\mu}\Big)^{y} \tag{1.1}$$

for $y = 0, 1, \ldots, 0 < \mu < \infty, 0 < c < \infty$. Here $E(y) = \mu$ and $Var(y) = \mu(1 + c\mu)$. Evidently the $NB(\mu, c)$ distribution becomes the Poisson distribution in the limit when $c \to 0$. Further properties of $NB(\mu, c)$ are given in Paul and Plackett(1978).

However, in many practical data analysis situations, sometimes, a full distributional assumption becomes restrictive. More robust analysis are performed using some semi-parametric model, such as the extended quasi-likelihood (Nelder and Pregibon, 1987) and the double extended quasi-likelihood (Lee and Nelder, 2001) (these are semi-parametric models as these require assumption of only the first two moments). Intrinsic in the semi-parametric analysis of count data is the assumption of the variance function. The most popular variance function is that given by the Negative binomial, namely $Var(Y) = \mu(1+c\mu) = \mu + c\mu^2$ (see, for example, Paul and Banerjee, 1998). Other variance functions can also be used. For example, Bartlet (1936) uses a variance function $Var(Y) = c_1\mu + c_2\mu^2$ to analyze counts for field experiments, where c_1 and c_2 are parameters to be estimated from the data. A similar expression is $Var(Y) = c_3\mu^b$. Both Armitage (1957) and Finney (1976) find by study of examples that $1 < b < 2$. The variance function $Var(Y) = c_3\mu^b$ with $b = 2$, that is, $Var(Y) = c_3\mu^2$ is identical to the second part of the negative binomial variance function, although, in practical data analysis, the value of c in $Var(Y) = \mu + c\mu^2$ and c_3 in $Var(Y) = c_3\mu^2$ would probably be different. Finney (1976) provide reference to Rodbard and Hutt (1974) who commented from personal experience that for immunoradiometric assays the linear term in $Var(Y) = c_1\mu + c_2\mu^2$ can be omitted, therefore proving support for the variance function $Var(Y) = c_3\mu^2$.

A more generalized variance function is $Var(Y) = \mu(1 + c\mu^b)$ obtained from a three parameter generalization of the negative binomial distribution developed by Cameron and Trivedi(1986). The probability function of this distribution, denoted by $NB(\mu, c, b)$, is

$$Pr(Y=y) = \frac{\Gamma(y+c^{-1}\mu^{1-b})}{y!\,\Gamma(c^{-1}\mu^{1-b})}\Big(\frac{1}{1+c\mu^b}\Big)^{c^{-1}\mu^{1-b}}\Big(\frac{c\mu^b}{1+c\mu^b}\Big)^{y}. \tag{1.2}$$

Note that the $NB(\mu, c, b)$ distribution reduces to the $NB(\mu, c)$ distribution for b=1. Therefore the three parameter generalized model can be used to test the goodness of fit of the negative binomial distribution, that is, to test the variance function $Var(Y) = \mu + c\mu^2$ against $Var(Y) = \mu(1 + c\mu^b)$.

The purpose of this paper is to choose an appropriate variance function (mean-variance relationship) which can be used in the semi-parametric analysis of count data.

Relationship between mean and variance, in count data, data in the form of proportions and also for some continuous data, has been of interest in many fields. For instance ecologists use power function relationship between the variance and mean number of organisms that reflects the spatial heterogeneity of a population within it's habitat (Kendal, 2004). The mean-variance relationship is helpful in finding the underlying distributions, for instance, if the variance seems to be proportional to the square of the mean, then the family consisting of gamma, log normal or the Weibull distributions might be interest and if the variance seems to be proportional to the cubic power of the mean then an inverse Gaussian model may be appropriate. A proper specification of the mean variance relationship helps in reliable inferences. Kilpatrick and Ives (2003) explains the relationship between mean and variance by means of probabilistic models based on negative interactions among species and spatial heterogeneity. Recently, Gaffeo, et al. (2008) extended the concept to a national industrial system that is, a single taxonomic group and discussed mean variance relationship of the firm size distribution. The mean-variance relationship is also of interest in binomial data (see Williams 1982).

We use hypothesis testing approach through a broader class of models, data analytic approach and some simulations. The models considered are the three parameter negative binomial distribution and the extended quasi-likelihood. In this context some theoretical and computational problems of the three parameter negative binomial distribution are discussed and some insights are given.

In Section 2 we discuss some variance functions. In Section 3 inference procedures for the parameters of the $NB(\mu, c, b)$ distribution are discussed. A score test statistic and a likelihood ratio statistic to test the fit of a two parameter negative binomial distribution $N(m, c)$ against its three parameter generalization $NB(\mu, c, b)$ are derived in Section 4. Some simulations and data analysis are performed in Section 5 to choose between these two models. Section 6 is devoted to extended quasi-likelihood analysis to choose an appropriate variance function. A Discussion follows in Section 7.

2 The Variance Functions

The Variance functions discussed in Section 1 are:

$$(a)\ Var(Y) = \mu(1 + c\mu), \qquad (b)\ Var(Y) = c_1\mu + c_2\mu^2,$$
$$(c)\ Var(Y) = c_3\mu^2 \text{ and} \qquad (d)\ Var(Y) = \mu(1 + c\mu^b).$$

Note that Variance function (a) is that of the negative binomial distribution and Variance function (d) is that of the three parameter generalized negative binomial distribution. The negative binomial variance function (a) is the most popular (Collings, 1981 and Fisher, 1941) which is a special case of the variance function (d). Variance function (b) is also a special case of the variance function (d) with $c_1 = 1$, $c_2 = c$, $b = 2$.

3 Inference for the $NB(\mu, c, b)$ Model Parameters

Suppose $Y_1, \ldots, Y_n$ is a random sample of size n from the three parameter negative binomial model $NB(\mu, c, b)$. Now, the kernel of the log-likelihood, after simplification, can be written as

$$l = n\bar{y}\log(\mu) + \sum_{i=1}^{n}\sum_{j=1}^{y_i}\log(1+(j-1)c\mu^{b-1}) - n(\bar{y}+c^{-1}\mu^{1-b})\log(1+c\mu^b).$$

Then, the maximum likelihood estimates of the parameters μ, c and b can be obtained by solving the following likelihood score equations

$$\begin{aligned}\frac{\partial\ell}{\partial\mu} &= \frac{n\bar{y}}{\mu} + c(b-1)\mu^{b-1}\sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{(j-1)}{\mu\left(1+(j-1)\,c\mu^{b-1}\right)} - \frac{n\bar{y}cb\mu^b}{\mu\left(1+c\mu^b\right)}\\ &\quad + \frac{n(b-1)\mu^{-b}\ln\left(1+c\mu^b\right)}{c} - \frac{nb}{\left(1+c\mu^b\right)} = 0, \qquad (3.1)\end{aligned}$$

$$\frac{\partial\ell}{\partial c} = \sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{(j-1)\,\mu^{b-1}}{1+(j-1)\,c\mu^{b-1}} - \frac{n\mu^b(c\bar{y}+\mu^{1-b})}{c(1+c\mu^b)} + \frac{n\mu^{-b+1}\ln\left(1+c\mu^b\right)}{c^2} = 0. \qquad (3.2)$$

and

$$\begin{aligned}\frac{\partial\ell}{\partial b} &= \sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{(j-1)\,c\mu^{b-1}\ln(\mu)}{1+(j-1)\,c\mu^{b-1}} - \frac{c\mu^b\ln(\mu)\sum_{i=1}^{n}y_i}{1+c\mu^b} + \frac{n\mu^{-b+1}\ln(\mu)\ln\left(1+c\mu^b\right)}{c}\\ &\quad - \frac{n\mu^{-b+1}\mu^b\ln(\mu)}{1+c\mu^b} = 0 \qquad (3.3)\end{aligned}$$

which, after taking $c\log(\mu)$ common, can be further written as

$$\begin{aligned}\frac{\partial\ell}{\partial b} &= c\ \log(\mu)\left[\sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{(j-1)\,\mu^{b-1}}{1+(j-1)\,c\mu^{b-1}} - \frac{n\mu^b(c\bar{y}+\mu^{1-b})}{c(1+c\mu^b)} + \frac{n\mu^{-b+1}\ln\left(1+c\mu^b\right)}{c^2}\right]\\ &= c\ \log(\mu)\frac{\partial\ell}{\partial c} \qquad (3.4)\end{aligned}$$

However, there is a problem in solving these equations. It is clear that $\partial\ell/\partial b = 0$ and $\partial\ell/\partial c = 0$ produce the same estimating equation. Thus all three parameters can not be estimated simultaneously. It seems that the parameters c and b can not be distinguished.

However, for fixed value of b the parameters μ and c can be estimated by solving $\frac{\partial \ell}{\partial \mu} = 0$ and $\frac{\partial \ell}{\partial c} = 0$ and the three parameter model $NB(\mu, c, b)$ can still be used to test goodness of fit of the two parameter negative binomial model $NB(\mu, c)$. For further insight see Section 5.

4 The Test Statistics

In this section we develop a score test and a likelihood ratio test for testing the negative binomial variance function (a) against the generalized negative binomial variance function (d), that is, for testing $H_0 : b = 1$ against $H_1 : b \neq 1$.

4.1 The Score Test

With the kernel of the log-likelihood l given in Section 3 define $\psi = \frac{\partial \ell}{\partial b} \mid_{b=1}$, $D = E[-\frac{\partial^2 \ell}{\partial b^2}]_{b=1}$, $A = (E[-\frac{\partial^2 \ell}{\partial b \partial c}]_{b=1},\ E[-\frac{\partial^2 \ell}{\partial b \partial \mu}]_{b=1})$ and

$$B \;=\; \begin{pmatrix} E[-\frac{\partial^2 \ell}{\partial \mu^2}]_{b=1} & E[-\frac{\partial^2 \ell}{\partial \mu \partial c}]_{b=1} \\ E[-\frac{\partial^2 \ell}{\partial c \partial \mu}]_{b=1} & E(-\frac{-\partial^2 \ell}{\partial c^2}]_{b=1} \end{pmatrix}.$$

Then, if, in ψ, A, B and D, the parameters μ and c of the negative binomial $NB(\mu, c)$ model are replaced by their maximum likelihood estimates $\hat{\mu}$ and $\hat{c}$ the score test statistic (Rao, 1947) for testing $H_0 : b = 1$ against $H_A : b \neq 1$, is $S = \frac{\psi^2}{Var(\psi)}$, where $Var(\psi) = D - AB^{-1}A^{'}$. The statistic S has asymptotically, as $n \to \infty$, a chi-square distribution with one degree of freedom. Detailed calculations and simplification in Appendix A show that $\hat{\mu} = \bar{y}$, $S = \hat{\psi}^2(\hat{a}_1 - \frac{\hat{a}_2^2}{\hat{a}_3})^{-1}$, where, $\hat{\psi} = \frac{n \log(\bar{y}) \log(1+\hat{c}\bar{y})}{\hat{c}} + \hat{c}\ \log(\bar{y})(\sum_{i=1}^{n} \sum_{j=1}^{y_i} \frac{(j-1)}{(1+(j-1)\hat{c})} - \hat{c} \sum_{i=1}^{n} \frac{(y_i+\hat{c}^{-1})}{1+\hat{c}\bar{y}})$, $\hat{a}_1 = 2n \log(\bar{y})\{\frac{\log(1+\hat{c}\bar{y})}{\hat{c}} - \frac{\bar{y}}{1+\hat{c}\bar{y}} - \hat{c}H\}$, $\hat{a}_2 = \frac{n \log(\bar{y})}{\hat{c}}\{\frac{\log(1+\hat{c}\bar{y})}{\hat{c}} - \frac{\bar{y}}{1+\hat{c}\bar{y}} - \hat{c}H\}$, $\hat{a}_3 = \frac{n}{\bar{y}(1+\hat{c}\bar{y})}$ and $H = E\{\sum_{i=1}^{n} \sum_{j=1}^{y_i} \frac{(j-1)}{(1+(j-1)\hat{c})^2}\}$. The maximum likelihood estimate $\hat{c}$ of c is obtained by solving the estimating equation $\sum_{i=1}^{n} \sum_{j=1}^{y_i} \frac{j-1}{c(1+(j-1)c)} + \frac{n \log(1+c\bar{y})}{c^2} - \frac{n(c\bar{y}+1)\bar{y}}{c(1+c\bar{y})} = 0$ or more simply by solving $\sum_{i=1}^{n} \sum_{j=1}^{y_i} \frac{1}{1+(j-1)c} - \frac{\log(1+c\bar{y})}{c} = 0$.

4.2 The likelihood Ratio Test

The likelihood ratio test statistic for testing $H_0 : b = 1$ against $H_1 : b \neq 1$ is given by $G^2 = 2(\hat{l}_1 - \hat{l}_0)$, where

$$\hat{l}_0 = n\bar{y}\log(\bar{y}) + \sum_{i=1}^{n} \sum_{j=1}^{y_i} \log(1 + (j-1)\hat{c}) - n(\bar{y} + \hat{c}^{-1}) \log(1 + \hat{c}\hat{y})$$

and

$$\hat{l}_1 = n\bar{y}\log(\tilde{\mu}) + \sum_{i=1}^{n} \sum_{j=1}^{y_i} \log(1 + (j-1)\tilde{c}\tilde{\mu}^{b-1}) - n(\bar{y} + \tilde{c}^{-1}\tilde{\mu}^{1-b}) \log(1 + \tilde{c}\tilde{\mu}^{b}),$$

where $\tilde{\mu}$ and $\tilde{c}$ are the maximum likelihood estimates of the $NB(\mu, c, b)$ model obtained by solving the estimating equations (3.1) and (3.2) for given b. Note that, as we discussed earlier, there are theoretical as well as computational difficulties in obtaining maximum likelihood estimates for all three parameters. See details of the likelihood ratio test in the context of data analysis in the next section.

5 Analysis Using the Score Test and the Likelihood Ratio Statistic

In this section we analyze five data sets using the score test statistic and the likelihood ratio statistic discussed in section 4. Conclusion from all five data sets were similar, so we present results of analysis of only two data sets. The details of the first two data sets are given here. Further details of the other three datasets analyzed can be obtained from the following references: Bohing, Dietz and Schlattmann(1999) for decayed, missing and filled tooth (DMFT) index data, Sellar, Stoll and Chavas (1985) for the data on number of recreational boating trips to Lake Somerville, Texas in 1980 and Leroux and Puterman (1992) for the data on the number of movements made by a fetal lamb in each of 240 consecutive five second intervals. Note that as discussed in section 4.2 the likelihood ratio test can be performed only for fixed values of the parameter b. Results of a simulation study is also reported here.

Embryonic Death Count Data: McCaughran and Arnold (1976), modelled data in which they referred to counts of embryonic deaths in a control group and two treatment groups. Here we consider data for one dose level which refers to the number of embryonic deaths in the treatment group related to dose level 2. The counts are summarized in Table 1. For this data set the value of the score test statistic $S = 0.00013$ and the value of likelihood ratio statistic is 1 for all values of b ($0 \leq b \leq 2$) given in Table 2. In Table 2, for different values of the parameter, we present values of $-\hat{l}$, estimates $\tilde{\mu}$ and $\tilde{c}$ with their standard errors in parenthesis; and an estimate of the variance function with its standard error in parenthesis. We have chosen values of b with these limits as previous studies have found that for count data these limits are reasonable. The value of $b = 0$ represents a model in which the variance is proportional to the mean. Armitage (1957) and Finney (1976) found by study of examples that $1 \leq b \leq 2$. Other values of b within $1 \leq b \leq 2$ were chosen to see whether property of any of these quantities remain the same.

European Red Mites Data: Bliss and Fisher (1953) presented data which concerned counts of the number of European red mites on apple leaves from Garman (1951) of the Connecticut Agricultural Experiment Station. This dataset was also analyzed by Clark and Perry (1949). There were six Macintosh apple trees which were given the same spray treatment in a single orchard. Garman (1951) selected 25 leaves at random from each of the six trees and counted the number of adult female mites on each leaf. The data in the form of frequencies of mites on the 150 leaves, are given in Table 3. For this dataset the value of the score test statistic is $S = 0.0104$ and the value of likelihood ratio is 1 for all values of b ($0 \le b \le 2$) given in Table 4. In Table 4 we give, for different values of b, values of $-\hat{l}$, estimates of $\tilde{\mu}$ and $\tilde{c}$ with their standard error in parenthesis and an estimate of the variance function with its standard error in parenthesis.

Results in Table 2 and Table 4 show that values of $-l$, $\hat{\mu}$, its standard error, the estimate of the variance function and its standard error are the same for all values of b. The only thing that changes is the estimate of c which does not show any impact on anything else. The reason is that in the variance function $Var(Y) = \mu(1 + c\mu^b)$ of the generalized negative binomial model, the parameters c and b are confounded, meaning that both these parameters are over-dispersion parameters. For the same amount of over-dispersion in the data, if the value of one of these parameters changes, then the value of the other parameter also changes, keeping the overall variance the same.

We also did an extensive simulation study with the score test and we found that, as in the analysis of the five data sets above, the value of S is very small indicating that the three parameter generalization of the negative binomial distribution dose not improve over its two parameter counterpart namely, the negative binomial distribution. Also, as we indicated earlier the three parameter generalized negative binomial model has theoretical as well as computational problems.

6 Semi-parametric Analysis Using the Extended Quasi-likelihood

6.1 The Extended Quasi-likelihood

The score test statistic and the likelihood ratio statistic analyze only tests $H_0 : b = 1$ against $H_1 : b \neq 1$ using the parametric model (1.2). As such we can not test other variance functions using this model. However, for comparing certain variance functions we can use the extended quasi likelihood. The extended quasi-likelihood was proposed by Nelder and Pregibon (1987) and Godambe and Thompson (1989), as an extension of the quasi-likelihood to incorporate the extra variation. For any variance function $V(Y)$ and data $y_1, y_2, \dots, y_n$, the extended quasi-likelihood is defined as

$$Q^{\dagger}(y; \mu, \phi) = -\sum_{i=1}^{n}\{\frac{1}{2}\log\{2\pi V(y_i)\} - \frac{1}{2}D(y_i; \mu)\}, \qquad (6.1)$$

where $D(y_i; \mu)$ is the deviance and is given by

$$D(y_i; \mu) = -2\sum_{i=1}^{n}\{\int_{y_i}^{\mu}\frac{y_i - u}{V(u)}du\}. \qquad (6.2)$$

The extended quasi-likelihoods for the variance functions (a) to (c) are given in Table 5. The extended quasi-likelihood for the variance function (d), that is, the variance function of the $NB(\mu, c, b)$ distribution, does not have a closed form. So, this is omitted from further consideration.

6.2 Evaluation of the Variance Functions Using Data Analysis

Using data analysis and simulations in section 5 we reported that a test of goodness of fit of the negative binomial distribution $NB(\mu, c)$ against its three parameter generalization $NB(\mu, c, b)$ either by a score test or a likelihood ratio analysis produces very insignificant values of the test statistics. This indicates that in practical data analysis the three parameter generalization does not improve in fit over the two parameter negative binomial model $NB(\mu, c)$.

Here we compare the variance functions (a) to (c) using the extended quasi-likelihood through data analysis. As in Section 5 we analyzed five published data sets. However, conclusion was found to be the same for all data sets, so here we present an analysis for two data sets. Through these data analysis we find that only one of the parameters of the variance function $c_1\mu + c_2\mu^2$ is estimable using the extended quasi-likelihood given in Table 5. The theoretical reason for this is unknown to us and will be investigated in a future study. So, we omit this variance function from further consideration.

Results of the extended quasi-likelihood analysis for the data sets in Table 1 and Table 3 are given in Table 6 and Table 7, respectively, for the remaining variance functions $\mu + c\mu^2$ and $c_3\mu^2$. In both tables we give values of $-\hat{q}$, where $\hat{q}$ is the estimated extended quasi log-likelihood, with estimates of parameters μ, c and c_3 along with their standard errors in parenthesis.

A common theme that is seen from the semi-parametric analysis of these real data sets (including those that are not shown here) is that the extended quasi likelihood is larger for the variance function $c_3\mu^2$ than that for the negative binomial variance function. To see whether this is a general phenomenon we did some further simulations.

6.3 Simulations

We now conduct a simulation study to compare the negative binomial likelihood with the extended quasi likelihood values for the variance functions $\mu + c\mu^2$ and $c_3\mu^2$. Efficiencies of the estimates of μ and those of the two variance functions are also evaluated. Simulations are conducted by taking 10000 repeated samples of sizes n= 20, 30 and 50 from the Negative binomial (μ, c) distribution. We use all combinations of $n = 20, 30, 50$, $\mu = 2, 5, 10, 20$ and $c = 0.1, 0.2, 0.4, 0.6$ in our simulations. Results for $c = 0.4$ and 0.6 are similar, so, to save space, these for $c = 0.6$ have been omitted. In Table 8 we present the following: values of minus the negative-binomial log-likelihood $(-\hat{l})$, values of minus the extended quasi likelihood $(-\hat{q}_1)$ for variance function $v_1 = \mu + c\mu^2$, values of minus the extended quasi likelihood $(-\hat{q}_2)$ for variance function $v_2 = c_3\mu^2$, the relative efficiency of the estimates of μ under the variance function v_1, denoted by RE_1, the relative efficiency of the estimates of μ under the variance function v_2, denoted by RE_2, and the relative efficiency of the estimates of v_1 and v_2 denoted by $RE(v_1)$ and $RE(v_2)$ for all combinations of $n = 20, 30, 50$, $\mu = 2, 5, 10, 20$ and $c = 0.1, 0.2$ and 0.4. The relative efficiency of an estimator of a parameter θ is calculated as $MSE(\hat{\theta})/MSE(\tilde{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimate of θ under the negative binomial model and $\tilde{\theta}$ is the extended quasi-likelihood estimate of θ under certain variance function.

Results in Table 8 indicate that the extended quasi-likelihood estimate of the parameter μ using either of the two variance functions has almost full efficiency in every situation studied. Also, the estimate of the variance function v_1 has almost full efficiency. The estimate of the variance function v_2 has efficiency below 1 only when μ and c are small, for example for $\mu = 2, 5$, $c = 0.1$ and $\mu = 2$ and $c = 0.2$ efficiency falls far below 1. Otherwise, in general, efficiency of v_2 is larger than 1.

To check whether the behavior of the estimates of μ and those of the variance functions remain the same when data arise from other over-dispersed count data models, we extended our simulation study. For this we generated data from the log-normal (μ^*, σ^2) mixture of Poisson distribution, where $\mu^* = \log(\mu) - .5\log(1+c)$ *and* $\sigma^2 = \log(1+c)$. The behavior of the estimate of μ and the two variance functions under consideration remains similar (see, Table 9) irrespective of whether we generate data from the gamma mixture of a Poisson (negative binomial) or from the log-normal mixture of a Poisson.

7 Discussion

We have developed likelihood and extended quasi-likelihood methodologies to choose an appropriate variance function for the semi-parametric analysis (that is, without full distributional assumption) of count data. Data analysis and simulations indicate that the three parameter generalized negative binomial distribution $NB(\mu, c, b)$ does not improve in fit to count data over the simpler negative binomial $NB(\mu, c)$ distribution. So, for semi-parametric analysis we prefer the negative binomial variance function over the variance function of the $NB(\mu, c, b)$ distribution. Moreover, the variance function of the $NB(\mu, c, b)$ distribution does not have a closed form and is difficult to calculate numerically. Also, as in the likelihood analysis based on the $NB(\mu, c, b)$ distribution, only one of the parameters c and b may be estimable. So, the variance function of the $NB(\mu, c, b)$ distribution was not considered in the semi-parametric analysis.

Further, through extensive data analysis we found that only one of the parameters of the variance function $c_1\mu + c_2\mu^2$ is estimable using the extended quasi-likelihood. Moreover, Rodbard and Hutt (1974) reported from their experience that for immunoradiometric assays the linear term in the variance function $c_1\mu + c_2\mu^2$ can be omitted, and that the variance function $c_3\mu^2$ is a very satisfactory representation of the variance. So, the variance function $c_1\mu + c_2\mu^2$ was also not considered for extended quasi-likelihood analysis.

Further data analysis and simulations using extended quasi-likelihood indicate that the negative binomial variance function $v_1 = \mu + c\mu^2$ has almost full efficiency. The estimate of the variance function $v_2 = c_3\mu^2$ has efficiency below 1 only when μ and c are small, for example, for $\mu = 2$, c=0.1; $\mu = 5$, c=0.1 and $\mu = 2$, c=0.2 efficiency falls far below 1. Otherwise, in general, the efficiency of v_2 is larger than 1.

Acknowledgements

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] Anscombe, F.J. (1949). The statistical analysis of insect counts based on negative binomial distribution. *Biometrics*, **5**, 165-173.

[2] Armitage, P. (1957). Studies in Variability of pock counts. *Journal of Hygiene. Camb.*, **55**, 564-581.

[3] Barnwal, R.K. and Paul, S.R. (1988). Analysis of one-way layout of count data with negative binomial variation. *Biometrika*, **75**, 215-222.

[4] Bartlett, M.S. (1936). Some notes on insecticide tests in the laboratory and in the field *Journal of Royal Statistical Society*, Suppl. **3**, 185-194.

[5] Böhning, D., Dietz, E., Schlattmann, P., Mendonca, L., Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of Royal Statistical Society A* , **162**, 195-209.

[6] Bliss, C.I. and Fisher, R.A. (1953). Fitting the negative binomial distribution to biological data. *Biometrics*, **9**,176-200.

[7] Bliss, C.I. and Owen, A.R.G. (1958). Negative binomial distribution with common k. *Biometrika*, **45**,37-58.

[8] Byrd, R.H., Lu, P., Nocedal, J. and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput*, **16**, 11-29.

[9] Cameron, A.C. and Trivedi, P.K. (1986). Econometrics models based on count data. *Journal of Applied Statistics.*, **1**, 29-53.

[10] Collings, B.J. (1981). The negative binomial distribution an alternative to the Poisson. unpublished PhD thesis, University of North Carolina at Chapel Hill, USA.

[11] Collings, B.J. and Margolin, B.H. (1985). Testing goodness of fit for Poisson assumption when observations are not identically distributed. *J. Am. Statist. Assoc.*, **80**, 411-418.

[12] Clark, S.J and Perry, J.N. (1989). Estimation of negative binomial parameter k by maximum quasi-likelihood. *Biometrics*, **45**, 309-316.

[13] Fisher, R.A. (1941). The negative binomial distribution. *Annals of Eugenics*, **11**, 182-187.

[14] Finney, D.J. (1976). Radioligand assay. *Biometrics*, **32**, 721-740.

[15] Gaffeo, E., Guilmi, C.D., Gallegati, M. and Russo, A. (2008). On the mean/variance relationship of the firm size distribution: evidence and some theory. *Discussion Paper No. 5* , Department Di Economia, Universitá Degli Studi Di Trento, Italy.

[16] Garman, P. (1951). Original Data on European Redmite on apple leaves. Report. Connecticut.

[17] Godambe, V.P. and Thompson, M.E. (1989). An extension of quasi-likelihood estimation. *Journal of Statistical Planning and Inference* , **22**, 137-152.

[18] Johnson, N.L. and Kotz, S. *Discrete Distributions*. Wiley, New York, 1969.

[19] Kendal, W.S. (2004). Taylor's ecological power law as a consequence of scale invariant exponential dispersion models. *Ecological Complexity* , **1**, 193-209.

[20] Kilpatrick, A.M. and Ives, A.R. (2003). Species interactions can explain Taylor's power law for ecological time series. *Nature*, **422**, 65-68.

[21] Lee, Y. and Nelder, J. A. (2001). Hierarchical generalized linear models: a synthesis of generalized linear models, random effect models and structured dispersions. *Biometrika*, **88**, 987-1006.

[22] Leroux, B.G. and Puterman, M.L. (1992). Maximum-penalized-likelihood estimation for independent and Markov dependent mixture models. *Biometrics*, **48**, 545-558.

[23] McCaughran, D.A. and Arnold, D.W. (1976). Statistical models for numbers of implantation sites and embryonic deaths in mice. *Toxicology and Applied Pharmacology*, **38**, 325-333.

[24] Moran, P.A.P. (1970). On asymptotically optimal tests of composite hypotheses. *Biometrika*, **57**, 45-55.

[25] Nelder, J.A. and Pregibon, D. (1987). An extended quasi likelihood function. *Biometrika*, **74**, 221-232.

[26] Paul, S.R. and Plackett, R.L. (1978). Inference sensitivity for Poisson mixtures. *Biometrika*, **65**, 591-602.

[27] Paul, S.R. and Banerjee, T. (1998). Analysis of two-way layout of count data involving multiple counts in each cell. *Journal of the American Statistical Association*, **93**, 1419-1429.

[28] Piegorsch, W.W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics*, **46**, 863-867.

[29] Rao, C.R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, **44**, 50-57.

[30] Rodbard, D. and Hutt, D.M. (1974). Statistical analysis of Radioimmunoassay and immunoradiometric (labelled antibody) assays: a generalized weighted, iterative, least squares method for logistic curve fitting. *Radioimmunoassay and related procedures in Medicine*, **1**, International Atomic Energy Agency, Vienna, 165-192.

[31] Sellar, C., Stoll, J.R. and Chavas, J.P. (1985). Validation of empirical measures of welfare change: a comparison of nonmarket techniques. *Land Economics*, **61**, 156-175.

[32] Williams, D.A. (1982). Extra binomial variation in logistic models. *Applied Statistics*, **31**, 144-148.

A Expression for Score Statistic

Suppose $Y_1, \ldots, Y_n$ is a random sample of size n from the three parameter negative binomial model $NB(\mu, c, b)$ which has probability function

$$Pr(Y=y) = \frac{\Gamma(y + c^{-1}\mu^{1-b})}{y!\Gamma(c^{-1}\mu^{1-b})}\Big(\frac{1}{1+c\mu^b}\Big)^{c^{-1}\mu^{1-b}}\Big(\frac{c\mu^b}{1+c\mu^b}\Big)^{y}. \tag{A.1}$$

Now, for any non-negative integer y and any $x \neq 0, -1, -2, \ldots$,

$$\Gamma(y+x)/\Gamma(x) = \begin{cases} \Pi_{j=1}^{y}\{x + (j-1)\}, & \text{if } y > 0 \\ 1, \text{ if } y = 0. \end{cases} \tag{A.2}$$

Using this and other simplifications, we obtain

$$\begin{aligned} \log\Big(\frac{\Gamma(y + c^{-1}\mu^{1-b})}{\Gamma(c^{-1}\mu^{1-b})}\Big) &= \log\Big[\prod_{j=1}^{y}\{c^{-1}\mu^{1-b} + (j-1)\}\Big] \\ &= \sum_{j=1}^{y}\log\big(1 + (j-1)c\mu^{b-1}\big) - y\ \log(c\mu^{b-1}). \end{aligned} \tag{A.3}$$

Then, using the results in equation (A.3) and after simplification, the kernel of the log-likelihood can be written as

$$\begin{aligned} l &= \sum_{i=1}^{n}\log\Big(\frac{\Gamma(y + c^{-1}\mu^{1-b})}{\Gamma(c^{-1}\mu^{1-b})}\Big) - \log(1+c\mu^b)\sum_{i=1}^{n}\Big\{(c^{-1}\mu^{1-b}) + y_i\Big\} + \log(c\mu^b)\sum_{i=1}^{n}y_i \\ &= \log(\mu)\sum_{i=1}^{n}y_i + \sum_{i=1}^{n}\sum_{j=1}^{y_i}\log(1 + (j-1)c\mu^{b-1}) - \log(1+c\mu^b)\sum_{i=1}^{n}(y_i + c^{-1}\mu^{1-b}). \end{aligned} \tag{A.4}$$

Following Barnwal and Paul (1988) we define $\phi = b$ and $\theta = (\mu, c)$. We further define $\psi(\theta) = \frac{\partial \ell}{\partial b}|_{b=1}$, $D = E(-\frac{\partial^2 \ell}{\partial b^2})_{b=1}$, $A = [E(-\frac{\partial^2 \ell}{\partial b \partial c})_{b=1}, E(-\frac{\partial^2 \ell}{\partial b \partial \mu})_{b=1}]$ and

$$B = \begin{pmatrix} E(-\frac{\partial^2 \ell}{\partial \mu^2})_{b=1} & E(-\frac{\partial^2 \ell}{\partial \mu \partial c})_{b=1} \\ E(-\frac{\partial^2 \ell}{\partial c \partial \mu})_{b=1} & E(-\frac{-\partial^2 \ell}{\partial c^2})_{b=1} \end{pmatrix}.$$

Then the score test statistic (Rao, 1947) for testing $H_0 : b = 1$ against $H_A : b \neq 1$, is

$$S = \frac{\psi(\theta)^2}{Var(\psi(\theta))}, \tag{A.5}$$

where $Var(\psi(\theta)) = D - AB^{-1}A'$, which has asymptotically, as $n \to \infty$, a chi-square distribution with one degree of freedom. Note that θ is an unknown nuisance parameter. If θ is replaced by $\hat{\theta}$, the maximum likelihood estimator of θ under H_0 in ψ, A, B and D, then the distribution of $S = \frac{\psi(\theta)^2}{D - AB^{-1}A}$ is chi-square, asymptotically, as $n \to \infty$, with one degree of freedom (Moran, 1970). Now, using the log-likelihood (l) the required quantities for the score statistic S are obtained in what follows.

$$\psi = \Big|\frac{\partial \ell}{\partial b}\Big|_{b=1} = \frac{\log(\mu)}{c}\Big\{n \log(1 + c\mu) + \sum_{i=1}^{n}\sum_{j=1}^{y_i} \frac{c^2(j-1)}{(1 + (j-1)c)} - \sum_{i=1}^{n} \frac{c^2\mu(x_i + c^{-1})}{1 + c\mu}\Big\}. \tag{A.6}$$

$$\begin{aligned} E\Big[-\frac{\partial^2 \ell}{\partial b^2}\Big]_{b=1} &= E\Big[\frac{n\ \log(1+c\mu)}{c} - \frac{\mu}{(1+c\mu)^2}\Big\{n(1+2c\mu) - c\sum_{i=1}^{n} y_i\Big\} - \sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{c(j-1)}{(1+(j-1)c)^2}\Big]\log(\mu)^2 \\ &= \frac{\log(\mu)^2}{c}\Big\{n\log(1+c\mu) - \frac{nc\mu}{(1+c\mu)} - c^2 \sum_{i=1}^{n}\sum_{y=0}^{\infty}\sum_{j=1}^{y}\frac{(j-1)}{(1+(j-1)c)^2}Pr(y)\Big\} \\ &= a_1(say), \end{aligned} \tag{A.7}$$

where $Pr(y)$ is the negative binomial probability given by equation (1.1).

$$\begin{aligned} E\Big[-\frac{\partial^2 \ell}{\partial b \partial \mu}\Big]_{b=1} &= E\Big[\frac{n}{(1+c\mu)} + \frac{c\{1+\log(\mu)\}\sum_{i=1}^{n} y_i}{(1+c\mu)} - \frac{n\log(1+c\mu)}{c\mu} - \frac{\mu c^2\log(\mu)(\sum_{i=1}^{n} y_i + nc^{-1})}{(1+c\mu)^2} \\ &\qquad - \frac{c}{\mu}\Big\{\sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{(j-1)}{(1+(j-1)c)}\Big\}\Big] \\ &= \frac{n}{(1+c\mu)} + \frac{n\mu c(1+\log(\mu))}{1+c\mu} - \frac{n\log(1+c\mu)}{c\mu} - \frac{n\mu c^2\log(\mu)(\mu+c^{-1})}{(1+c\mu)^2} \\ &\qquad - \frac{c}{\mu}E\Big\{\sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{(j-1)}{(1+(j-1)c)}\Big\}. \end{aligned} \tag{A.8}$$

Now,

$$E\Big[\frac{\partial \ell}{\partial b}\Big]_{b=1} = E\Big[\sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{(j-1)}{(1+(j-1)c)} - \frac{n\mu}{c} + \frac{n\log(1+c\mu)}{c^2}\Big] = 0, \tag{A.9}$$

so that

$$E\left[\sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{(j-1)}{(1+(j-1)c)}\right] = \frac{n\mu}{c}-\frac{n\log(1+c\mu)}{c^2}. \tag{A.10}$$

Substituting equation (A.10) in equation (A.8) we obtain

$$\begin{aligned} E\left[-\frac{\partial^2\ell}{\partial b\partial\mu}\right]_{b=1} &= \frac{n}{(1+c\mu)}+\frac{n\mu c(1+\log(\mu))}{1+c\mu}-\frac{n\log(1+c\mu)}{c\mu}-\frac{n\mu c^2\log(\mu)(\mu+c^{-1})}{(1+c\mu)^2}- \\ &-n+\frac{n\log(1+c\mu)}{\mu c} \\ &= 0. \end{aligned} \tag{A.11}$$

$$\begin{aligned} E\left[-\frac{\partial^2\ell}{\partial b\partial c}\right]_{b=1} &= E\Bigg[\frac{n\log(\mu)\log(1+c\mu)}{c^2}-\frac{n\mu\log(\mu)}{c(1+c\mu)}-\log(\mu)\left\{\sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{(j-1)}{(1+(j-1)c)^2}\right\} \\ &-\frac{n\mu^2\log(\mu)}{(1+c\mu)^2}-\frac{c\mu^2\log(\mu)\sum_{i=1}^{n}y_i}{(1+c\mu)^2}+\frac{\mu\log(\mu)\sum_{i=1}^{n}y_i}{(1+c\mu)}\Bigg] \\ &= \frac{n\log(\mu)\log(1+c\mu)}{c^2}-\frac{n\mu\log(\mu)}{c(1+c\mu)}-\log(\mu)\sum_{i=1}^{n}\sum_{y=0}^{\infty}\sum_{j=1}^{y}\frac{(j-1)}{(1+(j-1)c)^2}Pr(y) \\ &= a_2(say). \end{aligned} \tag{A.12}$$

Therefore, we can write $A=(a_2,\ 0)$. Next we obtain

$$E\left[-\frac{\partial^2\ell}{\partial\mu^2}\right]_{b=1} = E\left[\frac{2c\mu\sum_{i=1}^{n}y_i+\sum_{i=1}^{n}y_i-nc\mu^2}{(1+c\mu)^2\mu^2}\right] = \frac{n}{\mu(1+c\mu)} = a_3(say), \tag{A.13}$$

$$E\left[-\frac{\partial^2\ell}{\partial\mu\partial c}\right]_{b=1} = E\left[\frac{-\sum_{i=1}^{n}y_i+n\mu}{(1+c\mu)^2}\right] = 0 \tag{A.14}$$

and $E(-\frac{\partial^2\ell}{\partial c^2})|_{b=1}=a_4(say)$. Then

$$B^{-1}=\begin{pmatrix} 1/a_3 & 0 \\ 0 & 1/a_4 \end{pmatrix} \tag{A.15}$$

and $D-AB^{-1}A'=a_1-a_2^2/a_3$.

Thus, the score test statistic for testing $H_0: b=1$ against $H_1: b\neq 1$, is given by

$$S=\psi(\hat{\theta})^2\Big(\hat{a}_1-\frac{\hat{a}_2^2}{\hat{a}_3}\Big)^{-1},$$

where, $\psi(\hat{\theta})=\frac{n\log(\hat{\mu})\log(1+\hat{c}\hat{\mu})}{\hat{c}}+\hat{c}\ \log(\hat{\mu})\sum_{i=1}^{n}\sum_{j=1}^{y_i}\frac{(j-1)}{(1+(j-1)\hat{c})}-\hat{c}\ \hat{\mu}\ \log(\hat{\mu})\sum_{i=1}^{n}\frac{(y_i+\hat{c}^{-1})}{1+\hat{c}\hat{\mu}}$, $\hat{a}_1=2n\log(\hat{\mu})\{\frac{\log(1+\hat{c}\hat{\mu})}{\hat{c}}-\frac{\hat{\mu}}{1+\hat{c}\hat{\mu}}-\hat{c}H\}$, $\hat{a}_2=\frac{n\log(\hat{\mu})}{\hat{c}}\{\frac{\log(1+\hat{c}\hat{\mu})}{\hat{c}}-\frac{\hat{\mu}}{1+\hat{c}\hat{\mu}}-\hat{c}H\}$, $\hat{a}_3=\frac{n}{\hat{\mu}(1+\hat{c}\hat{\mu})}$

where, $H = E\{\sum_{i=1}^{n} \sum_{j=1}^{y_i} \frac{(j-1)}{(1+(j-1)\hat{c})^2}\}$ and $\hat{\mu}$ and $\hat{c}$ are the maximum likelihood estimates under the null hypothesis. Now, under the null hypothesis $b = 1$, the kernel of the log-likelihood, that is, the kernel of log-likelihood of the $NB(\mu, c)$ model, is

$$l_0(\mu, c) = \sum_{i=1}^{n} y_i \log(\mu) + \sum_{i=1}^{n} \sum_{j=1}^{y_i} \log(1 + (j-1)c) - \sum_{i=1}^{n} (y_i + c^{-1}) \log(1 + c\mu). \quad \text{(A.16)}$$

Maximum likelihood estimates of μ and c are obtained by solving

$$\frac{\partial \ell_0}{\partial \mu} = \sum_{i=1}^{n} \left[\frac{y_i}{\mu} - \frac{1 + cy_i}{1 + c\mu} \right] = 0 \quad \text{(A.17)}$$

and

$$\frac{\partial \ell_0}{\partial c} = \sum_{i=1}^{n} \left[\frac{\log(1 + c\mu)}{c^2} - \frac{(y_i + c^{-1})\mu}{1 + c\mu} + \sum_{j=1}^{y_i} \frac{(j-1)}{(1 + (j-1)c)} \right] = 0. \quad \text{(A.18)}$$

Solution to (A.17) gives $\hat{\mu} = \bar{y}$. Simplifying (A.18) and putting $\hat{\mu} = \bar{y}$ the estimating equation for obtaining $\hat{c}$ is $\frac{\log(1+c\bar{y})}{c} - \sum_{i=1}^{n} \sum_{j=1}^{y_i} \frac{1}{(1+(j-1)c)} = 0$. This equation can be solved numerically using a subroutine, such as the FORTRAN subroutine ZBREN. Alternatively one can do numerical maximization of $l_0(\bar{y}, c)$ following Piegorsch (1990). Here, we used R function OPTIM to maximize $l_0(\bar{y}, c)$ for obtaining $\hat{c}$. The R function maximizes a function by the quasi-Newton method which uses function values and gradient (see Byrd, Lu, Nocedal and Zhu, 1995).

Table 1: Frequency distribution of counts of embryonic death

Number of deaths	0	1	2	3	4
frequency in treatment with dose level 2	4	2	3	0	1

Table 2: Values of estimate of minus the log-likelihood $(-l)$; estimates of μ and c and their standard errors in parenthesis; and estimate of the variance function with its standard error in parenthesis; for different values of b

b	$-\hat{l}$	$\tilde{\tilde{\mu}}$SE($\tilde{\tilde{\mu}}$)	$\tilde{\tilde{c}}$SE($\tilde{\tilde{c}}$)	$\tilde{\tilde{\mu}}(1+\tilde{\tilde{c}}\tilde{\tilde{\mu}}^b)$	SE $(\tilde{\tilde{\mu}}(1+\tilde{\tilde{c}}\tilde{\tilde{\mu}}^b))$
0.0	14.85	1.20(.41)	.39(.76)	1.67	1.34
0.4	14.85	1.20(.41)	.36(.70)	1.67	1.34
0.8	14.85	1.20(.41)	.34(.65)	1.67	1.34
1.0	14.85	1.20(.41)	.33(.62)	1.67	1.34
1.4	14.85	1.20(.41)	.30(.58)	1.67	1.34
1.8	14.85	1.20(.41)	.28(.55)	1.67	1.34
2.0	14.85	1.20(.41)	.27(.53)	1.67	1.34

Table 3: Frequency distribution of red mites on apple leaves.

Number of mites per leaf	0	1	2	3	4	5	6	7	8+
Number of leaves observed	70	38	17	10	9	3	2	1	0

Table 4: Values of estimate of minus the log-likelihood $(-l)$; estimates of μ and c and their standard errors in parenthesis; and estimate of the variance function with its standard error in parenthesis; for different values b

b	$-\hat{l}$	$\tilde{\tilde{\mu}}$SE($\tilde{\tilde{\mu}}$)	$\tilde{\tilde{c}}$SE($\tilde{\tilde{c}}$)	$\tilde{\tilde{\mu}}(1+\tilde{\tilde{c}}\tilde{\tilde{\mu}}^b)$	SE $(\tilde{\tilde{\mu}}(1+\tilde{\tilde{c}}\tilde{\tilde{\mu}}^b))$
0.0	222.44	1.15(.13)	1.12(.33)	2.43	.29
0.4	222.44	1.15(.13)	1.06(.29)	2.43	.29
0.8	222.44	1.15(.13)	1.00(.27)	2.43	.29
1.0	222.44	1.15(.13)	.98(.26)	2.43	.29
1.4	222.44	1.15(.13)	.92(.25)	2.43	.29
1.8	222.44	1.15(.13)	.87(.25)	2.43	.29
2.0	222.44	1.15(.13)	.85(.25)	2.43	.29

Table 5: Variance function and the extended quasi likelihood for the variance functions (a) to (c)

Variance Function	Extended Quasi Likelihood
$\mu + c\mu^2$	$\sum_i^n \{-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\frac{(y_i+1/6)(1+cy_i)^2(1+c/6)}{(1+cy_i+c/6)})$ $+(y_i + c^{-1})\log(\frac{1+cy_i}{1+c\mu}) - y_i\log(\frac{y_i}{\mu})\}$
$c_1\mu + c_2\mu^2$	$-\frac{n}{2}\log(2\pi) - \frac{\Sigma\log(y_i)}{2} + \frac{\log(\mu)}{c_1}\Sigma y_i + (\frac{1}{c_2} - \frac{1}{2})\Sigma\log(c_1 + c_2 y_i)$ $-\log(c_1 + c_2\mu)(\frac{\Sigma y_i}{c_1} + \frac{n}{c_2}) + \frac{\Sigma y_i\log(c_2 + \frac{c_1}{y_i})}{c_1}$
$c_3\mu^2$	$-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(c_3) - \Sigma\log(y_i) - \frac{\sum y_i}{c_3\mu} + \frac{1}{c_3}\Sigma\log(y_i/\mu) + \frac{n}{c_3}$

Table 6: Values of estimate of minus the extended quasi log-likelihood ($-q$), estimates of μ, c, c_3 and variance function (VF) along with their standard errors in parenthesis for different variance functions for the embryo data set

VF	$-\hat{q}$	$\hat{\mu}$	$\hat{c}$	$\hat{c_3}$	$\hat{b}$	$\widehat{VF}$
$\mu + c\mu^2$	14.63	1.20(.41)	.33(.64)	-	-	1.68(1.18)
$c_3\mu^2$	10.99	1.20(.40)	-	1.10(.49)	-	1.59(1.27)

Table 7: Values of estimate of minus the extended quasi log-likelihood ($-q$), estimates of μ, c, c_3 and variance function (VF) along with their standard errors in parenthesis for different variance functions for the red mite data set

VF	$-\hat{q}$	$\hat{\mu}$	$\hat{c}$	$\hat{c_3}$	$\hat{b}$	$\widehat{VF}$
$\mu + c\mu^2$	219.47	1.15(.13)	1.00(.27)	-	-	2.47(.55)
$c\mu^2$	154.85	1.15(.11)	-	1.34(.15)	-	1.76(.39)

Table 8: Values of estimates of negative-binomial log-likelihood (l), extended quasi likelihood for variance function $v_1 = \mu + c\mu^2$ (q_1), extended quasi likelihood for variance function $v_2 = c_3\mu^2$ (q_2) and relative efficiency of estimates of μ and v_1, v_2 with respect to maximum likelihood estimates for all combinations of n=20, 30, 50, μ=2, 5, 10, 20 and c=.1, .2, 4; data from $NB(\mu, c)$ distribution

c	n	μ	$-\hat{l}$	$-\hat{q}_1$	$-\hat{q}_2$	RE_1	RE_2	$RE(v_1)$	$RE(v_2)$
0.1	20	2	34.51	34.55	31.09	.99	.99	.98	.17
		5	46.66	46.65	46.01	.99	.99	.99	.50
		10	56.45	56.45	56.18	1.00	1.00	.99	1.93
		20	67.62	67.62	67.37	.99	.99	.99	6.8
	30	2	51.86	51.92	46.80	.99	.99	.98	.14
		5	73.01	70.01	69.51	1.00	1.00	.99	.47
		10	85.15	85.15	84.80	1.00	1.00	.99	2.03
		20	101.88	101.87	101.53	1.00	1.00	.99	7.72
	50	2	87.15	87.31	78.90	.99	.99	.98	.12
		5	117.83	117.82	116.97	1.00	1.00	.99	.47
		10	142.92	142.91	142.41	1.00	1.00	.99	2.13
		20	170.58	170.58	170.06	.99	.99	.99	8.38
0.2	20	2	35.14	35.20	31.30	1.00	1.00	.98	.43
		5	48.38	48.38	47.53	.99	.99	.99	1.27
		10	59.77	59.76	59.27	1.00	1.00	.99	3.99
		20	71.78	71.78	71.37	1.00	1.00	.99	10.89
	30	2	53.22	53.30	47.41	1.00	1.00	.99	.40
		5	73.74	73.74	72.45	1.00	1.00	.99	1.27
		10	90.23	90.22	89.56	.99	.99	.99	4.44
		20	108.56	108.49	107.91	.99	.99	.99	12.63
	50	2	89.30	89.44	79.48	.99	.99	.98	1.22
		5	123.29	123.29	121.22	.99	.99	.99	1.34
		10	151.21	151.19	150.16	.99	.99	.99	5.29
		20	181.56	181.55	180.53	1.00	1.00	.99	17.02
0.4	20	2	36.29	36.35	31.44	.99	.99	.98	1.22
		5	51.12	51.12	49.25	.99	.97	.99	2.70
		10	63.49	63.48	62.23	1.00	1.00	.99	5.59
		20	76.14	76.12	75.30	.99	.99	.99	9.38
	30	2	54.63	54.73	47.38	.99	.99	.98	1.37
		5	79.97	76.97	74.24	1.00	1.00	.99	3.11
		10	95.40	95.38	93.85	1.00	1.00	.99	7.02
		20	114.94	114.91	113.72	1.00	1.00	.99	13.41
	50	2	92.29	92.46	80.10	1.00	1.00	.98	1.34
		5	129.14	129.14	124.44	1.00	1.00	.99	3.60
		10	160.31	160.27	157.77	1.00	1.00	.99	8.55
		20	192.27	192.23	190.27	.99	.99	.99	17.95

Table 9: Values of estimates of negative-binomial log-likelihood (l), extended quasi likelihood for variance function $v_1 = \mu + c\mu^2$ (q_1), extended quasi likelihood for variance function $v_2 = c_3\mu^2$ (q_2) and relative efficiency of estimates of μ and v_1, v_2 with respect to maximum likelihood estimates for all combinations of n=20, 30, 50, μ=2, 5, 10, 20 and c=.1, .2, .4; data from log-normal mixture of Poisson

c	n	μ	$-\hat{l}$	$-\hat{q}_1$	$-\hat{q}_2$	RE_1	RE_2	$RE(v_1)$	$RE(v_2)$
0.1	20	2	34.36	34.41	30.95	.99	.99	.98	.18
		5	46.49	46.49	45.90	1.00	1.00	.99	.48
		10	56.49	56.48	56.15	1.00	1.00	.99	1.99
		20	67.25	67.25	66.97	1.00	1.00	.99	7.10
	30	2	51.95	52.02	46.84	.99	.99	.98	.16
		5	69.89	69.88	69.21	.99	.99	.99	.50
		10	84.86	84.86	84.46	.99	.99	.99	2.12
		20	101.18	101.18	100.77	1.00	1.00	.99	8.77
	50	2	87.19	87.28	79.02	.99	.99	.99	.12
		5	117.53	117.52	116.68	1.00	1.00	.99	.49
		10	142.16	142.15	141.51	1.00	1.00	.99	2.35
		20	169.81	169.81	169.11	.99	.99	.99	9.47
0.2	20	2	35.12	35.17	31.25	.99	.99	.98	.44
		5	48.26	48.26	47.37	.98	.99	.99	1.33
		10	59.31	59.31	58.75	.99	.99	.99	5.04
		20	71.18	71.18	70.72	.99	.99	.99	11.99
	30	2	53.16	53.24	47.40	.99	.99	.98	.42
		5	72.64	72.64	71.32	1.00	1.00	.99	1.47
		10	89.36	89.35	88.57	1.00	1.00	.99	5.86
		20	107.41	107.39	106.69	.99	.99	.99	17.78
	50	2	89.17	89.29	79.53	.99	.99	.99	.40
		5	123.29	123.38	120.37	1.00	1.00	.99	1.54
		10	149.75	149.73	148.39	.99	.99	.99	7.31
		20	179.62	179.61	178.43	.99	.99	.99	23.71
0.4	20	2	35.80	35.87	31.16	.99	.99	.98	1.31
		5	50.57	50.57	48.97	1.00	1.00	.99	2.96
		10	62.17	62.16	61.22	.99	.99	.99	8.25
		20	74.66	74.65	73.93	.99	.99	.99	12.11
	30	2	54.58	54.67	47.61	1.00	1.00	.98	1.36
		5	76.27	76.26	73.86	.99	.99	.99	3.87
		10	94.04	94.02	92.61	1.00	1.00	.99	10.31
		20	113.07	113.06	111.92	.99	.99	.99	16.51
	50	2	91.67	91.82	79.92	.99	.99	.98	1.41
		5	128.02	128.01	124.00	1.00	1.00	.99	4.38
		10	157.63	157.59	155.24	.99	.99	.99	14.71
		20	189.22	189.18	187.306	.99	.99	.99	28.64