Statistica Sinica **15**(2005), 257-276

SCORE TESTS FOR ZERO-INFLATION AND OVER-DISPERSION IN GENERALIZED LINEAR MODELS

Dianliang Deng and Sudhir R. Paul

University of Regina and University of Windsor

Abstract: Discrete data in the form of counts often exhibit extra variation that cannot be explained by a simple model, such as the binomial or the Poisson. Also, these data sometimes show more zero counts than what can be predicted by a simple model. Therefore, a discrete generalized linear model (Poisson or binomial) may fail to fit a set of discrete data either because of zero-inflation, because of over-dispersion, or because there is zero-inflation as well as over-dispersion in the data. Previous published work deals with goodness of fit tests of the generalized linear model against zero-inflation and against over-dispersion separately. In this paper we deal with the class of zero-inflated over-dispersed generalized linear models and propose procedures based on score tests for selecting a model that fits such data. For over-dispersion we consider a general over-dispersion model and specific over-dispersion models. We show that in certain cases and under certain conditions, the score tests derived using the general over-dispersion model and those developed under specific over-dispersion models are identical. Empirical level and power properties of the tests are examined by a limited simulation study. Simulations show that the score tests, in general, hold nominal levels well and have good power properties. Two illustrative examples and a discussion are presented.

Key words and phrases: Binomial model, generalized linear model, over-dispersion, Poisson model, score test, zero-inflation.

1. Introduction

A discrete generalized linear model (Poisson or binomial) may fail to fit a set of data having a lot of zeros either because of zero-inflation, because of over-dispersion, or because there is zero-inflation as well as over-dispersion in the data. The purpose of this paper is to propose procedures based on score tests for selecting a model that fits such data. Dean (1992) develops score tests to detect over-dispersion in a generalized linear model. Broek (1995) obtains a score test to test whether the number of zeros is too large for a Poisson distribution to fit the data well. Deng and Paul (2000) develop score tests to detect zero-inflation in a generalized linear model and obtain Broek's results as special cases.

In this paper we obtain score tests (i) for zero-inflation in presence of over-dispersion, (ii) for over-dispersion in presence of zero-inflation, and (iii) simultaneously for zero-inflation and over-dispersion. We first develop score tests using a zero-inflated over-dispersed generalized linear model in which the over-dispersed generalized linear model is of the form considered by Cox (1983) and Dean (1992). For Poisson and binomial data these score tests are then obtained as special cases. Further, score tests are obtained using a zero-inflated negative binomial model for Poisson data and a zero-inflated beta-binomial model for binomial data. These score tests are then compared with those obtained from the zero-inflated over-dispersed generalized linear model considered earlier. We show that for Poisson data the score test statistics, for testing over-dispersion in the presence of zero-inflation and that for testing simultaneously for zero-inflation and over-dispersion, using the zero-inflated over-dispersed generalized linear model with log-link is identical to the corresponding score test statistics using the zero-inflated negative binomial model. Also, for binomial data the score test statistics, for testing over-dispersion in the presence of zero-inflation and that for testing simultaneously for zero-inflation and over-dispersion, using the zero-inflated over-dispersed generalized linear model with logit link is identical to the corresponding score test statistics using the zero-inflated beta-binomial model. However, for testing zero-inflation in presence of over-dispersion the score test statistics obtained as special cases do not coincide with those obtained from the specific over-dispersed zero-inflated models. The reasons are discussed. Some simulation results, two illustrative examples and a discussion are presented.

The score test (Rao (1947)) is a special case of the more general $C(\alpha)$ test (Neyman (1959)) in which the nuisance parameters are replaced by maximum likelihood estimates which are $\sqrt{N}$ (N=number of observations used in estimating the parameters) consistent estimates. The score test is particularly appealing as we have only to study distribution of the test statistic under the null hypothesis which is that of the basic model. It often maintains, at least approximately, a preassigned level of significance (Bartoo and Puri (1967)) and often produces a statistic which is simple to calculate. Other asymptotically equivalent tests, such as the likelihood ratio test and the Wald test (Moran (1970); Cox and Hinkley (1974)), can be considered. However, both of these tests require estimates of the parameters under the alternative hypotheses and often show liberal or conservative behaviour in small samples (See, for example, Barnwal and Paul (1988), Thall (1992) and Paul and Banerjee (1998)). To check this point in the present context we conducted a small scale simulation experiment to study the small sample behaviours of the likelihood ratio statistic and the Wald test statistic in testing simultaneously for no zero-inflation and no over-dispersion. The results, which are not presented in this paper, show that the likelihood ratio test is too liberal and the Wald test is too conservative. Therefore, these large sample tests are not considered in this paper.

In Section 2 we introduce the zero-inflated generalized linear model. Score tests for selecting a model from the class of zero-inflated over-dispersed generalized linear models are developed in Section 3. Results for Poisson data based on the zero-inflated over-dispersed generalized linear model with log-link and the zero-inflated negative binomial model are obtained and compared in Section 4, and results for binomial data based on the zero-inflated over-dispersed generalized linear model with logit link and the zero-inflated beta-binomial model are obtained and compared in Section 5. Results of some simulation experiments are presented in Section 6, and two illustrative examples are presented in Section 7.

2. The Zero-Inflated Over-Dispersed Generalized Linear Model

Consider the natural exponential family distribution with probability density function

$$f(y;\theta) = \exp\{a(\theta)y - g(\theta) + c(y)\}, \tag{2.1}$$

where y represents the response variable and θ is an unknown parameter. This family includes both the Poisson and binomial distribution. Departure from the generalized linear model (Poisson or binomial) may be because of having a lot of zeros in the data or because the data are over-dispersed.

The exponential family distribution with zero-inflation has probability density

$$f_1(y;\theta) = \begin{cases} \omega + (1-\omega)f(0;\theta), & \text{if } y = 0, \\ (1-\omega)f(y;\theta), & \text{if } y > 0, \end{cases} \tag{2.2}$$

where ω is the zero-inflation(deflation) parameter which can take negative values. Note that a zero-inflated model has $\omega > 0$ and a zero-deflated model has $\omega < 0$. The mean and variance of Y under (2.2) are $E(Y) = (1-\omega)(a'(\theta))^{-1}g'(\theta)$ and $\text{Var}\,(Y) = ((1-\omega)(a'(\theta))^2\{g''(\theta) - a''(\theta)(a'(\theta))^{-1}g'(\theta) + \omega(g'(\theta))^2\}$.

Now, suppose that for given θ^*, y has the exponential family model with probability density function $f(y;\theta^*) = \exp\{a(\theta^*)y - g(\theta^*) + c(y)\}$, where θ^* is continuous independent random variate with $E(\theta^*) = \theta(x;\beta)$, $\text{Var}\,(\theta^*) = \tau b(\theta) > 0$, $\alpha_r = E(\theta^* - \theta)^r$, where β is the $p \times 1$ vector of regression parameters and τ is the over-dispersed parameter. Then following Cox (1983), Chesher (1984) and Dean (1992) the probability function of the over-dispersed exponential family model is

$$f_2(y;\theta,\tau) = f(y;\theta)\Big\{1 + \sum_{r=2}^{\infty}\frac{\alpha_r}{r!}D_r(y;\theta)\Big\} \equiv f(y;\theta)\{1 + D(y;\theta,\tau)\}, \tag{2.3}$$

where $D_r(y,\theta) = \{\partial^{(r)}/\partial\theta^{*(r)})f(y;\theta^*)|_{\theta^*=\theta}\}\ \{f(y;\theta)\}^{-1}$ and $D(y,\theta,\tau) = \sum_{r=2}^{\infty}$ $(\alpha_r/r!)D_r(y;\theta)$. Further, for small τ, we assume that $\alpha_r = o(\tau)$ for $r \geq 3$ and

$$f_2(y;\theta,\tau) = f(y;\theta)\{1 + \frac{\alpha_2}{2!}D_2(y;\theta)\} = f(y;\theta)\{1 + \frac{1}{2}\tau b(\theta)D_2(y;\theta)\}.$$

The zero-inflated over-dispersed exponential family model then can be written as

$$f_3(y;\theta,\tau,\omega) = \begin{cases} \omega + (1-\omega)f_2(0;\theta,\tau) & \text{if } y = 0 \\ (1-\omega)f_2(y;\theta,\tau) & \text{if } y > 0. \end{cases} \tag{2.4}$$

Obviously (2.4) generalizes (2.1), (2.2) and (2.3). The mean and variance of y for (2.4) do not have closed forms.

3. Model Selection in the Zero-Inflated Over-Dispersed Generalized Linear Model

For discrete data in the form of counts or proportions, one of following discrete generalized linear models may fit the data: (i) a generalized linear (a Poisson or a binomial) model; (ii) a zero-inflated generalized linear model; (iii) a over-dispersed generalized linear model; (iv) a zero-inflated over-dispersed generalized linear model. Under (2.3), Dean (1992) develops score tests to detect over-dispersion in the generalized linear model. She then obtains score tests to detect over-dispersion in Poisson and binomial data separately, as special cases of the results she obtains for the generalized linear model. Broek (1995) obtains a score test to test whether the number of zeros is too large for a Poisson distribution to fit the data well. Using (2.2), Deng and Paul (2000) develop score tests to detect zero-inflation in generalized linear model and obtain score tests to test for zero-inflation in Poisson and binomial data separately, as special cases of the results they obtain for the generalized linear model. They show that for the Poisson data their results are identical to those obtained by Broek (1995).

In this section, we derive score test statistics, using (2.4) to test (i) for over-dispersion in presence of zero-inflation, (ii) for zero-inflation in presence of over-dispersion, and (iii) simultaneously for zero-inflation and over-dispersion.

Let Y_i, $i = 1,\ldots,n$, be a sample of independent observations from (2.4) with θ_i a function of $p\times 1$ vector of covariates X_i and a vector of regression parameters β; that is, $\theta_i = \theta_i(X_i;\beta)$, $i = 1,\ldots,n$. Then the likelihood function is

$$L(\omega,\tau,\theta;y) = \prod_{i=1}^{n}\{(\omega + (1-\omega)f_2(0;\theta,\tau))I_{\{y_i=0\}} + (1-\omega)f_2(y_i;\theta,\tau)I_{\{y_i>0\}}\}.$$

Writing $\gamma = \omega/(1-\omega)$ the log likelihood $l = l(\gamma,\tau,\theta;y)$ can be written as

$$l(\gamma,\tau,\theta;y) = \sum_{i=1}^{n} l_i(\gamma,\tau,\theta_i;y_i)$$

$$= \sum_{i=1}^{n} \{-\log(1+\gamma) + I_{\{y_i=0\}} \log(\gamma + f_2(0;\theta,\tau)) + I_{\{y_i>0\}} \log f_2(y_i;\theta,\tau)\}$$
$$= \sum_{i=1}^{n} \{-\log(1+\gamma) + I_{\{y_i=0\}} \log(\gamma + f(0;\theta_i)\{1 + D(y,\theta,\tau)\})$$
$$+ I_{\{y_i>0\}}(a(\theta_i)y_i - g(\theta_i) + c(y_i) + \log\{1 + D(y,\theta,\tau)\})\}.$$

Now, define the parameter vector $\delta = (\beta', \gamma, \tau)'$. Partition $\delta = (\delta_1', \delta_2')'$. Suppose we want to test $H_0 : \delta_2 = 0$ against $H_A : \delta_2 > 0$. The dimension of the parameter vector δ_2 will depend on the null hypothesis to be tested. For example, for testing $H_0 : \tau = 0, \delta_1 = (\beta', \gamma)', \delta_2 = \tau$ and the dimension of δ_2 is 1. Similarly, for testing $H_0 : (\tau, \gamma) = (0, 0), \delta_1 = \beta, \delta_2 = (\tau, \gamma)$ and the dimension of δ_2 is 2. Further define the likelihood score $S = (\partial l/\partial \delta_2)|_{\delta_2=0}$ and the expected mixed second partial derivative matrices, $I_{11} = E(-(\partial^2 l/\partial \delta_1 \partial \delta_1')|_{\delta_2=0}), I_{12} = E(-(\partial^2 l/\partial \delta_1 \partial \delta_2')|_{\delta_2=0})$ and $I_{22} = E(-(\partial^2 l/\partial \delta_2 \partial \delta_2')|_{\delta_2=0})$. Then, under some conditions for the application of the Central Limit Theorem to score components and the regularity conditions of maximum likelihood estimates, the score test statistic for testing $H_0 : \delta_2 = 0$ is $T = \hat{S}'(\hat{I}_{22} - \hat{I}_{12}'\hat{I}_{11}^{-1}\hat{I}_{12})^{-1}\hat{S}$, which, asymptotically, has a chi-square distribution with d degrees of freedom , where d is dimension of δ_2, $\hat{S} = S(\hat{\delta}_1), \hat{I}_{11} = I_{11}(\hat{\delta}_1), \hat{I}_{12} = I_{12}(\hat{\delta}_1)$, $\hat{I}_{22} = I_{22}(\hat{\delta}_1)$ and $\hat{\delta}_1$ is the maximum likelihood estimate of δ_1 under the null hypothesis.

We now give the score test statistics for the three null hypotheses $H_0 : \tau = 0, H_0 : \gamma = 0$ and $H_0 : (\tau, \gamma) = (0, 0)$ in Theorem 1, Theorem 2 and Theorem 3 respectively. The derivations are given in the Appendix. In what follows the dependence on θ_i of the functions $\mu(\theta_i), \sigma^2(\theta_i), a(\theta_i), b(\theta_i), g(\theta_i)$ and $D_r(\theta_i)$ will be suppressed for simplicity of natation. For convenience, we replace $f_2(0;\theta_i), a(\theta_i), g(\theta_i), b(\theta_i), D_2(y_i, \theta_i)$ and $D(y_i; \theta_i, \tau)$ with f_0, a, g, b, D_2 and D respectively.

Theorem 3.1. *Let* $\mathbf{1}$ *be an* $n \times 1$ *unit vector,* U *an* $n \times p$ *matrix with ir-element* $\partial\theta_i/\partial\beta_r$, W_1^τ, W_2^τ *and* W_3^τ *diagonal matrices with ith elements* $W_{1i}^\tau = g'' - a''Ey_i - (f_0\gamma/(\gamma + f_0)(1+\gamma))g'^2 - (\gamma/(1+\gamma))g'', W_{2i}^\tau = -(f_0/(\gamma + f_0)(1+\gamma))g'$ *and* $W_{3i}^\tau = [(1/2)g'bD_2(f_0\gamma/(\gamma+f_0)(1+\gamma)) + (1/2)(bD_2)'(\gamma/(1+\gamma))]|_{y_i=0} - (1/2)E[(bD_2)']$ *respectively. Further, let* $S_\tau = \sum_{i=1}^n -(\gamma I_{\{y_i=0\}}/(\gamma + f_0)) + (1/2)$ bD_2, $I_{\tau\tau}^\tau = \sum_{i=1}^n[(1/4)E(bD_2)^2 - ((\gamma^2 + 2f_0\gamma)/(\gamma+f_0)(1+\gamma))((1/2)bD_2)^2|_{y_i=0}]$, $I_{\gamma\gamma}^\tau = \sum_{i=1}^n((1-f_0)/(1+\gamma)^2(\gamma+f_0)|_{y_i=0}$ *and* $I_{\gamma\tau}^\tau = \sum_{i=1}^n((1/2)bD_2f_0/(\gamma + f_0)$ $(1+\gamma))|_{y_i=0}$. *Then the score test statistic for over-dispersion in the zero-inflated over-dispersed generalized linear model is* $T_1 = \hat{S}_\tau^2/\hat{V}_\tau$, *where,* $\hat{S}_\tau = S_\tau(\hat{\theta_1}, \ldots, \hat{\theta_n}; \hat{\gamma})$ *and* $\hat{V}_\tau = V_\tau(\hat{\theta_1}, \ldots, \hat{\theta_n}; \hat{\gamma})$ *with*

$$V_\tau = I_{\tau\tau} - \mathbf{1}^T W_3 U (U^T W_1 U)^{-1} U^T W_3 \mathbf{1}$$
$$-(I_{\gamma\tau} - (\mathbf{1}^T W_3 U)(U^T W_1 U)^{-1}(U^T W_2 \mathbf{1}))^2 (I_{\gamma\gamma} - \mathbf{1}^T W_2 U (U^T W_1 U)^{-1} U^T W_2 \mathbf{1})^{-1}.$$

$\hat{\theta}_i$ and $\hat{\gamma}$ are the maximum likelihood estimates of θ_i and γ under the zero-inflated generalized linear model. The statistic T_1 has an asymptotic χ^2 distribution with one degree of freedom.

Theorem 3.2. *Let $\mathbf{1}$ be an $n \times 1$ unit vector, U an $n \times p$ matrix with ir-element $(\partial\theta_i/\partial\beta_r)$, W_1^γ, W_2^γ and W_3^γ diagonal matrices with ith elements $W_{1i}^\gamma = g'' - a''Ey_i - E\{\partial^2/\partial\theta_i^2)\log(1+D)\}, W_{2i}^\gamma = -E\{(\partial^2/\partial\theta_i\partial\tau)\log(1+D)\}, W_{3i}^\gamma = -g' + \{(\partial/\partial\theta_i)\log(1+D)\}|_{y_i=0}$ respectively. Further, let $S_\gamma = \sum_{i=1}^n (I_{\{y_i=0\}}/f_2(0;\theta_i,\tau))$ $-1), I_{\tau\tau}^\gamma = \sum E\{-(\partial^2/\partial\tau^2)\log(1+D)\}, I_{\tau\gamma}^\gamma = \sum(\partial/\partial\tau)\log(1+D)|_{y_i=0}$ and $I_{\gamma\gamma}^\gamma = \sum(1/f_2(0;\theta_i,\tau) - 1)$. Then the score test statistic for zero-inflation in the zero-inflated over-dispersed generalized linear model is $T_2 = \hat{S}_\gamma^2/\hat{V}_\gamma$, where $\hat{S}_\gamma = S_\gamma(\hat{\theta}_1,\ldots,\hat{\theta}_n;\hat{\tau})$ and $\hat{V}_\gamma = V_\gamma(\hat{\theta}_1,\ldots,\hat{\theta}_n;\hat{\tau})$ with*

$$V_\gamma = I_{\gamma\gamma} - \mathbf{1}^T W_3 U (U^T W_1 U)^{-1} U^T W_3 \mathbf{1} - (I_{\tau\gamma} - (\mathbf{1}^T W_3 U)(U^T W_1 U)^{-1} \times (U^T W_2 \mathbf{1}))^2 (I_{\tau\tau} - \mathbf{1}^T W_2 U (U^T W_1 U)^{-1} U^T W_2 \mathbf{1})^{-1}$$

and $\hat{\theta}_i$ and $\hat{\tau}$ are the maximum likelihood estimates of θ_i and τ under the over-dispersed generalized linear model. The statistic T_2 has an asymptotic χ^2 distribution with one degree of freedom.

Theorem 3.3. *Let $\mathbf{1}$ be an $n \times 1$ unit vector, U an $n \times p$ matrix with ir-element $\partial\theta_i/\partial\beta_r$, W_1, W_2 and W_3 diagonal matrices with ith elements $W_{1i} = g_i'' - a_i''Ey_i, W_{2i} = -g', W_{3i} = (1/2)b_i[g_i'\{(h_i')^2 - h_i''\} - 2g_i''h_i' + g_i''']$ respectively. Further, let $S_1 = \sum_{i=1}^n((I_{\{y_i=0\}}/f(0;\theta_i)) - 1), S_2 = \sum_{i=1}^n(1/2)b_i(a_i')^2\{(y_i-\mu_i)^2 - (a_i')^{-2}(g_i'' - a_i''y_i)\}, I_{\gamma\gamma} = \sum_{i=1}^n(1/f(0;\theta_i) - 1), I_{\gamma\tau} = \sum_{i=1}^n(1/2)b_i[(g_i')^2 - g_i'']$ and $I_{\tau\tau} = (1/4)b_i^2[g_i'\{5h_i'h_i'' - 3(h_i')^3 - h_i'''\} + 2(h_i'g_i' - g_i'')^2 + g_i''\{6(h_i')^2 - 4h_i''\} - 4h_i'''h_i' + g_i'''']$. Then the score test statistic for zero-inflation and over-dispersion in the zero-inflated over-dispersed generalized linear model is*

$$T_3 = \frac{\hat{V}_{22}\hat{S}_1^2 + \hat{V}_{11}\hat{S}_2^2 - 2\hat{V}_{12}\hat{S}_1\hat{S}_2}{\hat{V}_{22}\hat{V}_{11} - \hat{V}_{12}^2},$$

which has an asymptotic χ^2 distribution with two degrees of freedom, where $V_{11} = I_{\gamma\gamma} - \mathbf{1}^T W_2 U (U^T W_1 U)^{-1} U^T W_2 \mathbf{1}, V_{12} = I_{\gamma\tau} - \mathbf{1}^T W_2 U (U^T W_1 U)^{-1} U^T W_3 \mathbf{1}, V_{22} = I_{\tau\tau} - \mathbf{1}^T W_3 U (U^T W_1 U)^{-1} U^T W_3 \mathbf{1}, \hat{S}_1 = S_1(\hat{\theta}), \hat{S}_2 = S_2(\hat{\theta}), \hat{V}_{11} = V_{11}(\hat{\theta}), \hat{V}_{22} = V_{22}(\hat{\theta})$ and $\hat{V}_{12} = V_{12}(\hat{\theta})$ and $\hat{\theta}$ is the maximum likelihood estimate of $\theta = (\theta_1,\ldots,\theta_n)$ under the generalized linear model.

4. Score Test for Poisson Data

In Sections 4.1−4.3, score tests are obtained from Theorems 3.1−3.3 with $\theta = \log\mu = X\beta$, where X is a $n \times p$ matrix of covariates and β is a $p \times 1$ vector of regression parameters, $a(\theta) = \theta, g(\theta) = e^\theta$ and $h(\theta) = 0$. Note that in this case,

$U = X$. Further, score tests are also obtained using the zero-inflated negative binomial model

$$P(Y=0) = \omega + (1-\omega)(\frac{1}{1+c\mu})^{c^{-1}},$$
$$P(Y=y) = (1-\omega)\frac{\Gamma(y+c^{-1})}{y!\Gamma(c^{-1})}(\frac{c\mu}{1+c\mu})^{y}(\frac{1}{1+c\mu})^{c^{-1}}$$

with $E(Y) = (1-\omega)\mu$ and $\mathrm{Var}\,(Y) = (1-\omega)\mu + (1-\omega)(\omega+c)\mu^2$.

4.1. Testing for over-dispersion

Using (2.4) the null hypothesis to be tested is $H_0 : \tau = 0$. Then, from Theorem 3.1, the score test statistic for testing for over-dispersion can be written in a simplified form as

$$Z_1 = \frac{(\sum_{i=1}^{n} \frac{1}{2}((y_i - \hat{\mu_i})^2 - \hat{\mu_i}) - \frac{\hat{\gamma}(\mu_i^2-\mu_i)I_{\{y_i=0\}}}{2(\hat{\gamma}+e^{-\hat{\mu}_i})})^2}{\hat{V}_\tau}.$$

Here $\hat{\mu}_i = \exp(\sum_{j=1} X_{ij}\hat{\beta}_j)$ and $\hat{\beta}_j$ and $\hat{\gamma}$ are the maximum likelihood estimates of β_j and γ under the zero-inflated Poisson model obtained by solving the estimating equations

$$\sum_{i=1}^{n}\left(\frac{-1}{1+\gamma} + \frac{I_{\{y_i=0\}}}{\gamma+e^{-\mu_i}}\right) = 0, \tag{4.1}$$

$$\sum_{i=1}^{n}\left(I_{\{y_i=0\}}\frac{\gamma\mu_i}{\gamma+e^{-\mu_i}} + (y_i-\mu_i)\right)X_{ij} = 0, \text{ for } j = 1,2,\ldots,p, \tag{4.2}$$

and where $\hat{V}_\tau = V_\tau(\hat{\mu},\hat{\gamma})$ with

$$V_\tau = I^\tau_{\tau\tau} - \mathbf{1}^T W^\tau_3 X(X^T W^\tau_1 X)^{-1} X^T W^\tau_3 \mathbf{1} - (I^\tau_{\gamma\tau} - \mathbf{1}^T W^\tau_3 X(X^T W^\tau_1 X)^{-1} X^T W^\tau_2 \mathbf{1})^2$$
$$\times (I^\tau_{\gamma\gamma} - \mathbf{1}^T W^\tau_2 X(X^T W^\tau_1 X)^{-1} X^T W^\tau_2 \mathbf{1})^{-1},$$

$$I^\tau_{\gamma\gamma} = \sum_{i=1}^{n}\frac{e^{\mu_i}-1}{(1+\gamma)^2(1+\gamma e^{\mu_i})}, \qquad I^\tau_{\gamma\tau} = \sum_{i=1}^{n}\frac{\mu_i^2-\mu_i}{2(1+\gamma)(1+\gamma e^{\mu_i})},$$

$$I^\tau_{\tau\tau} = \sum_{i=1}^{n}[\frac{2\mu_i^2+\mu_i}{4(1+\gamma)} - \frac{\gamma(\mu_i^2-\mu_i)^2}{(1+\gamma)(1+\gamma e^{\mu_i})}, \qquad W^\tau_{1i} = \frac{\mu_i}{1+\gamma} - \frac{\gamma\mu_i^2}{(1+\gamma)(1+\gamma e^{\mu_i})},$$

$$W^\tau_{2i} = -\frac{\mu_i}{(1+\gamma)(1+\gamma e^{\mu_i})}, \qquad W^\tau_{3i} = \frac{\mu_i}{2(1+\gamma)} + \frac{1}{2}\frac{\gamma\mu_i(\mu_i^2-\mu)}{(1+\gamma)(1+\gamma e^{\mu_i})}.$$

Using the zero-inflated negative binomial model the null hypothesis to be tested is $H_0 : c = 0$. The score test statistic, in this case, can be shown to be

$$Z_2 = \frac{(\sum_{i=1}^{n} \frac{1}{2}((y_i - \hat{\mu}_i)^2 - y_i) - I_{\{y_i=0\}}\frac{\hat{\mu}_i^2\hat{\gamma}}{2(\hat{\gamma}+e^{-\hat{\mu}_i})})^2}{\hat{V}_c}.$$

Here $\hat{\mu}_i = \exp(\sum_{j=1}^{p} X_{ij}\hat{\beta}_j)$ and $\hat{\beta}_j$ and $\hat{\gamma}$ are the same as those obtained from (4.1) and (4.2), where $\hat{V}_c = V_c(\hat{\mu}, \hat{\gamma})$ with

$$V_c = I_{cc} - \mathbf{1}^T W_3 X (X^T W_1 X)^{-1} X^T W_3 \mathbf{1} - (I_{c\gamma} - \mathbf{1}^T W_3 X (X^T W_1 X)^{-1} X^T W_2 \mathbf{1})^2 \times (I_{\gamma\gamma} - \mathbf{1}^T W_2 X (X^T W_1 X)^{-1} X^T W_2 \mathbf{1})^{-1},$$

where $W_1 = W_1^\tau, W_2 = W_2^\tau, W_3 = W_3^\tau - (1/2)W_1^\tau, I_{\gamma\gamma} = I_{\gamma\gamma}^\tau, I_{\gamma c} = I_{\gamma\tau}^\tau - (1/2)\sum W_{2i}^\tau$ and $I_{cc} = I_{\tau\tau}^\tau + (1/4)\sum W_{1i}^\tau - \sum W_{3i}^\tau$. After simplification, it can be shown that $V_c = V_\tau$.

It is interesting to note that when the regression model involves an intercept term, that is, when $X_{i1} = 1$ for $i = 1, \ldots, n$, the statistics Z_1 and Z_2 are identical, seen as follows.

From (4.2), we have $\sum_{i=1}^n \hat{\mu}_i = \sum_{i=1}^n \{-I_{\{y_i=0\}}(\hat{\gamma}\hat{\mu}_i/(\hat{\gamma} + e^{-\hat{\mu}_i})) + y_i\}$, from which

$$\sum_{i=1}^{n} \frac{1}{2}((y_i - \hat{\mu}_i)^2 - \hat{\mu}_i) - \frac{\hat{\gamma}(\hat{\mu}_i^2 - \hat{\mu}_i) I_{\{y_i=0\}}}{2(\hat{\gamma} + e^{-\hat{\mu}_i})} = \sum_{i=1}^{n} \frac{1}{2}((y_i - \hat{\mu}_i)^2 - y_i) - I_{\{y_i=0\}} \frac{\hat{\mu}_i^2 \hat{\gamma}}{2(\hat{\gamma} + e^{-\hat{\mu}_i})},$$

that is, the numerators of Z_1 and Z_2 are equal. Since $V_c = V_\tau$, $Z_1 = Z_2$.

Ridout, Hinde and Demetrio (2001) obtain a score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. We obtain a more general version. In particular, for no covariate case, we have $\mu_i = \mu$ and

$$Z_2^* = \frac{\left\{\sum_{i=1}^n ((y_i - \hat{\mu})^2 - y_i) - \frac{n_0 \hat{\mu}^2 \hat{\gamma}}{(\hat{\gamma} + e^{-\hat{\mu}})}\right\}^2}{\frac{\hat{\mu}^2}{1+\hat{\gamma}}(2 - \frac{\hat{\mu}^2}{e^{\hat{\mu}} - 1 - \hat{\mu}})},$$

where n_0 is the number of zeros in the values of y_i, which agrees with Ridout et al. (2001).

4.2. Testing for zero-inflation

The score test statistic given in Theorem 3.2 is derived using (2.4) in which we use (2.3). This statistic does not produce a simplified form for Poisson or binomial data and it poses computational difficulty. So, as in Cox (1983), we assume that for small τ that $\alpha_r = o(\tau)$, for $r = 3, \ldots, \infty$. Then, under the null hypothesis $H_0 : \gamma = 0$, the score statistic is obtained in a simple form as

$$Z_3 = \Big(\sum_{i=1}^{n} \frac{I_{\{y_i=0\}} e^{\hat{\mu}_i}}{1 + \frac{1}{2}\hat{\tau}((y_i - \hat{\mu}_i)^2 - \hat{\mu}_i)} - 1\Big)^2 / \hat{V}_\gamma.$$

Here $\hat{\mu}_i = \exp(\sum_{j=1} X_{ij}\hat{\beta}_j)$ and $\hat{\beta}_j$ and $\hat{\tau}$ are the maximum likelihood estimates of β_j and τ under (2.3) with $\alpha_r = o(\tau)$ for $r = 3, \ldots, \infty$, and $\hat{V}_\gamma = V_\gamma(\hat{\mu}, \hat{\tau})$ with

$$V_\gamma = I^\gamma_{\gamma\gamma} - \mathbf{1}^T W^\gamma_3 X(X^T W^\gamma_1 X)^{-1} X^T W^\gamma_3 \mathbf{1} - (I^\gamma_{\tau\gamma} - \mathbf{1}^T W^\gamma_3 X(X^T W^\gamma_1 X)^{-1} \times X^T W^\gamma_2 \mathbf{1})^2 (I_{\tau\tau} - \mathbf{1}^T W^\gamma_2 X(X^T W^\gamma_1 X)^{-1} U^T W^\gamma_2 \mathbf{1})^{-1},$$

$$I^\gamma_{\tau\tau} = \frac{1}{4}\sum_{i=1}^n (2\mu_i^2 + \mu_i), \qquad I^\gamma_{\gamma\tau} = \frac{1}{2}\sum_{i=1}^n (\mu_i^2 - \mu_i), \qquad I^\gamma_{\gamma\gamma} = \sum_{i=1}^n \Big(\frac{e^{\mu_i}}{1 + \frac{1}{2}\tau(\mu_i^2 - \mu_i)} - 1\Big),$$

$$W^\gamma_{1i} = \mu_i - \frac{1}{2}\tau(2\mu_i^2 - \mu_i), \quad W^\gamma_{2i} = \frac{1}{2}\mu_i - \frac{1}{2}\tau\mu_i^2, \quad W^\gamma_{3i} = -\mu_i + \frac{1}{2}\tau(2\mu_i^2 - \mu_i).$$

Further, using the zero-inflated negative binomial model, we obtain the score test statistic for testing the null hypothesis $H_0 : \gamma = 0$ as

$$Z_4 = \frac{(\sum_{i=1}^n I_{\{y_i=0\}}(1 + c\hat{\mu}_i)^{c^{-1}} - 1)^2}{\hat{V}_\gamma}.$$

Here $\hat{\mu}_i = \exp(\sum_{j=1} X_{ij}\hat{\beta}_j)$ and $\hat{\beta}_j$ and $\hat{c}$ are the maximum likelihood estimates of β_j and c under the negative binomial model, and $\hat{V}_\gamma = V_\gamma(\hat{\mu}, \hat{c})$ with

$$V_\gamma = I_{\gamma\gamma} - \mathbf{1}^T W_3 X(X^T W_1 X)^{-1} X^T W_3 \mathbf{1} - (I_{\gamma c} - \mathbf{1}^T W_3 X(X^T W_1 X)^{-1} X^T W_2 \mathbf{1})^2 \times (I_{cc} - \mathbf{1}^T W_2 X(X^T W_1 X)^{-1} X^T W_2 \mathbf{1})^{-1},$$

$$I_{\gamma\gamma} = \sum_{i=1}^n [(1 + c\mu_i)^{c^{-1}} - 1], \qquad I_{\gamma c} = \sum_{i=1}^n \Big[\frac{1}{c^2}\log(1 + c\mu_i) - \frac{\mu_i}{c(1 + c\mu_i)}\Big],$$

$$I_{cc} = \sum_{i=1}^n \Big\{E\Big[\sum_{l=1}^{y_i} \frac{(l-1)^2}{(1 + (l-1)c)^2}\Big] - \frac{2\log(1 + c\mu_i)}{c^3} - \frac{2\mu_i + c\mu_i^2}{c^2(1 + c\mu_i)}\Big\},$$

$$W_{1i} = \frac{\mu_i}{(1 + c\mu_i)}, \qquad W_{2i} = 0, \qquad W_{3i} = -\frac{\mu_i}{(1 + c\mu_i)}.$$

Theoretically, a relationship between the statistics Z_3 and Z_4 cannot be established, because the over-dispersion parameter c under the negative binomial model is not the same as that under (2.3) with $\alpha_r = o(\tau)$, for $r = 3, \ldots, \infty$. Further, simulations, results of which are not reported here, show extremely poor performance of the statistic Z_3. Note that the statistic Z_4 has a simple form and we show by simulations in Section 6 that it holds its level even when the over-dispersed model is not negative binomial. We recommend its use in practice for testing for zero-inflation in the presence of over-dispersion.

4.3. Testing for over-dispersion and zero-inflation

Using the results in Theorem 3.3 the score test statistic for testing the hypothesis $H_0 : (\gamma, \tau) = 0$ is

$$Z_5 = \frac{(\sum_{i=1}^n ((y_i - \hat{\mu}_i)^2 - \hat{\mu}_i))^2}{2\sum_{i=1}^n \hat{\mu}_i^2} + \frac{(\sum_{i=1}^n (\frac{I_{\{y_i=0\}}}{e^{-\hat{\mu}_i}} - 1 - \frac{1}{2}(y_i - \hat{\mu}_i)^2 + \frac{1}{2}\hat{\mu}_i))^2}{(\sum_{i=1}^n e^{\hat{\mu}_i} - 1 - \hat{\mu}_i - \frac{1}{2}\hat{\mu}_i^2)},$$

where $\hat{\mu_i}$ is the maximum likelihood estimate of the parameter μ_i under the Poisson model.

Further, using the zero-inflated negative binomial model, the score test statistic for testing the hypothesis $H_0 : (\gamma, c) = 0$ is:

$$Z_6 = \frac{(\sum_{i=1}^n((y_i - \hat{\mu_i})^2 - y_i))^2}{2\sum_{i=1}^n \hat{\mu_i}^2} + \frac{(\sum_{i=1}^n(\frac{I_{\{y_i=0\}}}{e^{-\hat{\mu_i}}} - 1 - \frac{1}{2}(y_i - \hat{mu_i})^2 + \frac{1}{2}y_i))^2}{(\sum_{i=1}^n e^{\hat{\mu_i}} - 1 - \hat{\mu_i} - \frac{1}{2}\hat{\mu_i}^2)},$$

where $\hat{\mu_i}$ is the maximum likelihood estimate of the parameter μ_i, also, under the Poisson model.

Note that under the Poisson model we have $\sum \hat{\mu}_i = \sum y_i$. Thus the two statistics Z_5 and Z_6 are identical. Further, note that the test statistic is the sum of two terms. The first term is the score test statistic for testing for over-dispersion in Poisson data (see, Dean (1992)). The second term is closely related, but not identical, to the score test statistic for testing for zero-inflation in Poisson data (see, Deng and Paul (2000)). There seems to be some confounding effect between zero-inflation and over-dispersion. In connection with discrete exponential mixture models, Lindsay and Roeder (1992) find such confounding of mixtures with over-dispersion.

5. Score Test for Binomial Data

In Sections 5.1−5.3 we obtain score tests for binomial data. Here score tests are obtained from Theorems 3.1−3.3 using $\theta = \pi$, $a(\theta) = \log\{\pi/(1-\pi)\}$, $g(\theta) = -m\log(1-\theta)$, $b(\theta) = \theta(1-\theta)$ and $h(\theta) = -\log\theta - \log(1-\theta)$(in this case $U = \text{diag}(\pi_i(1-\pi_i))X$), and also by using the zero-inflated beta binomial model

$$P(Y_i = 0) = \omega + (1-\omega)\frac{\prod_{r=0}^{m-1}(1 + r\phi - \pi)}{\prod_{r=0}^{m-1}(1 + r\phi)},$$
$$P(Y_i = y_i) = (1-\omega)\binom{m}{y_i}\frac{\prod_{r=0}^{y-1}(\pi + r\phi)\prod_{r=0}^{m-y-1}(1 - \pi + r\phi)}{\prod_{r=0}^{m-1}(1 + r\phi)},$$

with $E(Y) = (1-\omega)m\pi$ and $\text{Var}\,(Y) = (1-\omega)m\pi(1-\pi)\frac{1+m\phi}{1+\phi} + (1-\omega)\omega m^2\pi^2$.

5.1 Testing for over-dispersion

Here the null hypothesis to be tested is $H_0 : \tau = 0$. Using the results in Theorem 3.1 the score test statistic for testing for over-dispersion is

$$Z_7 = \frac{\left(\sum_{i=1}^{n}\left(-\frac{I_{\{y_i=0\}}\hat{\gamma}}{\hat{\gamma}+(1-\hat{\pi}_i)^{m_i}}\frac{\hat{\pi}_i m_i(m_i-1)}{2(1-\hat{\pi}_i)} + \frac{1}{2\hat{\pi}_i(1-\hat{\pi}_i)}((y_i - m_i\hat{\pi}_i)^2 + (y_i - m_i\hat{\pi}_i)\hat{\pi}_i - y_i(1-\hat{\pi}_i))\right)\right)^2}{\hat{V}_\tau}.$$

Here $\hat{\pi}_i = (\exp(\sum X_{ij}\hat{\beta}_j))/(1+\exp(\sum X_{ij}\hat{\beta}_j))$ and $\hat{\beta}_j$ and $\hat{\gamma}$ are the maximum likelihood estimates of β_j and γ under the null hypothesis, and $\hat{V}_\tau = V_\tau(\hat{\pi},\hat{\gamma})$ with

$$
\begin{aligned}
V_\tau(\pi,\gamma) &= I^\tau_{\tau\tau} - \mathbf{1}^T W^\tau_3 U(U^T W^\tau_1 U)^{-1} U^T W^\tau_3 \mathbf{1} \\
&\quad -(I^\tau_{\gamma\tau} - \mathbf{1}^T W^\tau_3 U(U^T W^\tau_1 U)^{-1} U^T W^\tau_2 \mathbf{1})^2 (I^\tau_{\gamma\gamma} - \mathbf{1}^T W^\tau_2 U(U^T W^\tau_1 U)^{-1} U^T W^\tau_2 \mathbf{1})^{-1}, \\
I^\tau_{\gamma\gamma} &= \sum_{i=1}^n \frac{1-(1-\pi_i)^{m_i}}{(1+\gamma)^2(\gamma+(1-\pi_i)^{m_i})}, \qquad I^\tau_{\gamma\tau} = \sum_{i=1}^n \left(\frac{(1-\pi_i)^{m_i-1}\pi_i m_i(m_i-1)}{2(1+\gamma)(\gamma+(1-\pi_i)^{m_i})}\right), \\
I^\tau_{\tau\tau} &= \sum_{i=1}^n \Big(\frac{m_i(m_i-1)}{2} - \frac{m_i^2(m_i-1)^2}{4(1-\pi_i)^2}\frac{(2(1-\pi_i)^{m_i}+\gamma)\gamma}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i}}\Big), \\
W^\tau_{1i} &= \frac{m_i}{\pi_i(1-\pi_i)} - \frac{\gamma m_i}{(1+\gamma)(1-\pi_i)^2} - \frac{\gamma(1-\pi_i)^{m_i-2}m_i^2}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i})}, \\
W^\tau_{2i} &= -\frac{m_i(1-\pi_i)^{m_i-1}}{(1+\gamma)(\gamma+(1-\pi_i)^{m_i})}, \\
W^\tau_{3i} &= \frac{\gamma\pi_i m_i^2(m_i-1)(1-\pi_i)^{m_i-2}}{2(1+\gamma)(\gamma+(1-\pi_i)^{m_i})} + \frac{\gamma m_i(m_i-1)}{2(1+\gamma)(1-\pi_i)^2}.
\end{aligned}
$$

The maximum likelihood estimates $\hat{\beta}_j$ and $\hat{\gamma}$ are obtained by solving the estimating equations:

$$
\begin{aligned}
&\sum_{i=1}^n \Big(\frac{-1}{1+\gamma} + \frac{I_{\{y_i=0\}}}{\gamma+(1-\pi_i)^{m_i}}\Big) = 0, \\
&\sum_{i=1}^n \left(I_{\{y_i=0\}}\frac{\gamma m_i}{(\gamma+(1-\pi_i)^{m_i})(1-\pi_i)} + \frac{(y_i - m_i\pi_i)}{\pi_i(1-\pi_i)}\right)\pi_i(1-\pi_i)X_{ij} = 0, \text{ for } j=1,\ldots,p.
\end{aligned}
$$

Using the zero-inflated beta-binomial model the null hypothesis to be tested is $H_0:\phi=0$. As in Section 4.1, after a lot of algebra it can be shown that the score test statistic obtained here is exactly the same as Z_7. The proof is omitted.

5.2. Testing for zero-inflation

Here we deal with the hypothesis $H_0:\gamma=0$ in the zero-inflated overdispersed binomial model. As in Section 4.2 , the score statistic poses problems so we derive the score statistic by using the zero-inflated beta-binomial model, which is

$$
Z_8 = \frac{\left(\sum_{i=1}^n\big(\frac{\prod_{r=0}^{m_i-1}(1+r\hat{\phi})}{\prod_{r=0}^{m_i-1}(1+r\hat{\phi}-\hat{\pi}_i)}I_{\{y_i=0\}} - 1\big)\right)^2}{\hat{V}_\gamma}.
$$

Here $\hat{\pi}_i = \pi_i(\sum X_{ij}\hat{\beta}_j)$ and $\hat{\beta}_j$ and $\hat{\phi}$ are the maximum likelihood estimate of β_j and ϕ under the beta-binomial model, and $\hat{V}_\gamma = V_\gamma(\hat{\pi}, \hat{\phi})$ with

$$V_\gamma = V_\gamma(\pi,\phi) = I^\gamma_{\gamma\gamma} - \mathbf{1}^T W^\gamma_3 U(U^T W^\gamma_1 U)^{-1} U^T W^\gamma_3 \mathbf{1}$$
$$-(I^\gamma_{\gamma\phi} - \mathbf{1}^T W^\gamma_3 U(U^T W^\gamma_1 U)^{-1} U^T W^\gamma_2 \mathbf{1})^2 (I^\gamma_{\phi\phi} - \mathbf{1}^T W^\gamma_2 U(U^T W^\gamma_1 U)^{-1} U^T W^\gamma_2 \mathbf{1})^{-1},$$

$$I^\gamma_{\gamma\gamma} = \sum_{i=1}^n \left(\frac{\prod_{r=0}^{m_i-1}(1+r\phi)}{\prod_{r=0}^{m_i-1}(1+r\phi-\pi_i)} - 1 \right), \qquad I^\gamma_{\gamma\phi} = \sum_{i=1}^n \sum_{r=1}^{m_i-1} \left(\frac{r}{1+r\phi-\pi_i} - \frac{r}{1+r\phi}\right),$$

$$I^\gamma_{\phi\phi} = \sum_{i=1}^n \sum_{r=1}^{m_i-1} \left\{ \frac{r^2 P(Y>r)}{(\pi_i + r\phi)} + \frac{r^2 P(Y<m_i-r)}{(1+r\phi-\pi_i)} + \frac{r^2}{(1+r\phi)} \right\},$$

$$W^\gamma_{1i} = \sum_{r=0}^{m_i-1} \left\{ \frac{P(Y>r)}{(\pi_i+r\phi)^2} + \frac{P(Y<m_i-r)}{(1+r\phi-\pi_i)} \right\},$$

$$W^\gamma_{2i} = \sum_{r=0}^{m_i-1} \left\{ \frac{rP(Y>r)}{(\pi_i+r\phi)^2} - \frac{rP(Y<m_i-r)}{(1+r\phi-\pi_i)^2} \right\}, \qquad W^\gamma_{3i} = \sum_{r=1}^{m_i-1} \frac{-1}{(1+r\phi-\pi_i)}.$$

5.3. Testing for over-dispersion and zero-inflation

The hypothesis to be tested is $H_0 : (\gamma, \phi) = 0$. The score statistic obtained using Theorem 3.3 is identical to the score statistic obtained using the over-dispersed beta-binomial model. Thus, the score statistic to test $H_0 : (\gamma, \phi) = 0$ is

$$Z_9 = \frac{\hat{V}_{TT}\hat{S}^2 + \hat{V}_{SS}\hat{T}^2 - 2\hat{V}_{ST}\hat{S}\hat{T}}{\hat{V}_{TT}\hat{V}_{SS} - \hat{V}^2_{ST}},$$

where

$$S(\pi) = \sum_{i=1}^n (I_{\{y_i=0\}}(1-\pi_i)^{-m_i} - 1),$$

$$T(\pi) = \sum_{i=1}^n \frac{(y_i - m_i\pi_i)^2 + \pi_i(y_i - m_i\pi_i) - y_i(1-\pi_i)}{2\pi_i(1-\pi_i)},$$

$$V_{SS}(\pi) = \sum_{i-1}^n ((1-\pi_i)^{-m_i} - 1), \qquad V_{ST}(\pi) = \sum_{i=1}^n \frac{\pi_i m_i(m_i-1)}{2(1-\pi_i)},$$

$$V_{TT}(\pi) = \sum_{i=1}^n \frac{1}{2} m_i(m_i-1) - \mathbf{1}^T W_3 U(U^T W_1 U)^{-1} U^T W_3 \mathbf{1},$$

with $W_{1i} = m_i/[\pi_i(1-\pi_i)]$, $W_{3i} = -m_i/(1-\pi_i)$, $\hat{S} = S(\hat{\pi}), \hat{T} = T(\hat{\pi}), \hat{V}_{SS} = V_{SS}(\hat{\pi}), \hat{V}_{TT} = V_{TT}(\hat{\pi}), \hat{V}_{ST} = V_{ST}(\hat{\pi})$ and $\hat{\pi}$ is the maximum likelihood estimate of the parameter $\pi = (\pi_1, \ldots, \pi_n)$ under the binomial model.

6. Simulations

A simulation study was conducted to examine the empirical size and power of the score statistics. The simulation study was limited only to examining performance of the score test statistics Z_1, Z_4 and Z_5 for Poisson data. Covariates were not considered. Each simulation experiment was based on 10,000 simulations.

6.1. Testing for over-dispersion in presence of zero-inflation in Poisson data

Samples of size $n = 20, 50, 100, 200$, were taken from the zero-inflated Poisson (μ, ω) distribution, where μ is the Poisson mean and ω is the zero-inflation parameter. We conducted simulations for values of $\mu = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$ and $\omega = 0.01, 0.10, 0.20, 0.30, 0.40, 0.50$. However, we present results for only $\mu = 2.0$ in Table 1. Conclusions for other values of μ are similar. Results in Table 1 show that, in general, the statistic Z_1 maintains nominal level well for all values of μ and ω considered. Only for small n, μ and α does one find some liberal behaviour.

Table 1. Empirical levels of score test statistic Z_1 for testing over-dispersion in presence of zero inflation; data are generated from $P(\mu,\omega)$ distribution with no covariates; based on 10,000 replications; $\mu = 2.0$

α	n	$\omega = 0.01$	$\omega = 0.10$	$\omega = 0.20$	$\omega = 0.30$	$\omega = 0.40$	$\omega = 0.50$
0.01	20	0.044	0.025	0.043	0.023	0.017	0.012
	50	0.019	0.014	0.013	0.012	0.010	0.011
	100	0.010	0.010	0.012	0.011	0.013	0.011
	200	0.010	0.010	0.011	0.012	0.010	0.010
0.05	20	0.061	0.045	0.060	0.038	0.030	0.024
	50	0.048	0.042	0.041	0.036	0.036	0.033
	100	0.045	0.044	0.047	0.042	0.042	0.041
	200	0.047	0.046	0.048	0.051	0.046	0.045
0.10	20	0.091	0.076	0.092	0.068	0.055	0.048
	50	0.090	0.086	0.085	0.077	0.078	0.071
	100	0.092	0.091	0.094	0.085	0.083	0.085
	200	0.098	0.096	0.093	0.094	0.094	0.092

A power study was conducted for $\mu = 2.0$, $\alpha = 0.05$, $c = 0.01, 0.05, 0.10, 0.15, 0.25, 0.35, 0.45, 0.55$, and for the zero-inflated parameter ω =0.05, 0.15, 0.35, 0.55. The data for the power analysis were generated from the zero-inflated $NB(\mu, c)$ distribution, and the results are summarized in Table 2. For fixed values of ω power increases as c increases. However, for fixed values of c, power decreases as ω increases.

Table 2. Empirical power of score test statistic Z_1 for testing over-dispersion in presence of zero-inflation; data are generated from zero-inflated $NB(\mu, c)$; based on 10,000 replications; $\mu = 2.0$; $\alpha = 0.05$.

ω	n	$c = 0.01$	$c = 0.05$	$c = 0.1$	$c = 0.15$	$c = 0.25$	$c = 0.35$	$c = 0.45$	$c = 0.55$
0.05	20	0.050	0.061	0.085	0.109	0.163	0.225	0.287	0.347
	50	0.047	0.076	0.120	0.183	0.317	0.454	0.564	0.657
	100	0.050	0.086	0.180	0.285	0.516	0.703	0.824	0.894
	200	0.055	0.123	0.282	0.479	0.780	0.927	0.975	0.993
0.15	20	0.059	0.073	0.086	0.110	0.165	0.215	0.271	0.327
	50	0.047	0.069	0.110	0.166	0.290	0.423	0.525	0.618
	100	0.049	0.085	0.165	0.271	0.487	0.666	0.784	0.868
	200	0.054	0.112	0.259	0.436	0.746	0.900	0.963	0.987
0.35	20	0.036	0.048	0.063	0.081	0.122	0.169	0.213	0.250
	50	0.040	0.065	0.102	0.139	0.244	0.352	0.437	0.510
	100	0.044	0.075	0.139	0.219	0.401	0.567	0.680	0.773
	200	0.053	0.103	0.220	0.366	0.642	0.822	0.912	0.958
0.55	20	0.026	0.034	0.049	0.060	0.088	0.121	0.150	0.177
	50	0.037	0.056	0.082	0.115	0.186	0.261	0.331	0.389
	100	0.048	0.072	0.124	0.178	0.316	0.446	0.544	0.638
	200	0.053	0.092	0.182	0.286	0.507	0.692	0.802	0.872

Table 3. Empirical levels of score test statistic Z_4 for testing zero-inflation in presence of over-dispersion; (a) data are simulated from $NB(\mu, c)$, (b) data are simulated from $LMP(\mu, c)$, with no covariates; based on 10,000 replications; $\mu = 2.0$.

		(a) $NB(\mu, c)$				(b) $LMP(\mu, c)$			
α	n	$c = 0.05$	$c = 0.10$	$c = 0.20$	$c = 0.50$	$c = 0.05$	$c = 0.10$	$c = 0.20$	$c = 0.50$
0.01	20	0.010	0.008	0.007	0.008	0.009	0.008	0.009	0.014
	50	0.008	0.009	0.008	0.009	0.010	0.010	0.010	0.021
	100	0.010	0.008	0.010	0.011	0.009	0.009	0.013	0.028
0.05	20	0.049	0.048	0.046	0.044	0.048	0.046	0.047	0.055
	50	0.046	0.048	0.044	0.047	0.046	0.051	0.052	0.072
	100	0.047	0.047	0.048	0.049	0.048	0.052	0.052	0.090
0.10	20	0.097	0.095	0.097	0.092	0.100	0.098	0.097	0.108
	50	0.096	0.098	0.092	0.096	0.097	0.101	0.103	0.128
	100	0.094	0.093	0.097	0.100	0.098	0.102	0.103	0.154

6.2. Testing for zero-inflation in presence of over-dispersion in Poisson data

Samples of size n = 20, 50, 100 were taken from negative binomial (μ,c) distribution. Simulations were conducted for values of $\mu = 0.5$, 1.0, 1.5, 2.0, 2.5, 3.0 and $c = 0.05$, 0.10, 0.20, 0.50. Here also we present results for only $\mu = 2.0$

as conclusions for other values of μ are similar. The results presented in Table 3 show that the statistic Z_4 maintains nominal level well.

Since the statistic Z_4 was derived using the zero-inflated negative binomial distribution, we conducted further simulations to examine robustness of this statistic when data are generated from a over-dispersed model other than the negative binomial distribution. So, we repeated the simulation by taking samples from the log-normal mixture of Poisson LMP(μ,c) distribution. The results of the simulations are also reported in Table 3. The level properties, in this case, remains the same.

A power study was conducted for $\mu = 2.0$, $\alpha = 0.05$, $c = 0.01$, 0.05, 0.15, 0.25 and the zero-inflation parameter $\omega = 0.01$, 0.05, 0.10, 0.15, 0.25, 0.35, 0.45, 0.55. The results are given in Table 4. Power increases as the zero-inflation parameter increases except for large values of ω in which case there seems to be some reversal.

Table 4. Empirical power of score test statistic Z_4 for testing zero-inflation in presence of over-dispersion; data are generated from zero-inflated $NB(\mu, c)$ distribution; based on 10,000 replications; $\mu = 2.0$; $\alpha = 0.05$.

c	n	$\omega = 0.01$	$\omega = 0.05$	$\omega = 0.10$	$\omega = 0.15$	$\omega = 0.25$	$\omega = 0.35$	$\omega = 0.45$	$\omega = 0.55$
0.01	20	0.046	0.053	0.072	0.086	0.133	0.157	0.193	0.251
	50	0.076	0.116	0.194	0.279	0.458	0.575	0.617	0.675
	100	0.211	0.322	0.454	0.609	0.804	0.888	0.876	0.828
	200	0.556	0.631	0.738	0.852	0.961	0.984	0.966	0.857
0.05	20	0.044	0.052	0.061	0.082	0.109	0.125	0.132	0.114
	50	0.046	0.062	0.112	0.171	0.283	0.352	0.343	0.288
	100	0.050	0.088	0.184	0.311	0.542	0.674	0.652	0.481
	200	0.055	0.138	0.352	0.576	0.858	0.946	0.889	0.590
0.15	20	0.044	0.047	0.054	0.063	0.095	0.108	0.118	0.116
	50	0.049	0.060	0.091	0.133	0.218	0.268	0.260	0.245
	100	0.056	0.087	0.155	0.244	0.429	0.540	0.490	0.353
	200	0.066	0.125	0.273	0.448	0.731	0.858	0.701	0.319
0.25	20	0.044	0.046	0.048	0.057	0.070	0.082	0.083	0.078
	50	0.048	0.056	0.082	0.104	0.166	0.207	0.184	0.154
	100	0.049	0.070	0.125	0.192	0.333	0.414	0.341	0.204
	200	0.051	0.098	0.207	0.356	0.620	0.744	0.486	0.161

6.3. Testing for over-dispersion and zero-inflation in Poisson data

Here, samples of size $n = 20$, 50, 100 were generated from Poisson (μ) distribution for $\mu = 1.0$, 2.0, 3.0. The results are presented in Table 5. The results show that the score test statistic Z_5 for testing simultaneously for zero-inflation and over-dispersion maintains nominal level reasonably well, although for small n there seems to be some conservative behaviour.

A power study was conducted for $\mu = 2.0$, $\alpha = 0.05$, and for some selected values of (ω, c). The results are given in Table 6. The statistic shows good power against departure from the Poisson model. In general, power increases as sample size increases. Power also increases as either or both ω and c increase.

Table 5. Empirical levels of score test statistic Z_5 for testing zero-inflation and over-dispersion; data are simulated from $P(\mu)$ with no covariates; based on 10,000 replications.

α	n	$\mu = 1.0$	$\mu = 2.0$	$\mu = 3.0$
0.01	20	0.015	0.013	0.014
	50	0.017	0.015	0.012
	100	0.015	0.011	0.011
0.05	20	0.040	0.040	0.039
	50	0.049	0.048	0.044
	100	0.051	0.049	0.046
0.10	20	0.071	0.075	0.072
	50	0.090	0.092	0.090
	100	0.094	0.096	0.096

Table 6. Empirical power of score test statistic Z_5 for testing zero-inflation and over-dispersion; data are generated from zero-inflated $NB(\mu, c)$ distribution; based on 10,000 replications; $\mu = 2.0$; $\alpha = 0.05$.

	(ω, c)									
n	(0.01, 0.05)	(0.05, 0.1)	(0.05, 0.15)	(0.05, 0.25)	(0.1, 0.1)	(0.1, 0.15)	(0.1, 0.25)	(0.15, 0.1)	(0.15, 0.15)	(0.15, 0.25)
20	0.074	0.161	0.213	0.345	0.218	0.282	0.419	0.300	0.376	0.497
50	0.098	0.277	0.397	0.627	0.431	0.551	0.745	0.604	0.706	0.841
100	0.135	0.472	0.644	0.876	0.712	0.827	0.955	0.878	0.936	0.985

7. Examples

Example 1. The data are from a prospective study of dental status of school-children from Bohning, Dietz and Schlattmann (1999). The children were all 7 years of age at the beginning of the study. Dental status were measured by the decayed, missing and filled teeth(DMFT) index. Only the eight deciduous molars were considered so the smallest possible value of the DMFT index is 0 and the largest is 8. The prospective study was for a period of two years. The DMFT index was calculated at the beginning and at the end of the study. The DMFT index data at the beginning of the study are: (index, frequency): (0,172), (1,73), (2,96), (3,80), (4,95), (5,83), (6,85), (7,65), (8,48). We now separately fit the Poisson model(P), over-dispersed Poisson model (negative binomial model)(NB), zero-inflated Poisson model(ZIP) and zero-inflated negative binomial model(ZINB) to the data. The values of the maximized log-likelihood under these four models are −1998.88, −1833.86, −1761.20 and −1756.80 respectively, which indicate that the zero-inflated negative binomial model fits the data far better than other models. Note that the likelihood ratio statistic for testing the fit of the zero-inflated Poisson model(ZIP) against the zero-inflated negative binomial model(ZINB)is $2(-1756.8 + 1761.2) = 8.8$ which is highly significant on the $\chi^2(1)$ scale.

We now use the score tests to select an appropriate model for the data. The values of the score test statistic for the goodness of fit of a model (such as the Poisson model) against another model (such as the negative binomial model) are given in Table 7. The tests overwhelmingly rejects the fit of all other models in favour of the zero-inflated negative binomial model.

Table 7. Results of the goodness of fit tests (score tests) for the DMFT index data.

Test	Score statistic with no covariate (p-value)	Score statistic with covariates (p-value)
P vs NB	394.44(0.00000)	354.51(0.00000)
P vs ZIP	847.20(0.00000)	709.89(0.00000)
ZIP vs ZINB	7.87(0.00503)	6.67(0.00978)
NB vs ZINB	45.51(0.00000)	41.87(0.00000)
P vs ZINB	890.17(0.00000)	761.75(0.00000)

To see how covariates affect goodness of fit of the models we expanded the above analysis to include the school variable. There were six treatments and six schools. The six treatments were randomized to the six schools, so that all children of a given school received the same treatment. Note in the previous analysis μ, c and ω were all common for all treatment groups, whereas now we have different μ's, but common c and ω. The values of the maximized log-likelihood under the four models P, NB, ZIP and ZINB are −1980.78, −1826.95, −1756.33 and −1752.50 respectively. The values of the score test statistics are given in the last column of Table 7. The conclusions here are essentially the same as those obtained earlier.

Example 2. The data given in Berry (1987) pertain to twelve patients who experienced frequent premature ventricular contractions (PVCs) and were administered a drug with antiarrhythmic properties. One-minute EKG recordings were taken before and after drug administration. The PVCs were counted on both recordings. The values of the maximized log-likelihood under binomial(B), beta-binomial(BB), zero-inflated binomial(ZIB) and zero-inflated beta-binomial(ZIBB) are −40.69, −19.18, −18.87 and −18.02 respectively. The values of the score test statistics are given in Table 8. The maximized log-likelihood values indicate that the zero-inflated beta-binomial model fits the data slightly better than the zero-inflated binomial model, and fits the data far better than other models. The values of the score test statistics and their corresponding p-values indicate that either the zero-inflated binomial or the beta-binomial model adequately fit the data. However, the value of the score test statistic for testing the fit of the zero-inflated binomial model against the zero-inflated beta-binomial model is smaller than that for testing the fit of the beta-binomial model against the zero-inflated beta-binomial model. Therefore, the zero-inflated binomial model is the model of choice for the data.

Table 8. Results of the goodness of fit tests (score tests) for the PVC counts data.

Test	Score statistic	p-value
B vs BB	236.4964	0.0000
B vs ZIB	931.0414	0.0000
ZIB vs ZIBB	1.2883	0.2564
BB vs ZIBB	1.6997	0.1613
B vs ZIBB	990.2721	0.0000

Acknowledgement

This research was partially supported by the Natural Science and Engineering Research Council of Canada. The editor and two referees made helpful suggestions which have improved the presentation of the paper.

Appendix A. Proofs

A.1. Information matrix

Partition the Fisher information matrix $I(\beta,\tau,\gamma)$ for the parameters β,τ and γ as

$$I(\beta,\tau,\gamma)=\begin{bmatrix} I_{\beta\beta} & I_{\beta\tau} & I_{\beta\gamma} \\ I_{\tau\beta} & I_{\tau\tau} & I_{\tau\gamma} \\ I_{\gamma\beta} & I_{\gamma\tau} & I_{\gamma\gamma} \end{bmatrix},$$

where $I_{\beta\beta}, I_{\gamma\gamma}, I_{\beta\gamma}, I_{\gamma\gamma}, I_{\gamma\tau}$ and $I_{\beta\tau}$ are $p\times p, 1\times 1, 1\times p, 1\times 1, 1\times 1$ and $1\times p$ matrices respectively and have usual meanings. To obtain the elements of the matrix $I(\beta,\tau,\gamma)$, we first obtain the first and the second partial derivatives of the log likelihood function $l_i(\gamma,\tau,\theta_i;y_i)$ with respect to the parameters γ,τ,θ_i.

$$\frac{\partial l_i}{\partial\gamma}=\frac{-1}{1+\gamma}+I_{\{y_i=0\}}\frac{1}{\gamma+f_0}, \qquad \frac{\partial l_i}{\partial\tau}=I_{\{y_i=0\}}\frac{f_0}{\gamma+f_0}\frac{\frac{\partial D}{\partial\tau}}{1+D}+I_{\{y_i>0\}}\frac{\frac{\partial D}{\partial\tau}}{1+D},$$

$$\frac{\partial l_i}{\partial\theta_i}=I_{\{y_i=0\}}\frac{f_0}{\gamma+f_0}\Big[-g'+\frac{\frac{\partial D}{\partial\theta_i}}{1+D}\Big]+I_{\{y_i>0\}}[a'y_i-g']+I_{\{y_i>0\}}\frac{\frac{\partial D}{\partial\theta_i}}{1+D},$$

$$\frac{\partial^2 l_i}{\partial \gamma^2} = \frac{1}{(1+\gamma)^2} - I_{\{y_i=0\}}\frac{1}{(\gamma+f_0)^2}, \qquad \frac{\partial^2 l_i}{\partial \tau \partial \gamma} = -I_{\{y_i=0\}}\frac{f_0}{(\gamma+f_0)^2}\frac{\frac{\partial D}{\partial \tau}}{1+D},$$

$$\frac{\partial^2 l_i}{\partial \theta_i^2} = \frac{-I_{\{y_i=0\}}}{(\gamma+f_0)^2}\Big[f_0(-g') + f_0\frac{\frac{\partial D}{\partial \theta_i}}{1+D}\Big]^2 + \frac{f_0 I_{\{y_i=0\}}}{\gamma+f_0}\Big[(-g' + \frac{\frac{\partial D}{\partial \theta_i}}{1+D})^2 - g''\Big]$$

$$+\frac{f_0 I_{\{y_i=0\}}}{\gamma+f_0}\Big[\frac{\frac{\partial^2 D}{\partial \theta_i^2}}{1+D} + \frac{-(\frac{\partial D}{\partial \theta_i})^2}{(1+D)^2}\Big] + I_{\{y_i>0\}}\Big[(a'' y_i - g'') + \frac{\frac{\partial^2 D}{\partial \theta_i^2}}{1+D} + \frac{-(\frac{\partial D}{\partial \theta_i})^2}{(1+D)^2}\Big],$$

$$\frac{\partial^2 l_i}{\partial \tau^2} = I_{\{y_i=0\}}\frac{-f_0^2}{(\gamma+f_0)^2}\Big(\frac{\frac{\partial D}{\partial \tau}}{1+D}\Big)^2 + I_{\{y_i=0\}}\frac{f_0}{\gamma+f_0}\frac{\frac{\partial^2 D}{\partial \tau^2}}{1+D}$$

$$+I_{\{y_i>0\}}\Big[\frac{(\frac{\partial D}{\partial \tau})^2}{(1+D)^2} + \frac{\frac{\partial^2 D}{\partial \tau^2}}{1+D}\Big],$$

$$\frac{\partial^2 l_i}{\partial \theta_i \partial \gamma} = -I_{\{y_i=0\}}\frac{f_0}{(\gamma+f_0)^2}\Big(-g' + \frac{\frac{\partial D}{\partial \theta_i}}{1+D}\Big),$$

$$\frac{\partial^2 l_i}{\partial \theta_i \partial \tau} = \frac{I_{\{y_i=0\}} f_0 \frac{\partial D}{\partial \tau}}{(\gamma+f_0)(1+D)}\Big(-g' + \frac{\frac{\partial D}{\partial \theta_i}}{1+D}\Big) + \frac{I_{\{y_i=0\}} f_0}{\gamma+f_0}\Big[\frac{\frac{\partial^2 D}{\partial \theta_i \partial \tau}}{1+D} + \frac{-\frac{\partial D}{\partial \tau}\frac{\partial D}{\partial \theta_i}}{(1+D)^2}\Big],$$

$$+\frac{-I_{\{y_i=0\}} f_0^2 \frac{\partial D}{\partial \tau}}{(1+D)(\gamma+f_0)^2}\Big[-g' + \frac{\frac{\partial D}{\partial \theta_i}}{1+D}\Big] + I_{\{y_i>0\}}\Big[\frac{\frac{\partial^2 D}{\partial \theta_i \partial \tau}}{1+D} + \frac{-\frac{\partial D}{\partial \tau}\frac{\partial D}{\partial \theta_i}}{(1+D)^2}\Big].$$

A.2. Proof of Theorems

We give the proof of Theorem 3.1. Let U be an $n \times p$ matrix with ir-element $\partial\theta_i/\partial\beta_r$, $\mathbf{1}$ an $n \times 1$ unit vector, W_1^τ, W_2^τ and W_3^τ be diagonal matrices with ith element W_{1i}^τ, W_{2i}^τ and W_{3i}^τ respectively. Then, using the results in Section A.1, we obtain

$$S_i^\tau(\theta_i,\gamma) = \frac{\partial l_i}{\partial \tau}|_{\tau=0} = -\frac{\gamma I_{\{y_i=0\}}}{\gamma+f_0} + \frac{1}{2}bD_2,$$

$$W_{1i}^\tau = E\{-\frac{\partial^2 l_i}{\partial \theta_i^2}\}|_{\tau=0} = g'' - a'' E y_i - \frac{f_0\gamma}{(\gamma+f_0)(1+\gamma)}g'^2 - \frac{\gamma}{1+\gamma}g'',$$

$$W_{2i}^\tau = E\{-\frac{\partial^2 l_i}{\partial \theta_i \partial \gamma}\}|_{\tau=0} = -\frac{f_0}{(\gamma+f_0)(1+\gamma)}g',$$

$$W_{3i}^\tau = E\{-\frac{\partial^2 l_i}{\partial \theta_i \partial \tau}\}|_{\tau=0} = [\frac{1}{2}g' bD_2\frac{f_0\gamma}{(\gamma+f_0)(1+\gamma)} + \frac{1}{2}(bD_2)'\frac{\gamma}{1+\gamma}]|_{y_i=0} - \frac{1}{2}E[(bD_2)'],$$

$$I_{\tau\tau}^\tau = \sum[\frac{1}{4}E(bD_2)^2 - \frac{(\gamma^2+2f_0\gamma)}{(\gamma+f_0)(1+\gamma)}(\frac{1}{2}bD_2)^2|_{y_i=0}],$$

$$I_{\gamma\gamma}^\tau = \sum\frac{1-f_0}{(1+\gamma)^2(\gamma+f_0)}|_{y_i=0}, \qquad I_{\gamma\tau}^\tau = \sum\frac{\frac{1}{2}bD_2 f_0}{(\gamma+f_0)(1+\gamma)}|_{y_i=0}.$$

From the Fisher information matrix, the asymptotic variance of the likelihood score $S_\tau = \sum_{i=1}^n((1/2)bD_2 - (\gamma I_{\{y_i=0\}}/(\gamma+f_0)))$ is

$$V_\tau = I_{\tau\tau}^\tau - (\mathbf{1}^T W_3^\tau U, I_{\tau\gamma}^\tau)\begin{pmatrix} U^T W_1^\tau U & U^T W_2^\tau \mathbf{1} \\ \mathbf{1}^T W_2^\tau U & I_{\gamma\gamma}^\tau \end{pmatrix}^{-1}\begin{pmatrix} U^T W_3^\tau \mathbf{1} \\ I_{\gamma\tau}^\tau \end{pmatrix}.$$

Then, the score test statistics for testing H_0: $\tau=0$ is $\hat{S}_\tau^2/\hat{V}_\tau$, where $\hat{S}_\tau=S_\tau(\hat{\theta}_1,\ldots,\hat{\theta}_n;\hat{\gamma})$, $\hat{V}_\tau = V_\tau(\hat{\theta}_1,\ldots,\hat{\theta}_n;\hat{\gamma})$ and $\hat{\theta}_i$ and $\hat{\gamma}$ are the maximum likelihood estimates of θ_i and γ under the zero-inflated generalized linear model, which, under mild regularity conditions, has, asymptotically, a χ^2 distribution with one degree of freedom.

Proofs of Theorem 3.2 and Theorem 3.3 are omitted as these follow similar steps as those above.

References

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

Bartoo, J. B. and Puri, P. S. (1967). On optimal asymptotic tests of composite statistical hypothesis. *Ann. Math. Statist.* **38**, 1845-52.

Berry, D. A. (1987). logarithmic transformations in ANOVA. *Biometrics* **43**, 439-56.

Bohning, D., Dietz, E. and Schlattmann, P. (1999). The zero-inflation Poisson and the decayed, missing and filled teeth index in dental epidemiology. *J. Roy. Statist. Soc. Ser. A* **162**, 195-209.

Breslow, N. (1990). Tests of hypotheses in over-dispersed Poisson regression and other quasi-likelihood models. *J. Amer. Statist. Assoc.* **85**, 565-571.

Broek, V. J. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics* **51**, 738-743.

Chesher, A. (1984). Testing of neglected heterogeneity. *Econometrika* **52**, 865-872.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika* **70**, 269-274.

Dean, C. B. (1992). Testing for overdispersion in Poisson and binomial regression models, *J. Amer. Statist. Assoc.* **87**, 451-457.

Deng, D. and Paul, S. R. (2000). Score tests for zero-inflation in generalized linear models. *Canad. J. Statist.* **27**, 563-570.

Lindsay, B. G. and Roeder, K. (1992). Residual diagnostics for mixture models. *J. Amer. Statist. Assoc.* **87**, 785-794.

Neyman, J. (1959). Optimal asymptotic tests for composite hypothesis. In *Probability and Statistics: The Harold Cramer Volume* (Edited by U. Grenander), Wiley, New York.

Neyman, J. and Scott, E. L. (1966). On the use of C(α) Optimal tests of composite hypotheses. *Bull. Inst. Internat. Statist.* **41**, 477-497.

Paul, S. R. and Banerjee, T. (1998). Analysis of two-way layout of count data involving multiple counts in each cell. *J. Amer. Statist. Assoc.* **93**, 1419-1429.

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

Ridout, M., Hinde, J. and Demetrio, C. G. B. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics* **57**, 219-223.

Thall, P. F. (1992). Score tests in the two-way layout of counts. *Comm. Statist. Theory Methods* **21**, 3017-3036.

Department of Mathematics and Statistics, University of Regina, Sask., Canada.

E-mail: deng@math.uregina.ca

Department of Mathematics and Statistics, University of Windsor, Ont., Canada.

E-mail: smjp@uwindsor.ca

(Received November 2002; accepted March 2004)