Statistica Sinica **20** (2010), 323-341

TESTING FOR ZERO-MODIFICATION IN COUNT REGRESSION MODELS

Aleksey Min and Claudia Czado

Technische Universität München, Germany

Abstract: Count data often exhibit overdispersion and/or require an adjustment for zero outcomes with respect to a Poisson model. Zero-modified Poisson (ZMP) and zero-modified generalized Poisson (ZMGP) regression models are useful model classes for such data. In the literature so far only score tests are used for testing the necessity of this adjustment. We address this problem by using Wald and likelihood ratio tests. We show how poor the performance of the score tests can be in comparison to the performance of Wald and likelihood ratio (LR) tests through a simulation study. In particular, the score test in the ZMP case results in a power loss of 47% compared to the Wald test in the worst case, while in the ZMGP case the worst loss is 87%. Therefore, regardless of the computational advantage of score tests, the loss in power compared to the Wald and LR tests should not be neglected and these much more powerful alternatives should be used instead. We prove consistency and asymptotic normality of the maximum likelihood estimates in ZGMP regression models, which form the basics of Wald and likelihood ratio tests. The usefulness of ZGMP models is illustrated in a data example.

Key words and phrases: Generalized Poisson distribution, likelihood ratio test, maximum likelihood estimate, overdispersion, score test, Wald test, zero-inflation, zero-modification.

1. Introduction

Poisson regression models are plain-vanilla models for count data. However they are often too simple to capture such complex structures of count data as overdispersion. Overdispersion is present if the count regression data has a higher variability than it is allowed by the model. In particular, the equality of mean and variance for the count data analyzed under Poisson assumption is often violated. Various sources such as missing covariates, correlation among measurements, and excess of zero-outcomes with respect to standard Poisson regression models make counts overdispersed. Excellent surveys on overdispersion and its treatment can be found in Cameron and Trivedi (1998, Chap. 4) and Winkelmann (2008, Chap. 3 and 4).

Since positive counts may still be overdispersed with respect to the zero-truncated Poisson distribution, in the last decade zero-inflated generalized Poisson (ZIGP) regression models have been found useful for the analysis of count data with a large amount of zero-outcomes (see e.g., Famoye and Singh (2003), Gupta, Gupta and Tripathi (2004), Joe and Zhu (2005), Bae, Famoye, Wulu, Bartolucci and Singh (2005), and Famoye and Singh (2006)). It is a large class of regression models that contains zero-inflated Poisson (ZIP), generalized Poisson (GP), and Poisson regressions (see Mullahy (1986), Lambert (1992), Consul and Famoye (1992), and Famoye (1993)). Recently Czado, Erhardt, Min and Wagner (2007) introduced flexible ZIGP models with regression effects on the mean, dispersion, and zero-inflation (ZI) level, and applied them to patent outsourcing rates. Their findings showed the superiority with regard to data fit of their ZIGP models over Poisson, GP, ZIP, and even over ZIGP with constant overdispersion and/or constant ZI based on the Vuong's test (see Vuong (1989)).

Score tests are widely used for testing misspecifications in count regression models because they require one to fit the model only under the null hypothesis. For regression models with constant ZI and/or constant overdispersion they have been developed by Dean and Lawless (1989), Dean (1992), van den Broek (1995), Deng and Paul (2000, 2005), Ridout, Hinde and Demétrio (2001), Gupta et al. (2004). Considering only score tests for ZI, we observe that there is a confusion with regard to the limiting distribution of score test statistics or the interpretation of rejecting the null hypothesis in the literature. To discuss this point in more detail, first note that zero-inflated count regression models are based on mixtures of the Bernoulli distribution and a count distribution, i.e., so-called zero-inflated count distributions. Now the ZI parameter ω is interpreted as the probability of obtaining a zero-outcome from the Bernoulli distribution. However zero-inflated distributions are also well-defined for small negative values of the ZI parameter ω, which indicates that the probability of zero outcomes is smaller than the probability of zero outcomes for the count distribution used in the mixture. The negative lower bound for ω is derived from the necessity that the probability of zero-outcome for zero-inflated distributions is nonnegative (see e.g., van den Broek (1995)). Thus, negative values of the parameter ω are acceptable and correspond to zero-deflation. Further, score tests do not require an estimation of the ZI parameter and therefore the score tests for ZI have, in fact, a two-sided alternative hypothesis, i.e., the null and alternative hypotheses are given by

$$H_0 : \omega = 0 \quad \text{versus} \quad H_1 : \omega \neq 0. \tag{1.1}$$

The null hypothesis might now be rejected in favor of zero-modification (ZM), i.e., ZI or zero-deflation and not of only ZI. Therefore score tests for ZI known in the literature are mostly score tests for ZM. Zero-modified Poisson (ZMP) regression models have explicitly been introduced by Dietz and Böhning (2000) and we extend these models to more general count regression models. In order to derive a score test only for ZI, the problem of testing parameters on the boundary of the parameter space needs to be addressed. Consequently, the limiting distribution of the score statistic differs from a standard χ^2-distribution with one degree of freedom and should be corrected, according to Silvapulle and Silvapulle (1995). Under regularity assumptions they have shown, for tests with one sided alternative, that likelihood ratio and score test statistics have the same limiting distribution. One crucial point of their assumptions is that a score vector is well-defined in a small neighborhood of the null hypothesis H_0. We will see that this requirement is satisfied for $H_0 : \omega = 0$. For insightful discussions on this problem, we refer to Verbeke and Molenberghs (2003).

Nowadays, given modern computing power, the computational advantage of score tests has lost some of its original attraction. We think more attention should be paid to Wald and likelihood ratio (LR) tests for ZM though they have not been utilized for the testing problem (1.1) so far. In addition to numerical difficulties related to estimation of the ZM parameter ω, these tests also require the knowledge of the asymptotic distribution of the maximum likelihood estimates (MLE's) including the Fisher information matrix. We show that these additional theoretical and numerical efforts related to Wald and LR tests bring a gain in test power.

In this paper we introduce zero-modified generalized Poisson (ZMGP) regression models and consider the testing problem (1.1) for them. We show that MLE's in ZMGP regression models, on which the Wald and LR tests are based, are consistent and asymptotically normal. We investigate the performance of Wald, LR, and score tests for testing ZM. It should be noted that our theoretical results remain valid for GP and ZMP regression models subject to appropriate changes in assumptions.

There exists an alternative count regression model for overdispersed and zero-inflated data: a zero-inflated negative binomial (ZINB) regression (see Ridout et al. (2001) and Hall and Berenhaut (2002)). It is not a subject of this paper to compare ZIGP and ZINB models but we list most important differences, from our point of view, between these regression models.

The paper is organized as follows. In Section 2 we introduce the GP distribution and discuss its basic forms and properties. A ZMGP regression model with constant ZM and overdispersion is defined in Section 3. Section 4 gives the asymptotic existence, the consistency and the asymptotic normality of the MLE in a ZMGP regression model. In contrast to Czado et al. (2007), we provide a rigorous detailed proof of our asymptotic results contained in the appendices of the accompanying on-line supplement. In Section 5 we compare the performance of the score test, for detecting ZM in ZMP and ZMGP models, to the performance of the Wald and LR tests in a simulation study. In particular it is discovered that, using the score test, one may lose in power compared to the Wald test, up to 47% for the ZMP case and up to 87% for the ZMGP case. We also show that the score test for ZM in the analysis of the apple propagation data (see Ridout and Demétrio (1992)) does not always detect ZM, while the Wald and LR tests give strong evidence for ZM. The paper closes with a conclusion and discussion section. For brevity the Fisher information matrix of the ZMGP regression and the proof of Theorem 1 are given in the appendices of an on-line supplement at `http://www.stat.sinica.edu.tw/statistica`.

2. The GP Distribution

A random variable $\tilde{Y}$ has a GP distribution with parameters μ and φ, which we denote by $GP(\mu,\varphi)$, if its probability mass function is

$$P_{\mu,\varphi}(y) := \begin{cases} \dfrac{\mu(\mu+y(\varphi-1))^{y-1}\varphi^{-y}e^{-(\mu+y(\varphi-1))/\varphi}}{y!} & \text{for } y=0,1,\ldots \\ 0 & \text{for } y>m, \quad \text{when } \varphi<1. \end{cases} \tag{2.1}$$

The real-valued parameters μ and φ satisfy the two constraints:

(i) $\mu > 0$;

(ii) $\varphi \geq \max\{1/2, 1-\mu/m\}$, where m $(m \geq 4)$ is the largest natural number such that $\mu + m(\varphi-1) > 0$ when $\varphi < 1$.

If $\varphi < 1$, then (2.1) does not correspond to a probability distribution. However the lower limit, imposed on φ in this case, guarantees that the total error of truncation is less than 0.5% (see Consul and Shoukri (1985)). Since all discrete distributions are truncated under sampling procedures, this is a reasonable condition.

The GP distribution was first introduced by Consul and Jain (1970) and subsequently studied in detail by Consul (1989). One particular property of the GP distribution is that the variance is greater than, equal to, or less than the mean according to whether the second parameter φ is greater than, equal to, or less than 1. More precisely (for details see Consul (1989, p. 12)), if $\tilde{Y} \sim GP(\mu,\varphi)$ then the mean and variance of Y are

$$E(\tilde{Y}) = \mu \quad \text{and} \quad Var(\tilde{Y}) = \varphi^2\mu. \tag{2.2}$$

A NB distribution with mean μ and overdispersion parameter $a > 0$ (see Lawless (1987) for precise definition) also has a flexible variance function, $\mu(1 + a\mu)$. Thus the overdispersion in the GP case is independent of the mean while this is not the case for the NB distribution. This implies that overdispersion in the NB case might be present over and above that accounted for by a. This point was first discussed by Lawless (1987). Czado and Sikora (2002) also noted this and developed an approach based on $p-$value-curves to quantify overdispersion effects more precisely. Another significant difference between these two distributions is that the NB distribution belongs to the exponential family if the overdispersion parameter a is known, while this does not hold for the GP distribution. A comparison of GP and NB probability functions can be found in Joe and Zhu (2005) and Gschlößl and Czado (2008).

There is a form of the GP distribution obtained by assuming that $\varphi - 1 = \alpha\mu$ for $\alpha > 0$. In the literature it is known as a restricted generalized Poisson (RGP) distribution (see Consul (1989, p. 5)) and the relation between its mean and variance is given by $Var(\tilde{Y}) = (1 + \alpha E(\tilde{Y}))^2 E(\tilde{Y})$. Thus overdispersion in the RGP case is not independent of the mean. We deal here only with an unrestricted form (2.1) of the GP distribution.

3. ZMGP Regression

A ZMGP distribution is defined analogous to a ZMP distribution (see Dietz and Böhning (2000)) and its probability mass function is

$$P_{\mu,\varphi,\omega}(y) := P(Y = y) = \begin{cases} \omega + (1-\omega)P(\tilde{Y} = 0) & y = 0, \\ (1-\omega)P(\tilde{Y} = y) & y = 1, \dots, \end{cases} \tag{3.1}$$

where $\tilde{Y}$ is distributed according to the GP distribution with parameters φ and μ, and the parameter ω satisfies

$$\frac{-\exp(-\mu/\varphi)}{1 - \exp(-\mu/\varphi)} \leq \omega \leq 1. \tag{3.2}$$

Thus, this distribution has 3 parameters μ, φ and ω, and is further denoted by $ZMGP(\mu, \varphi, \omega)$. Condition (3.2) ensures that (3.1) defines a probability mass function for negative values of ω corresponding to zero-deflation. Positive values of the parameter ω correspond to ZI which mostly occurs in practice.

A simple calculation using (2.2) implies that the mean and variance of the ZMGP distribution are

$$E(Y) = (1-\omega)\mu \quad \text{and} \quad Var(Y) = E(Y)(\varphi^2 + \mu\omega). \tag{3.3}$$

One of the main benefits of considering a regression model based on the ZMGP distribution is that it gives a large class of regression models for count response data. In particular, it reduces to Poisson regression when $\varphi = 1$ and $\omega = 0$, to GP regression when $\omega = 0$, and to ZMP regression when $\varphi = 1$. Moreover, by virtue of (3.3), this regression can be used to fit zero-modified count regression data exhibiting overdispersion or underdispersion.

Analogous to the generalized linear models (GLM) framework, we now introduce a regression model with response Y_i and (known) explanatory variables $\mathbf{x}_i = (x_{i0}, \dots, x_{ip})^t$ with $x_{i0} = 1$ for $i = 1, \dots, n$.

1. *Random components:* $\{Y_i, 1 \le i \le n\}$ are independent with $Y_i \sim ZMGP(\mu_i, \varphi, \omega)$.
2. *Systematic components:* The linear predictors $\eta_i(\boldsymbol{\beta}) = \mathbf{x}_i^t \boldsymbol{\beta}$ for $i = 1, \ldots, n$, influence the response Y_i. Here $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^t$ is a vector of unknown regression parameters. The matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^t$ is called the design matrix.
3. *Parametric link components:* The linear predictors $\eta_i(\boldsymbol{\beta})$ are related to the parameter μ_i of Y_i by $\mu_i = \exp(\eta_i(\boldsymbol{\beta}))$ for $i = 1, \ldots, n$.

Here $\mathbf{A}^t$ and $\mathbf{a}^t$ denote the transpose of a matrix $\mathbf{A}$ and a vector $\mathbf{a}$, respectively. We call this the ZMGP regression model. The parameters φ and ω are fixed and (3.2) holds for all μ_i, $i = 1, \ldots, n$. We denote the joint vector of the regression parameters $\boldsymbol{\beta}$ and the parameters φ and ω of the ZMGP distribution by $\boldsymbol{\delta}$ and its MLE by $\hat{\boldsymbol{\delta}}$.

The following abbreviations, $i = 1, \ldots, n$, are used: $\mu_i(\boldsymbol{\beta}) := \exp(\mathbf{x}_i^t \boldsymbol{\beta})$, $f_i(\boldsymbol{\beta}, \varphi) := \exp(-\mu_i(\boldsymbol{\beta})/\varphi)$, $g_i(\boldsymbol{\delta}) := \omega + (1-\omega) f_i(\boldsymbol{\beta}, \varphi) = P_{\mu_i(\beta), \varphi, \omega}(0)$. For observations $y_1, \ldots, y_n$, the log-likelihood $l(\boldsymbol{\delta})$ derived from the ZMGP regression can be written as

$$l_n(\boldsymbol{\delta}) = \sum_{i=1}^n \mathbb{1}_{\{y_i=0\}} \log(g_i(\boldsymbol{\delta})) + \sum_{i=1}^n \mathbb{1}_{\{y_i>0\}} \Bigg(\log(1-\omega) + \mathbf{x}_i^t \boldsymbol{\beta} - \frac{1}{\varphi} \mu_i(\boldsymbol{\beta})$$
$$+ (y_i - 1) \log \Big[\mu_i(\boldsymbol{\beta}) + y_i(\varphi - 1) \Big] - y_i \log \varphi - y_i \frac{1}{\varphi} (\varphi - 1) - \log(y_i!) \Bigg).$$

The score vector can be written as $\mathbf{s}_n(\boldsymbol{\delta}) = (s_0(\boldsymbol{\delta}), \ldots, s_p(\boldsymbol{\delta}), s_{p+1}(\boldsymbol{\delta}), s_{p+2}(\boldsymbol{\delta}))^t$, where

$$s_r(\boldsymbol{\delta}) := \frac{\partial l_n(\boldsymbol{\delta})}{\partial \beta_r} = \sum_{i=1}^n s_{r,i}(\boldsymbol{\delta}),$$
$$s_{r,i}(\boldsymbol{\delta}) := -x_{ir} \Bigg[\mathbb{1}_{\{y_i=0\}} \frac{(1-\omega) f_i(\boldsymbol{\beta}, \varphi) \mu_i(\boldsymbol{\beta})}{\varphi g_i(\boldsymbol{\delta})} - \mathbb{1}_{\{y_i>0\}} \Bigg(1 + \frac{\mu_i(\boldsymbol{\beta})(y_i - 1)}{\mu_i(\boldsymbol{\beta}) + (\varphi - 1) y_i} - \frac{\mu_i(\boldsymbol{\beta})}{\varphi} \Bigg) \Bigg],$$

for $r = 0, \ldots, p$ and $i = 1, \ldots, n$;

$$s_{p+1}(\boldsymbol{\delta}) := \frac{\partial l_n(\boldsymbol{\delta})}{\partial \varphi} = \sum_{i=1}^n s_{p+1,i}(\boldsymbol{\delta}),$$

$$s_{p+1,i}(\boldsymbol{\delta}) := \mathbb{1}_{\{y_i=0\}} \frac{(1-\omega) f_i(\boldsymbol{\beta}, \varphi) \mu_i(\boldsymbol{\beta})}{\varphi^2 g_i(\boldsymbol{\delta})}$$
$$+ \mathbb{1}_{\{y_i>0\}} \left(\frac{y_i(y_i-1)}{\mu_i(\boldsymbol{\beta}) + (\varphi-1) y_i} - \frac{y_i}{\varphi} + \frac{\mu_i(\boldsymbol{\beta}) - y_i}{\varphi^2} \right);$$
$$s_{p+2}(\boldsymbol{\delta}) := \frac{\partial l_n(\boldsymbol{\delta})}{\partial \omega} = \sum_{i=1}^{n} s_{p+2,i}(\boldsymbol{\delta}),$$
$$s_{p+2,i}(\boldsymbol{\delta}) := \mathbb{1}_{\{y_i=0\}} \frac{1 - f_i(\boldsymbol{\beta}, \varphi)}{g_i(\boldsymbol{\delta})} - \mathbb{1}_{\{y_i>0\}} \frac{1}{1-\omega} \quad \text{for} \quad i = 1, \ldots, n.$$

The score vector $\mathbf{s}_n(\boldsymbol{\delta})$ is well-defined in a small neighborhood of $\omega = 0$. This indicates that, for testing against ZI, the general theory of Silvapulle and Silvapulle (1995) on one-sided score tests is applicable and, therefore, the limiting distribution of the corresponding score statistic differs from a standard χ^2-distribution with one degree of freedom.

4. Asymptotic Theory

Fahrmeir and Kaufmann (1985) proved consistency and asymptotic normality of the MLE in GLM's for canonical as well as noncanonical link functions, under mild assumptions. In fact, they presented a general tool for deriving an asymptotic distribution of MLE's in any regression model. The validity of their general assumptions can easily be verified in GLM's with compact regressors, stochastic regressors, and bounded responses, respectively. However, it is not easy to derive the asymptotic distribution of the MLE in a regression model under such easily verified assumptions. We show in the on-line supplementary version of the paper that the MLE in ZMGP regression models with compact regressors possesses similar asymptotic properties as in GLM's with compact regressors.

As in Fahrmeir and Kaufmann (1985), we use the Cholesky square root matrix for normalizing the MLE. The left Cholesky square root matrix $\mathbf{A}^{1/2}$ of a positive definite matrix $\mathbf{A}$ is the unique lower triangular matrix with positive diagonal elements such that $\mathbf{A}^{1/2}(\mathbf{A}^{1/2})^t = \mathbf{A}$ (see Stewart (1998, p. 188)). For convenience, set $\mathbf{A}^{t/2} := (\mathbf{A}^{1/2})^t$, $\mathbf{A}^{-1/2} := (\mathbf{A}^{1/2})^{-1}$ and $\mathbf{A}^{-t/2} := (\mathbf{A}^{t/2})^{-1}$. We deal only with the spectral norm of square matrices $\mathbf{A}$ is given by $\|\mathbf{A}\| :=$ (maximum eigenvalue of $\mathbf{A}^t\mathbf{A})^{1/2} = \sup_{\|\mathbf{u}\|_2=1} \|\mathbf{A}\mathbf{u}\|_2$, where $\|\cdot\|_2$ denotes the L^2- norm of vectors. We drop the subindex 2 in $\|\cdot\|_2$ since the spectral norm is generated by the L^2–norm of vectors, and arguments of considered norms are always clearly defined. The minimal eigenvalue of a square matrix $\mathbf{A}$ is further denoted by $\lambda_{min}(\mathbf{A})$, and the vector of true parameter values of the ZMGP regression is denoted as $\boldsymbol{\delta}_0$. Further, $\mathbf{F}_n(\boldsymbol{\delta})$ is used for the Fisher information matrix in a ZMGP regression evaluated at $\boldsymbol{\delta}$. We note that the entries of the Fisher information matrix in a ZMGP regression have a closed form (see Appendix 1 of the on-line supplement), while this is not the case in regression models associated with a NB distribution (see e.g., Lawless (1987)).

Now for $\varepsilon > 0$ denote a neighborhood of $\boldsymbol{\delta}_0$ by

$$N_n(\varepsilon) = \Big\{\boldsymbol{\delta} : \|\mathbf{F}_n^{t/2}(\boldsymbol{\delta}_0)(\boldsymbol{\delta} - \boldsymbol{\delta}_0)\| \le \varepsilon\Big\}. \tag{4.1}$$

For convenience, we drop the arguments $\boldsymbol{\delta}_0$, $\boldsymbol{\beta}_0$, and φ_0, as well as the subindex $\boldsymbol{\delta}_0$ in $\mu_i(\boldsymbol{\beta}_0)$, $f_i(\boldsymbol{\beta}_0, \varphi_0)$, $g_i(\boldsymbol{\delta}_0)$, P_{δ_0}, E_{δ_0} etc., and write μ_i, f_i, g_i, P, E etc. Constants are further denoted by C and c, with subindices or without them; they may depend on $\boldsymbol{\delta}_0$ but not on n. The same C's and c's in different places denote different constants. Finally, the d-dimensional unit matrix is denoted by $\mathbf{I}_d$, and an admissible set for a vector $\boldsymbol{\beta}$ of regression parameters is denoted by B. We make the following assumptions.

(A1) $n/\lambda_{min}(\mathbf{F}_n) \le C_1$ for all $n \ge 1$, where C_1 is a positive constant.

(A2) $\{\mathbf{x}_i, i \ge 1\} \subset K_x$, where $K_x \subset \mathbb{R}^{p+1}$ is a compact set.

(A3) $B \subset \mathbb{R}^{p+1}$ is an open set and $\boldsymbol{\delta}_0$ is an interior point of the set $K_\delta := B \times \Phi \times \Omega$, where $\Phi := (1, \infty)$ and $\Omega := (-c_\omega, 1)$. Here c_ω is a positive constant such that (3.2) holds for all $\mathbf{x} \in K_x$, $\boldsymbol{\beta} \in B$, and $\varphi \in \Phi$.

Our main result states that results of Theorem 4 of Fahrmeir and Kaufmann (1985) can be extended to ZMGP regression models.

Theorem 1. *Under* (A1)$-$(A3), *there exists a sequence of random variables* $\hat{\boldsymbol{\delta}}_n$ *such that*

(i) $P(\mathbf{s}_n(\hat{\boldsymbol{\delta}}_n) = 0) \to 1$ *as* $n \to \infty$ (*asymptotic existence*),

(ii) $\hat{\boldsymbol{\delta}}_n \xrightarrow{P} \boldsymbol{\delta}_0$ *as* $n \to \infty$ (*weak consistency*),

(iii) $\mathbf{F}_n^{t/2}(\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) \stackrel{\mathcal{D}}{\Longrightarrow} N_p(\mathbf{0}, \mathbf{I}_{p+3})$ *as* $n \to \infty$ (*asymptotic normality*).

The proof of Theorem 1 is based on several auxiliary lemmas. The lemmas, their proofs, as well as the proof of Theorem 1, are given in Appendices 2 and 3 of the accompanying on-line supplement at `http://www.stat.sinica.edu.tw/statistica`.

Remarks

(i) Assumption (A1) is more restrictive than the corresponding condition (D) of Fahrmeir and Kaufmann (1985). This is the price we have paid for deriving the asymptotic theory for ZMGP regression models. Assumption (A2) means that we deal with compact regressors.

(ii) If $\boldsymbol{\delta}_0$ lies on the boundary of the parameter space K_δ, i.e., (A3) is violated, then Theorem 1 does not hold anymore. This implies that we cannot test the adequacy of the GP regression. However, the asymptotic results of Theorem 1 remain valid in GP or ZMP regression models subject to appropriate changes to the log-likelihood, the score equations, and the Fisher information matrix, as well as in Assumption (A3).

(iii) We would like to especially note that $\omega = 0$ is not on the boundary of the parameter space in ZMGP and ZMP regression models, thus allowing a direct application of Wald, LR, and score tests.

5. Applications

5.1. Power comparison of score, Wald, and LR tests in ZMP and ZMGP models

Jansakul and Hinde (2002) investigated the performance of the score test for ZI in small and moderate sample sizes within the ZIP regression model. They noted that their score test compares the Poisson model to the ZMP model, thus avoiding the problem of testing on the boundary of ZI.

By virtue of Remarks (ii) and (iii), we can construct the Wald and LR tests for testing ZM in ZMP models and then compare their performance with the performance of the score test. Note this comparison is only feasible for models with a constant ZM parameter. In particular, Jansakul and Hinde (2002) considered models with $\omega = 0, 0.25, 0.45$, linear predictors $\eta_i(\boldsymbol{\beta}) = 0.25, 0.75$, and $\eta_i(\boldsymbol{\beta}) = 0.75 - 1.45x_i$ for $i = 1, \ldots, n$ and $n = 50, 100, 200$. Covariates x_i's were taken uniformly from $(0, 1)$. For each combination of sample size and model, they simulated 1,000 sets of responses from the working model. The simulation setup for the constant linear predictors η_i's implies that the corresponding Poisson distribution has approximately 28% ($\eta_i(\boldsymbol{\beta}) = 0.25$) and 12% ($\eta_i(\boldsymbol{\beta}) = 0.75$) zero responses. In the case of nonconstant linear predictors, the probability of obtaining zero outcomes from the Poisson distribution with parameter $\exp(\eta_i(\boldsymbol{\beta}))$ varies between 0.12 and 0.61 for $i = 1, \ldots, n$. We used their simulation setup to compare the performance of the three above mentioned tests in S-PLUS 7.0 on a Windows platform. The MLE's were determined with a help of the S-PLUS function ``‘‘nlminb’’`` which finds the minimum of a smooth nonlinear function subject to bound-constrained parameters.

The Wald statistic for testing $H_0 : \omega = 0$ versus $H_1 : \omega \neq 0$ is $W_\omega = \hat{\omega}^2/\hat{\sigma}^2_\omega$, where $\hat{\omega}$ is the MLE of ω in a ZMP regression and $\hat{\sigma}^2_\omega$ is the estimated variance of $\hat{\omega}$, the corresponding diagonal element of the inverse of the Fisher information matrix evaluated at $(\hat{\omega}, \hat{\boldsymbol{\beta}})$. The LR statistic for the same testing problem is $LR_\omega = -2(l_n^{\mathrm{P}}(\hat{\boldsymbol{\beta}}^{\mathrm{P}}) - l_n^{\mathrm{ZMP}}(\hat{\boldsymbol{\delta}}^{\mathrm{ZMP}}))$, where $l_n^{\mathrm{P}}(\cdot)$ and $\hat{\boldsymbol{\beta}}^{\mathrm{P}}$ denote, respectively, the log-likelihood and the MLE in a Poisson regression model. Further $l_n^{\text{ZMP}}(\cdot)$, and $\hat{\boldsymbol{\delta}}^{\text{ZMP}} = (\hat{\boldsymbol{\beta}}^{\text{ZMP}}, \hat{\omega}^{\text{ZMP}})$ denote, respectively, the log-likelihood and the MLE in a ZMP regression model. The score statistic for the above testing problem is derived in detail by Jansakul and Hinde (2002) and is not given here. Following them, the score statistic is denoted by S_ω.

Estimated upper tail probabilities for an α size test were computed by calculating the proportion of times that W_ω, LR_ω, or S_ω were greater than or equal to the critical value $\chi^2_{1,1-\alpha}$. For example, for the Wald test we determined $\#\{j : W_\omega^j \geq \chi^2_{1,1-\alpha}, j = 1, \ldots, 1,000\}/1,000$. Here $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)100\%$ quantile of a χ^2 distribution with 1 degree of freedom and W_ω^j denotes the value of W_ω in the $j-$th sample. Note that when samples are drawn from the Poisson distribution the estimated upper tail probabilities correspond to the estimated level of the test. For ZMP samples with ZM $\omega > 0$, the estimated upper tail probabilities give the estimated power function at ω. These values are given in Table 1 for all three tests in the case of nonconstant linear predictors $\eta_i(\boldsymbol{\beta}) = 0.75 - 1.45x_i$, $i = 1, \ldots, n$. We observe that the Wald and LR tests were conservative while the score test was often somewhat liberal. Despite this fact, the Wald test had higher power than the score test for samples of size $n = 50$ and $n = 100$, and especially at level $\alpha = 0.01$. For example when $\omega = 0.45$, $n = 50$ and level $\alpha = 0.01$ the power of the score test was 0.471, approximately 69% of the power (0.683) of the corresponding Wald test. Here and in the sequel, percents are rounded to integers. It should be noted that our results for the score test are in a good agreement with results in Table 2 from Jansakul and Hinde (2002). In general when $\eta_i(\boldsymbol{\beta}) = 0.75 - 1.45x_i$, $i = 1, \ldots, n$ the score test resulted in power loss between 15% (5%) and 38% (27%) compared to the Wald test for $n = 50$ ($n = 100$). For sample size $n = 200$ these tests were almost equally powerful. Simulation results for constants linear predictors are only briefly reported. In the case of the constant linear predictors $\eta_i(\beta) = 0.75$ all three tests performed about equally well. In contrast to this, the Wald test was more powerful than others for $\eta_i(\beta) = 0.25$. The power loss for the score test compared to the Wald test was between 15% (2%) and 43% (26%) for sample size $n = 50$ (n=100). Thus a higher percentage of zeros arising from the Poisson part resulted in a higher power loss for the score and LR tests compared to the Wald test. It should be noted that in our simulation for the ZMP case, the difference in power for the score and LR tests was always negligible for constant as well as nonconstant linear predictors (see e.g., Table 1).

As noted above, for the sample size $n = 200$, the Wald and the likelihood ratio tests did not achieve their nominal level $\alpha = 0.05$ or $\alpha = 0.01$ satisfactory, while the score test achieved its nominal level quite well. In order to shed light on this phenomenon, we plot the profile log-likelihood with respect to ω for two data

Table 1. Estimated upper tail probabilities for Wald (W_ω), LR (LR_ω) and score (S_ω) statistics at $\chi^2_{1,1-\alpha}$ based on 1,000 samples from the ZMP model with nonconstant linear predictors $\eta_i(\boldsymbol{\beta}) = 0.75 - 1.45x_i$.

Level of the tests		$\alpha = \mathbf{0.05}$			$\alpha = \mathbf{0.01}$		
		W_ω	LR_ω	S_ω	W_ω	LR_ω	S_ω
$n = 50$	$\omega = 0.00$	0.023	0.019	0.047	0.008	0.007	0.014
	$\omega = 0.25$	0.407	0.339	0.340	0.244	0.151	0.152
	$\omega = 0.45$	0.804	0.680	0.685	0.683	0.471	0.471
$n = 100$	$\omega = 0.00$	0.027	0.030	0.068	0.006	0.005	0.016
	$\omega = 0.25$	0.594	0.504	0.510	0.397	0.276	0.288
	$\omega = 0.45$	0.931	0.888	0.884	0.871	0.734	0.740
$n = 200$	$\omega = 0.00$	0.019	0.019	0.060	0.002	0.002	0.011
	$\omega = 0.25$	0.934	0.918	0.919	0.842	0.795	0.800
	$\omega = 0.45$	1.000	1.000	1.000	0.999	0.997	0.997

sets of responses and covariates of sample size $n = 200$ from the simulation setup with true ZM parameters, $\omega_{true} = 0$ and $\omega_{true} = 0.25$. The profile log-likelihood function for ω in ZMP models is their log-likelihoods as a function of ω evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{\text{ZMP}}$, i.e.,

$$l_n^{\text{ZMP}}\Big((\hat{\boldsymbol{\beta}}^{\text{ZMP}}, \omega)\Big). \tag{5.1}$$

On the left plot of Figure 1 we see that the profile log-likelihood (solid curve) based on the data set with true ZM parameter $\omega_{true} = 0$ substantially deviates from its quadratic approximation (dashed curve). The quadratic approximation to the profile is just the second order Taylor expansion of $l_n^{\text{ZMP}}((\hat{\boldsymbol{\beta}}^{\text{ZMP}}, \omega))$ at $\hat{\omega}^{\text{ZMP}}$. In contrast, in the right plot we see that the quadratic approximation (dashed curve) to the profile log-likelihood function (solid curve) based on the data set with true ZM parameter $\omega_{true} = 0.25$ is very accurate. It is well known (see e.g., Meeker and Escobar (1995) and Pawitan (2000)) that the Wald and LR tests may fail if the quadratic approximation to the log-likelihood function fails. That is exactly the case here and therefore the Wald and the LR tests do not maintain their nominal level α. It should be mentioned that for both data sets the lowest permissible negative value of ω (i.e., $-c_\omega$) is around -0.05, and the 99% confidence intervals for ω are determined by the horizontal solid lines in Figure 1.

We also conducted an extensive simulation study to compare the performance of score, Wald, and LR tests in ZMGP regression models for samples of size $n = 50, 100, 200$. For brevity we report only some results from this study. A ZMGP model with $\varphi = 2$, $\omega_j = 0.05j$ for $j = 0, \ldots, 9$ and linear predictors $\eta_i(\boldsymbol{\beta}) = 1 + 0.5x_i$ for $i = 1, \ldots, n$ and $n = 50, 100, 200$, was taken as a working model. Covariates x_i's were taken uniformly from $(0, 1)$. For each combination of sample

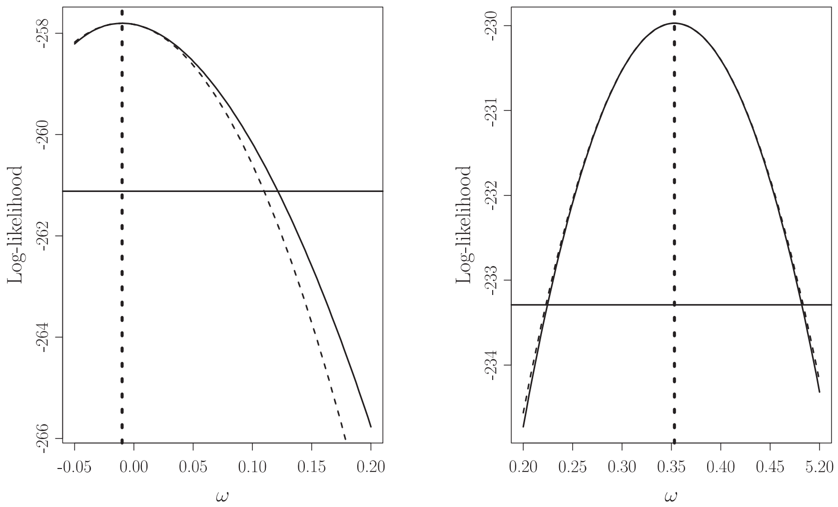

Figure 1. The profile log-likelihood for ω of a data set of size $n = 200$ simulated from the ZMP model with $\eta_i(\boldsymbol{\beta}) = 0.75 - 1.45x_i$ and the true ZM parameter $\omega_{true} = 0$ (left), or $\omega_{true} = 0.25$ (right) (Vertical dotted lines correspond to the MLE's, and horizontal solid lines indicate asymptotic 99% confidence intervals for ω. The solid curves correspond to the profile log-likelihood given by (5.1) and the dashed curves correspond to the quadratic approximation of the profile log-likelihood).

size and model, we simulated 1,000 sets of responses from the working model. This simulation setup implies that the probability of obtaining zero outcomes from the GP distribution with parameters $\varphi = 2$ and $\mu_i = \exp(\eta_i(\boldsymbol{\beta}))$ varies between 0.11 and 0.25 for $i = 1, \ldots, n$. We display our findings in Figure 2. The estimated power of the tests between two neighbouring knot points ω_j and ω_{j+1} for $j = 0, \ldots, 8$, is obtained by linear interpolation. From Figure 2 we see that all three tests maintained approximately their size, while the Wald test was much more powerful than the LR test, and even more powerful than the score test. A sample size of 50 was needed for the Wald test to achieve 80% power at $\omega = 0.40$ and level $\alpha = 0.05$, while for the score test a sample size of 100 was not sufficient. Taking the total cost for sampling and statistical inference, the Wald test appears much more effective than the score test: the power loss for the score test compared to the Wald test lies between 46% and 87% for $n = 50$, and between 22% and 73% for $n = 100$. In contrast to the ZMP case, for $n = 200$ the percent difference in the power for the score and Wald tests was still significant, between 2% and 56%. Thus the score test performed worse when an additional overdispersion parameter compared to the Poisson distribution was allowed. Moreover the LR test had significantly higher power than the score test,

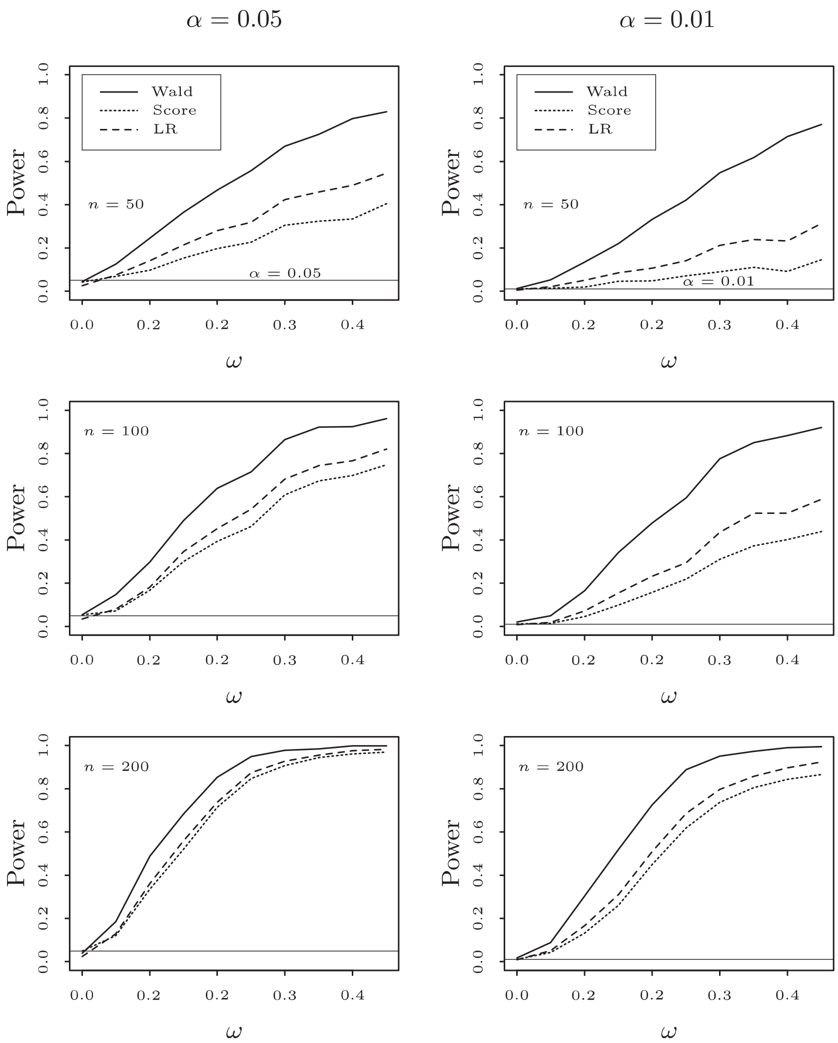

Figure 2. Estimated upper tail probabilities for Wald, LR, and score statistics at $\chi^2_{1,1-\alpha}$ in the ZMGP regression, based on 1,000 samples from the ZMGP model with linear predictors $\eta_i(\boldsymbol{\beta}) = 1 + 0.5x_i$ (horizontal line corresponds to α level).

not the case in ZMP regression. The percent difference in power for the score and LR tests was between 8% and 64% for $n = 50$, 8% and 36% for $n = 100$, 1% and 20% for $n = 200$. With regard to the Wald and LR tests, we observe that the LR test resulted in power loss up to 68% compared to the Wald test.

In contrast to the ZMP case, the Wald test for zero-modification maintained the nominal level α in ZMGP models much better. Table 2 displays empirical levels of all three tests for zero-modification from Figure 2 when $\omega = 0$. It should

Table 2. Empirical levels for Wald, LR, and score statistics in the ZMGP regression, based on 1000 samples from the ZMGP model with linear predictors $\eta_i(\boldsymbol{\beta}) = 1 + 0.5x_i$. They are the estimated upper tail probabilities from Figure 2 when $\omega = 0$.

Level of the tests	$\alpha = \mathbf{0.05}$			$\alpha = \mathbf{0.01}$		
Sample size	W_ω	LR_ω	S_ω	W_ω	LR_ω	S_ω
$n = 50$	0.044	0.026	0.043	0.012	0.005	0.010
$n = 100$	0.055	0.034	0.054	0.020	0.009	0.012
$n = 200$	0.040	0.025	0.049	0.017	0.009	0.011

be noted that the LR test was still conservative for $\alpha = 0.05$ and failed to achieve its nominal level. This could be again caused by a poor quadratic approximation of the log-likelihood function.

5.2. Apple propagation data

Ridout et al. (2001) analyzed data on the number of roots produced by 270 shoots of a certain apple cultivar. The shoots had been produced under an 8– or 16– hour photoperiod (Factor "P") in culture systems that utilized one of four different concentrations of cytokinin BAP (Factor "H") in the culture medium (for more details see Ridout and Demétrio (1992) and Marin, Jones and Hadlow (1993)). Note that the data contain a large number of zero responses for the 16–hour photoperiod. Ridout et al. (2001) derived a score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternative, and showed that the zero-inflated Poisson model is unsuitable for these data.

Here we consider two different ZMGP models for the entire data set and one ZMGP model for the subdata that were collected under 16–hour photoperiod. In the first model for the entire data (Model 1), μ may take different values only for two levels of Factor "P", while in the second model (Model 2), μ may take different values for each of the eight treatment combinations ("P∗H"). For the partial data we fit the ZMGP model similar to Model 2, i.e., μ takes different values for each four levels of Factor "H". This model is referred to as Model 3. The overdispersion parameter φ is taken to be constant in all models. We are interested in testing for ZM, i.e., $H_0 : \omega = 0$ against $H_1 : \omega \neq 0$.

The values of the corresponding score, Wald, and LR statistics for testing ZM are given in Table 3. The Wald and LR tests clearly indicate that a simple GP regression without ZM is not sufficient for the whole apple propagation data as well as for its part with a 16–hour photoperiod. The score test detects ZM only in the partial data and is not powerful enough to do so in the entire data set. Moreover we see that, for the partial data, the Wald test gives much higher

Table 3. The values of the score, Wald, and LR statistics for testing ZM in the apple propagation data. The corresponding p–values are given in parentheses.

Data	**Model**	**Score statistic**	**Wald statistic**	**LR statistic**
Complete	Model 1:	0.45	72.96	8.03
	Factor "P"	(0.50)	($< 10^{-16}$)	(0.005)
Complete	Model 2:	0.57	73.18	14.41
	Factor "P" $*$ Factor "H"	(0.45)	($< 10^{-16}$)	(10^{-4})
Partial	Model 3 :	26.84	104.49	46.23
	Factor "H"	($2 \cdot 10^{-7}$)	($< 10^{-16}$)	(10^{-11})

evidence for ZM than the LR and score tests due to the fact that the Wald test is much more powerful compared to them, as seen in our simulation.

For the partial data, the ZMGP model and the the corresponding GP model are compared with respect to their fit to the empirical mean $\widehat{E}(Y|H = i)$ and variance $\widehat{V}ar(Y|H = i)$ $(i = 1, \ldots 4)$ for the 4 different levels of Factor "H". Recall that the data contains replications for each level of Factor "H", therefore the $\widehat{E}(Y|H = i)$ and $\widehat{V}ar(Y|H = i)$ $(i = 1, \ldots 4)$ can be computed. The mean and variance in the GP and ZMGP regression models are

$$\begin{aligned}
E(Y|H = i) &= \exp(\mathbf{x}_i^t\boldsymbol{\beta}^{\text{GP}}), \qquad Var(Y|H = i) = (\varphi^{\text{GP}})^2 \exp(\mathbf{x}_i^t\boldsymbol{\beta}^{\text{GP}}),\\
E(Y|H = i) &= (1-\omega)\exp(\mathbf{x}_i^t\boldsymbol{\beta}^{\text{ZMGP}}),\\
Var(Y|H = i) &= (1-\omega)\exp(\mathbf{x}_i^t\boldsymbol{\beta}^{\text{ZMGP}})\Big((\varphi^{\text{GP}})^2 + \omega\exp(\mathbf{x}_i^t\boldsymbol{\beta}^{\text{ZMGP}})\Big),
\end{aligned}$$

respectively. Here $(\varphi^{\text{GP}}, \boldsymbol{\beta}^{\text{GP}})$ and $(\varphi^{\text{ZMGP}}, \omega, \boldsymbol{\beta}^{\text{ZMGP}})$ denote the parameters of the GP and ZMGP models, respectively. Hence 95% confidence intervals (CI's) for the mean and variance of the both regressions can be constructed and plotted for all covariates $\mathbf{x}_i$ $(i = 1, \ldots, 4)$ on the basis of basis of the Delta method (van der Vaart (1998, Chap. 3)) and normality of the MLE $\hat{\boldsymbol{\delta}}$ in ZMGP and GP regression models (Theorem 1 and Remark (ii)).

From Figure 3, we see that the 95% CI's in the ZMGP case are always shorter, and predicted values for mean and variance are closer to their empirical values than in the GP case. The only exception is the prediction of the mean in the case of Level 3 of Factor "H", where the GP regression better estimates the mean. This is caused by the fact that the frequency of observed zero responses is lower here compared to other levels of Factor "H" (40% ($H = 3$) versus 50% ($H = 1$), 53.3% ($H = 2$), and 47.5% ($H = 4$)). The MLE's and the corresponding asymptotic 95% CI's for the ZM parameter ω and overdispersion parameter φ given in Table 4 also support the necessity of ZM in GP models for the apple propagation data.

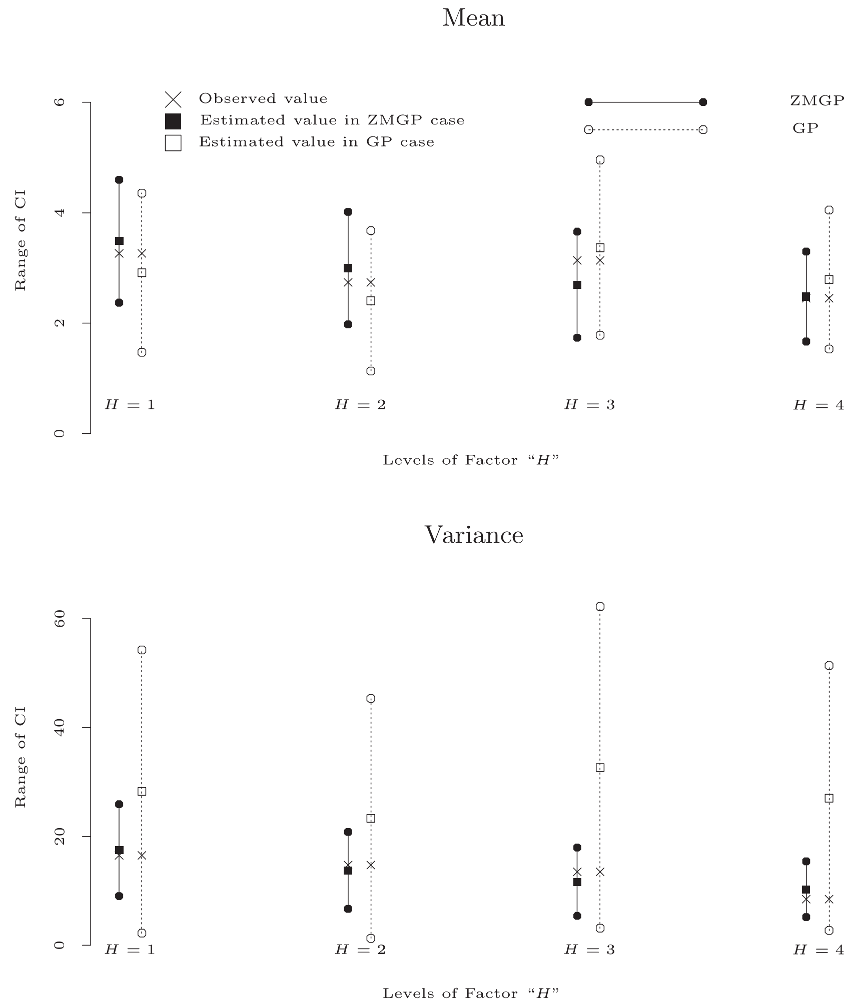

Figure 3. 95% confidence intervals for the mean (top panel) and variance (bottom panel) of the partial apple propagation data for ZMGP and GP models.

Gupta et al. (2004) also analyzed these data within the framework of a zero-inflated regression model associated with a RGP distribution. Their score tests strongly indicate that a zero-inflated RGP regression is suitable.

6. Conclusions and Discussions

This paper shows that the MLE's in ZMGP (GP, ZMP) regression models possess similar asymptotic properties as GLM regression models despite the fact that the ZMGP (GP, ZMP) distribution does not belong to the exponential

Table 4. MLE's and the corresponding 95% confidence intervals (CI) for ω and φ in the ZMGP regression for the apple propagation data.

Data	**Model**	$\hat{\omega}$	$\hat{\varphi}$	**CI for** ω	**CI for** φ
Complete	Model 1	0.2225	1.2782	(0.1714, 0.2735)	(1.1423, 1.4141)
Complete	Model 2	0.2231	1.2427	(0.1720, 0.2742)	(1.1118, 1.3736)
partial	Model 3	0.4638	1.4154	(0.3749, 0.5527)	(1.1327, 1.6981)

family. General results of Fahrmeir and Kaufmann (1985) for noncanonical links in GLM have been adopted for this purpose. A simulation study exhibits that the power of the score test for testing ZM in ZMP regression can be up to 43% lower than the power of the corresponding Wald test. In the case of ZMGP regression, this difference increases to 87%. The effect of the poor performance of the score test seen in our simulation studies can also be seen in the analysis of the apple propagation data. The score test does not detect any ZM despite the high proportion of zeros observed for one level of Factor "P". The superiority with regard to fit of ZMGP models over GP models is also illustrated on this data set. The zero-inflated count regression models have been found to be appropriate for these data by Ridout et al. (2001) and Gupta et al. (2004). We conclude that score test for testing ZM in ZMP and ZMGP models can be highly misleading, and the Wald and LR tests should be used instead.

It is often of interest to test whether ZI and/or overdispersion adjustments in ZIGP regression models are needed. In this testing problem the true parameter (or the parameter vector) lies on the boundary of a parameter space. To derive the corresponding Wald and LR tests, we have to deal with a delicate boundary problem as in Moran (1971), Self and Liang (1987) and Vu and Zhou (1997). Further the small sample performance of these boundary Wald, LR, and score tests needs to be investigated. These are the subjects of future research.

Acknowledgement

We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (Cz 86/1-3). We thank Marie Hušková and Axel Munk for helpful discussions. We also would like to express our gratitude to two anonymous referees and an associate editor for helpful remarks and suggestions.

References

Bae, S., Famoye, F., Wulu, J. T., Bartolucci, A. A. and Singh, K. P. (2005). A rich family of generalized Poisson regression models. *Math. Comput. Simulation* **69**, 4-11.

Cameron, A. and Trivedi, P. (1998). *Regression Analysis of Count Data*. Cambridge University Press.

Consul, P. C. (1989). *Generalized Poisson Distributions*. Marcel Dekker Inc., New York.

Consul, P. C. and Famoye, F. (1992). Generalized Poisson regression model. *Comm. Statist. Theory Methods* **21**, 89-109.

Consul, P. C. and Jain, G. C. (1970). On the generalization of Poisson distribution. *Ann. Math. Statist.* **41**, 1387.

Consul, P. C. and Shoukri, M. M. (1985). The generalized Poisson distribution when the sample mean is larger than the sample variance. *Comm. Statist. Simulation Comput.* **14**, 667-681.

Czado, C., Erhardt, V., Min, A. and Wagner, S. (2007). Zero-inflated generalized poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates. *Stat. Model.* **7**, 125-153.

Czado, C. and Sikora, I. (2002). Quantifying overdispersion effects in count regression data. Discussion paper 289 of SFB 386 (`http://www.stat.uni-muenchen.de/sfb386/`).

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

Dean, C. and Lawless, J. F. (1989).Tests for detecting overdispersion in Poisson regression models. *J. Amer. Statist. Assoc.* **84**, 467-472.

Deng, D. and Paul, S. R. (2000). Score tests for zero inflation in generalized linear models. *Canad. J. Statist.* **28**, 563-570.

Deng, D. and Paul, S. R. (2005). Score tests for zero-inflation and over-dispersion in generalized linear models. *Statist. Sinica* **15**, 257-276.

Dietz, E. and Böhning, D. (2000).On estimation of the Poisson parameter in zero-modified Poisson models, *Comput. Statist. Data Anal.* **34**, 441-459.

Fahrmeir, L. and Kaufmann, H. (1985).Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13**, 342-368.

Famoye, F. (1993). Restricted generalized Poisson regression model. *Comm. Statist. Theory Methods* **22**, 1335-1354.

Famoye, F. and Singh, K. P. (2003). On inflated generalized Poisson regression models. *Adv. Appl. Stat.* **3**, 145-158.

Famoye, F. and Singh, K. P. (2006). Zero-inflated generalized Poisson model with an application to domestic violence data. *Journal of Data Science* **4**, 117-130.

Gschlößl, S. and Czado, C. (2008). Modelling count data with overdispersion and spatial effects. *Statist. Papers* **49**, 531-552.

Gupta, P. L., Gupta, R. C. and Tripathi, R. C. (2004). Score test for zero inflated generalized Poisson regression model. *Comm. Statist. Theory Methods* **33**, 47-64.

Hall, D. B. and Berenhaut, K. S. (2002). Score tests for heterogeneity and overdispersion in zero-inflated Poisson and binomial regression models. *Canad. J. Statist.* **30**, 415-430.

Jansakul, N. and Hinde, J. P. (2002). Score tests for zero-inflated Poisson models, *Comput. Statist. Data Anal.* **40**, 75-96.

Joe, H. and Zhu, R. (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biom. J.* **47**, 219-229.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1-14.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *Canad. J. Statist.* **15**, 209-225.

Marin, J., Jones, O. and Hadlow, W. (1993). Micropropagetion of columnar apple trees. *Journal of Horticultiral Science* **68**, 289-297.

Meeker, W. Q. and Escobar, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *Amer. Statist.* **49**, 48-53.

Moran, P. A. P. (1971). Maximum-likelihood estimation in non-standard conditions. *Proc. Cambridge Philos. Soc.* **70**, 441-450.

Mullahy, J. (1986). Specification and testing of some modified count data models. *J. Econometrics* **33**, 341-365.

Pawitan, Y. (2000).A reminder on the fallibility of the Wald statistics: Likelihood explanation. *Amer. Statist.* **54**, 54-56.

Ridout, M. and Demétrio, C. G. B. (1992). Generalized linear models for positive count data. *Revista de Matematica e Estatistica* **10**, 139-148.

Ridout, M., Hinde, J. and Demétrio, 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.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82**, 605-610.

Silvapulle, M. J. and Silvapulle, P. (1995). A score test against one-sided alternatives. *J. Amer. Statist. Assoc.* **90**, 342-349.

Stewart, G. W. (1998). *Matrix Algorithms*. Vol. I, Society for Industrial and Applied Mathematics, Philadelphia, PA.

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

van der Vaart, A. W. (1998). *Asymptotic Statistics, Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics* **59**, 254-262.

Vu, H. T. V. and Zhou, S. (1997). Generalization of likelihood ratio tests under nonstandard conditions. *Ann. Statist.* **25**, 897-916.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica* **57**, 307-333.

Winkelmann, R. (2008). *Econometric Analysis of Count Data.*, 5th edition. Springer-Verlag, Berlin.

Zentrum Mathematik, Technische Universität München, Boltzmannstr. 3, D-85747 Garching bei München, Germany.

E-mail: aleksmin@ma.tum.de

Zentrum Mathematik, Technische Universität München, Boltzmannstr. 3, D-85747 Garching bei München, Germany.

E-mail: cczado@ma.tum.de

(Received April 2008; accepted October 2008)