Czado, Min:

Testing for zero-modification in count regression models

Sonderforschungsbereich 386, Paper 474 (2006)

Testing for zero-modification in count regression models

Claudia CZADO and **Aleksey MIN**[1]

Center for Mathematical Sciences
Munich University of Technology
Boltzmannstr. 3
D-85747 Garching, 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 classes of models for such data. In the literature so far only score tests are used for testing the necessity of this adjustment. For this testing problem we show how poor the performance of the corresponding score test 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 also prove consistency and asymptotic normality of the maximum likelihood estimators in the above mentioned regression models to give a theoretical justification for Wald and likelihood ratio tests.

Keywords: generalized Poisson distribution; likelihood ratio test; maximum likelihood estimator; overdispersion; score test; Wald test; zero-modification

[1]Corresponding author.
Fax: +49/89/289-17435
E-mail address: aleksmin@ma.tum.de (A. Min).

1 Introduction

Zero-inflated generalized Poisson (ZIGP) regression models have recently been found useful for the analysis of count data with a large amount of zeros (see e.g. Famoye and Singh (2003), Gupta et al. (2004), Joe and Zhu (2005), Bae et al. (2005) and Famoye and Singh (2006)). It is a large class of regression models which contains zero-inflated Poisson (ZIP), generalized Poisson (GP) and Poisson regressions (Mullahy (1986), Lambert (1992), Consul and Famoye (1992) and Famoye (1993)). The interest in this class of regression models is driven by the fact that it can handle overdispersion and/or zero-inflation which count data very often exhibit.

Score tests are widely used for testing misspecifications in count regression models because they require to fit the model only under the null hypothesis. In particular, van den Broek (1995) proposed a score test for testing zero-inflation in ZIP regression and Gupta et al. (2004) derived score tests for testing zero-inflation or overdispersion in ZIGP regression. The score test for zero-inflation considered by the above authors is as they noted the score test for zero-inflation or zero-deflation, i.e. for zero-modification in zero-modified Poisson (ZMP) regression (see Dietz and Böhning (2000)) and zero-modified generalized Poisson (ZMGP) regression. In order to derive a score test only for zero-inflation, the problem of testing parameters on the boundary of the parameter space needs to be addressed. Consequently, the limiting distribution of the score statistic will differ from a standard χ^2-distribution. For insightful discussions on this problem we would like to refer to Verbeke and Molenberghs (2003).

Nowadays, given modern computing power, the computational advantage of score tests has lost some of its original attractivity in many problems. Therefore we think that more attention should be paid to Wald and likelihood ratio (LR) tests for ZMP and ZMGP regressions. The objective of this paper is to derive the appropriate asymptotic theory for the ZMGP regression models and to investigate the performance of Wald, LR and score tests for testing zero-modification, i.e. zero-inflation or zero-deflation. Our theoretical results also remain valid for GP and ZMP regression models subject to appropriate changes in assumptions.

There is also a count regression for overdispersed and zero-inflated data based on a negative binomial (NB) distribution. This is a zero-inflated negative binomial (ZINB) regression (see Ridout et al. (2001) and Hall and Berenhaut (2002)). It is not a subject of the paper but we list most important differences, from our point of view, between regression models based on a NB and GP distributions in next sections.

In Section 2 we introduce the GP distribution and discuss its basic forms and properties. A ZMGP regression model is defined in Section 3. Section 4 gives the asymptotic existence, the consistency and the asymptotic normality of the ML estimator in a ZMGP regression model. In Section 5 we compare the performance of the score test for detecting zero-modification in ZMP and ZMGP models to the performance of the Wald and LR tests in a simulation study. In particular it is shown that using the score test one may lose in test power compared to the Wald test up to 47% for the ZMP case and up to 87% for the ZMGP case. We also illustrate that the score test for zero-modification in the analysis of the apple propagation data (see Ridout and Demétrio (1992)) does not always detect zero-modification while the Wald and LR tests give strong evidence for zero-modification. Thus the score test can result in misleading conclusions about the presence of zero-modification. The Fisher information matrix of the ZMGP regression and the proof of Theorem 1 is given in the Appendix.

2 The GP distribution

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

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

The real-valued parameters μ and φ are assumed to satisfy the following constraints:

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

If $\varphi < 1$ then (1) does not correspond to a probability distribution. However the lower limit, imposed on φ in this case, guarantees us 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 found to be a quite 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 of this distribution 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), page 12), if $\tilde{Y} \sim GP(\mu,\varphi)$ then the mean and variance of Y are given by

$$E(\tilde{Y}) = \mu \tag{2}$$

and

$$Var(\tilde{Y}) = \varphi^2\mu. \tag{3}$$

A NB distribution with mean μ and overdispersion parameter $a > 0$ (see Lawless (1987) for precise definition) also has a flexible variance function. Its variance is given by $\mu(1 + a\mu)$. 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; a fact concurred 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 whenever 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 (2005).

There is a form of the GP distribution obtained by assuming that $\varphi - 1$ is linearly proportional to μ, say $\varphi - 1 = \alpha\mu$ 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. To avoid the point indicated in the previous paragraph we deal here only with an unrestricted form (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 given by

$$P_{\mu,\varphi,\omega}(y) := P(Y = y) = \begin{cases} \omega + (1-\omega)P(\tilde{Y} = 0) & y = 0, \\ (1-\omega)P(\tilde{Y} = y) & y = 1, 2, \ldots, \end{cases} \tag{4}$$

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

$$\frac{-\exp(-\mu/\varphi)}{1-\exp(-\mu/\varphi)} \le \omega \le 1. \tag{5}$$

Thus, this distribution has 3 parameters μ, φ and ω and will be further denoted by $ZMGP(\mu, \varphi, \omega)$.

The above condition (5) ensures that (4) defines a probability mass function for negative values of ω corresponding to zero-deflation. Positive values of the parameter ω correspond to zero-inflation which mostly occurs in practice. In this case ω is a probability of zero outcome of a zero-inflating Bernoulli distribution.

A simple calculation using equations (2) and (3) imply that the mean and variance of the ZMGP distribution are given by

$$E(Y) = (1-\omega)\mu \tag{6}$$

and

$$Var(Y) \quad = \quad E(Y)\left(\varphi^2 + \mu\omega\right). \tag{7}$$

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 (6) and (7) 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 $\mathbf{x}_i = (x_{i0}, x_{i1}, \dots, x_{ip})^t$ with $x_{i0} = 1$ for $i = 1, \dots, n$:

1. *Random components:*
   $\{Y_i,\ 1 \leq i \leq n\}$ are independent where $Y_i \sim ZMGP(\mu_i, \varphi, \omega)$.

2. *Systematic components:*
   The linear predictors $\eta_i(\boldsymbol{\beta}) = \mathbf{x}_i^t\boldsymbol{\beta}$ for $i = 1, \dots, n$ influence the response Y_i. Here $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^t$ is a vector of unknown regression parameters. The matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^t$ 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 $i = 1, \dots, n$.

Here and in the subsequent sections, $\mathbf{A}^t$ and $\mathbf{a}^t$ denote the transpose of a matrix $\mathbf{A}$ and a vector $\mathbf{a}$, respectively. To stress the fact that the distribution of the responses Y_i's does not belong to the exponential family, this regression will be called the ZMGP regression model. It should be noted that parameter φ and ω are assumed to be constant and (5) now should hold for all μ_i, $i = 1, 2, \dots, n$. Further, we denote the joint vector of the regression parameters $\boldsymbol{\beta}$ and the parameters φ and ω of the ZMGP distribution by $\boldsymbol{\delta}$, i.e. $\boldsymbol{\delta} := (\boldsymbol{\beta}^t, \varphi, \omega)^t$, and its ML estimator by $\hat{\boldsymbol{\delta}}$.

The following abbreviations for $i = 1, \dots, n$ will be used throughout in the paper:

$$\begin{aligned}
\mu_i(\boldsymbol{\beta}) &:= \exp\left(\mathbf{x}_i^t\boldsymbol{\beta}\right), \\
f_i(\boldsymbol{\beta}, \varphi) &:= \exp\left(-\mu_i(\boldsymbol{\beta})/\varphi\right), \\
g_i(\boldsymbol{\delta}) &:= \omega + (1-\omega) f_i(\boldsymbol{\beta}, \varphi) = P_{\mu_i(\beta), \varphi, \omega}(0).
\end{aligned}$$

For observations $y_1, \dots, y_n$, the log-likelihood $l(\boldsymbol{\delta})$ derived from the ZMGP regression can be written as

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

Further the score vector, i.e. the vector of the first derivatives, has the following representation:

$$\mathbf{s}_n(\boldsymbol{\delta}) = \left(s_0(\boldsymbol{\delta}),\ldots,s_p(\boldsymbol{\delta}),s_{p+1}(\boldsymbol{\delta}),s_{p+2}(\boldsymbol{\delta})\right)^t, \tag{8}$$

where

$$s_r(\boldsymbol{\delta}) := \frac{\partial l_n(\boldsymbol{\delta})}{\partial\beta_r} = \sum_{i=1}^{n} s_{r,i}(\boldsymbol{\delta})$$

with

$$\begin{aligned} s_{r,i}(\boldsymbol{\delta}) &:= -x_{ir}\mathbb{1}_{\{y_i=0\}}\frac{(1-\omega)f_i(\boldsymbol{\beta},\varphi)\mu_i(\boldsymbol{\beta})}{\varphi g_i(\boldsymbol{\delta})} \\ &+ x_{ir}\mathbb{1}_{\{y_i>0\}}\left(1 + \frac{\mu_i(\boldsymbol{\beta})(y_i-1)}{\mu_i(\boldsymbol{\beta})+(\varphi-1)y_i} - \frac{\mu_i(\boldsymbol{\beta})}{\varphi}\right) \end{aligned} \tag{9}$$

for $r = 0,\ldots,p$,

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

with

$$\begin{aligned} s_{p+1,i}(\boldsymbol{\delta}) &:= \mathbb{1}_{\{y_i=0\}}\frac{(1-\omega)f_i(\boldsymbol{\beta},\varphi)\mu_i(\boldsymbol{\beta})}{\varphi^2 g_i(\boldsymbol{\delta})} \\ &+ \mathbb{1}_{\{y_i>0\}}\left(\frac{y_i(y_i-1)}{\mu_i(\boldsymbol{\beta})+(\varphi-1)y_i} - \frac{y_i}{\varphi} + \frac{\mu_i(\boldsymbol{\beta})-y_i}{\varphi^2}\right) \end{aligned} \tag{10}$$

and

$$s_{p+2}(\boldsymbol{\delta}) := \frac{\partial l_n(\boldsymbol{\delta})}{\partial\omega} = \sum_{i=1}^{n} s_{p+2,i}(\boldsymbol{\delta})$$

with

$$s_{p+2,i}(\boldsymbol{\delta}) := \mathbb{1}_{\{y_i=0\}}\frac{1-f_i(\boldsymbol{\beta},\varphi)}{g_i(\boldsymbol{\delta})} - \mathbb{1}_{\{y_i>0\}}\frac{1}{1-\omega}, \tag{11}$$

for $i = 1,\ldots,n$.

4 Asymptotic theory

Fahrmeir and Kaufmann (1985) proved consistency and asymptotic normality of the ML estimator in GLM for canonical as well as noncanonical link functions under mild assumptions. Their method can be adapted for proving similar results for the ZMGP regression.

As in Fahrmeir and Kaufmann (1985), we use the Cholesky square root matrix for normalizing the ML estimator. 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 $\mathbf{A}^{1/2}\left(\mathbf{A}^{1/2}\right)^t = \mathbf{A}$ (see Stewart (1998), p. 188). For convenience, set $\mathbf{A}^{t/2} := \left(\mathbf{A}^{1/2}\right)^t$, $\mathbf{A}^{-1/2} := \left(\mathbf{A}^{1/2}\right)^{-1}$ and $\mathbf{A}^{-t/2} := \left(\mathbf{A}^{t/2}\right)^{-1}$. In this paper we deal only with the spectral norm of square matrices denoted by $\|\cdot\|$. The spectral norm of a real-valued matrix $\mathbf{A}$ is given by

$$\|\mathbf{A}\| = \left(\text{maximum eigenvalue of } \mathbf{A}^t\mathbf{A}\right)^{1/2} = \sup_{\|\mathbf{u}\|_2=1} \|\mathbf{A}\mathbf{u}\|_2 \,,$$

where $\|\cdot\|_2$ denotes the L^2– norm of vectors. We drop 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 $\mathbf{A}$ will be further denoted by $\lambda_{min}(\mathbf{A})$ and the vector of true parameter values of the ZMGP regression will be denoted as $\boldsymbol{\delta}_0$. Further $\mathbf{F}_n(\boldsymbol{\delta})$ will stand for the Fisher information matrix in a ZMGP regression evaluated at $\boldsymbol{\delta}$. It should be noted that the entries of the Fisher information matrix in a ZMGP regression have a closed form (see Appendix) while this is not the case in regression models associated with a NB distribution (see e.g. Lawless (1987)).

Now denote a neighborhood of $\boldsymbol{\delta}_0$ by

$$N_n(\varepsilon) = \{\boldsymbol{\delta} : \|\mathbf{F}_n^{t/2}(\boldsymbol{\delta}_0)(\boldsymbol{\delta} - \boldsymbol{\delta}_0)\| \leq \varepsilon\} \tag{12}$$

for $\varepsilon > 0$.

For convenience, we drop the arguments $\boldsymbol{\delta}_0$, $\boldsymbol{\beta}_0$ and φ_0 as well as the subindex $\boldsymbol{\delta}_0$ in $\mu_i(\boldsymbol{\beta}_0)$, $f_i(\boldsymbol{\beta}_0, \varphi_0)$, $g_i(\boldsymbol{\delta}_0)$, $P_{\boldsymbol{\delta}_0}$, $E_{\boldsymbol{\delta}_0}$ etc. and write μ_i, f_i, g_i, P, E etc. Constants will be further denoted by C and c, with subindexes 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 n-dimensional unit matrix will be denoted by $\mathbf{I}_n$ and an admissible set for a vector $\boldsymbol{\beta}$ of regression parameter will be denoted by B.

We make the following assumptions.

(A1)

$$\frac{n}{\lambda_{min}(\mathbf{F}_n)} \leq C_1 \quad \forall\ n \geq 1,$$

where C_1 is a positive constant.

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

(A3) Assume that $B \subset \mathbb{R}^{p+1}$ 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 (5) holds for all $\mathbf{x} \in K_x$, $\boldsymbol{\beta} \in B$ and $\varphi \in \Phi$.

Now we state our main result which is the analogue to Theorem 4 of Fahrmeir and Kaufmann (1985).

THEOREM 1. *Under the assumptions* (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) $\mathbf{F}_n^{t/2}(\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \mathbf{I}_{p+3})$ *as* $n \to \infty$ *(asymptotic normality).*

Remarks

(i) Assumption (A1) is more restrictive than the corresponding condition (D) of Fahrmeir and Kaufmann (1985). Assumption (A2) means that we deal with compact regressors.

(ii) If $\boldsymbol{\delta}_0$ lies on the boundary of parameter space K_δ, i.e. (A3) is violated, then statements of Theorem 1 do not hold anymore. Particularly 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 be performed in the log-likelihood, the ML 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 for 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 zero-inflation 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 zero-inflation.

By virtue of Remarks (ii) and (iii) of Theorem 1, we can construct the Wald and LR tests for testing zero-modification 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 zero-modification parameter. In particular, Jansakul and Hinde (2002) considered models with $\omega = 0, 0.25, 0.45$ and 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 1000 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% ($\eta_i(\boldsymbol{\beta}) = 0.75$) of 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 ML estimators 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 $H_1 : \omega \neq 0$ has the following form

$$W_\omega = \frac{\hat{\omega}^2}{\hat{\sigma}^2_\omega},$$

where $\hat{\omega}$ is the ML estimator of ω in a ZMP regression and $\hat{\sigma}^2_\omega$ is the estimated variance of $\hat{\omega}$, which is 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 given by

$$LR_\omega = -2(l_n^{\mathrm{P}}(\hat{\boldsymbol{\beta}}^{\mathrm{P}}) - l_n^{\mathrm{ZMP}}(\hat{\boldsymbol{\delta}}^{\mathrm{ZMP}})),$$

where $l_n^{\mathrm{P}}(\cdot)$ and $\hat{\boldsymbol{\beta}}^{\mathrm{P}}$ denote, respectively, the log-likelihood and the ML estimator in a Poisson regression model, $l_n^{\mathrm{ZMP}}(\cdot)$ and $\hat{\boldsymbol{\delta}}^{\mathrm{ZMP}} = (\hat{\boldsymbol{\beta}}^{\mathrm{ZMP}}, \hat{\omega}^{\mathrm{ZMP}})$ denote, respectively, the log-likelihood and the ML estimator in a ZMP regression model. The score statistic for the above testing problem is derived in detail by Jansakul and Hinde (2002) and therefore it is not given here. Further following them, the score statistic is denoted by S_ω.

Estimated upper tail probabilities for an α size test are computed by calculating the proportion of times when W_ω, LR_ω or S_ω are greater than or equal to the critical value $\chi^2_{1,1-\alpha}$. For the Wald test we have for example

$$\frac{\#\{j : W_\omega^j \geq \chi^2_{1,1-\alpha}, j = 1, \ldots, 1000\}}{1000}.$$

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 zero-modification

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 1000 samples from the ZMP model with nonconstant linear predictors $\eta_i(\beta) = 0.75 - 1.45x_i$

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

$\omega > 0$ the estimated upper tail probabilities give the estimated power function at ω. These values are given in Table 1 for the all three tests in the case of nonconstant linear predictors $\eta_i(\boldsymbol{\beta}) = 0.75 - 1.45x_i$, $i = 1, \ldots, n$. Thus we observe that the Wald and LR tests are conservative while the score test is often somewhat liberal. Despite this fact the Wald test has the 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 is 0.471 which is 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 results 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 become 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 among others for $\eta_i(\beta) = 0.25$. The loss in power for the score test compared to the Wald test was between 15% (2%) and 43% (26%) for sample size $n = 50$ (n=100). This shows that a higher percentage of zeros arising from the Poisson part results in a higher loss of power for the score and LR tests compared to the Wald test. It should be noted that in our simulation for 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).

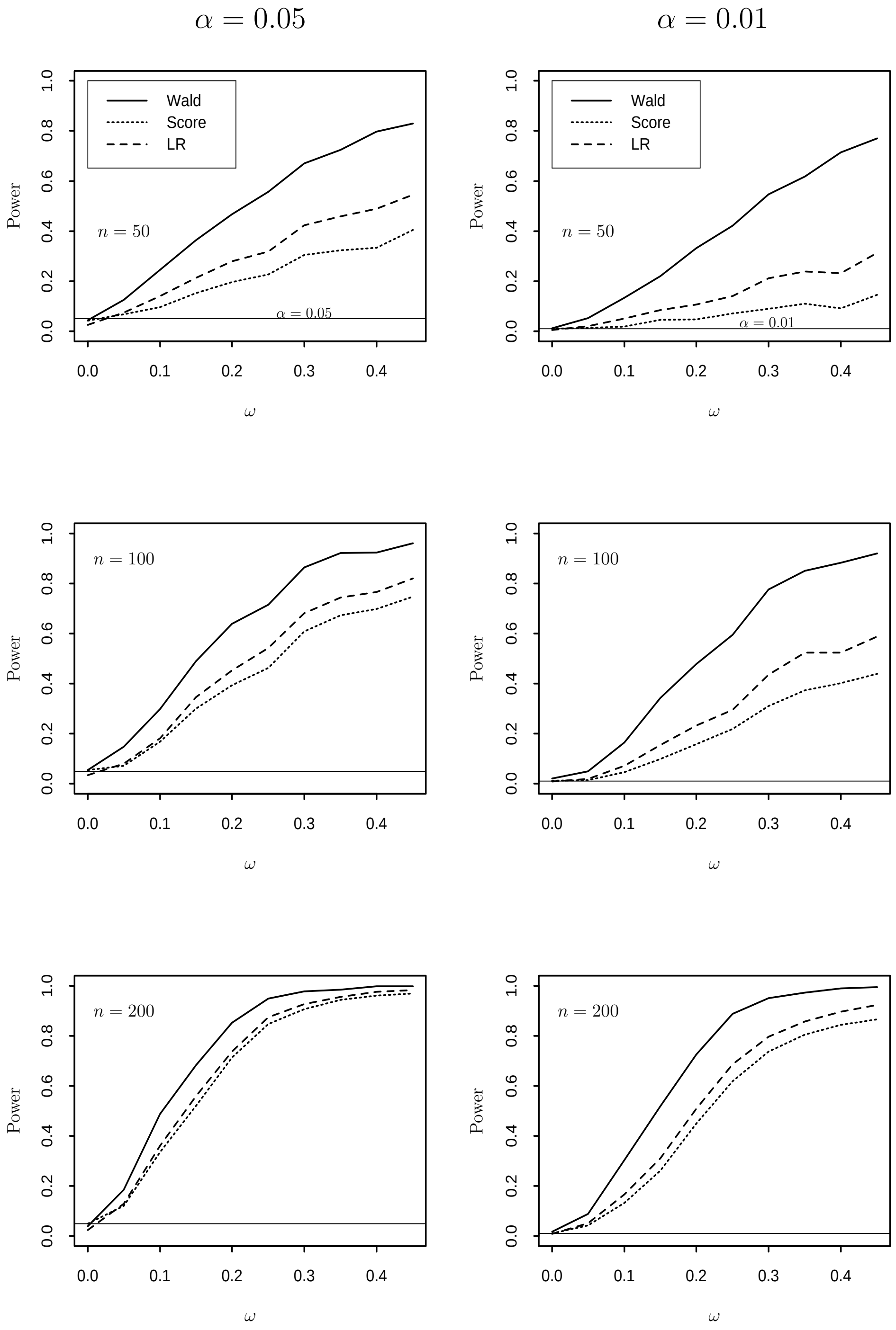
We also conducted an extensive simulation study to compare the performance of score, Wald and LR tests in ZMGP regression models for samples of size $n = 50, 100, 200$. For brevity we report only some results from this study. A ZMGP model with $\varphi = 2$, $\omega_j = 0.05j$ for $j = 0, \ldots, 9$ 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. As above, covariates x_i's were taken uniformly from $(0, 1)$. For each combination of sample size and model we simulated 1000 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 $i = 1, \ldots, n$. For a better visualization we displayed our findings in Figure 1. The power of the tests between two neighbour knot points ω_j and ω_{j+1} for $j = 0, \ldots, 8$ is obtained by linear interpolation. From Figure 1 we see that all three tests maintain approximately their size, while the Wald test is much powerful than the LR test and even more powerful than the score test. A sample size of 50 is 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 is not sufficient. Taking the total cost for sampling and statistical inference the Wald test will be much more effective than the score test. The loss in power for the score test compared to the Wald test lies between 46% and 87% for sample size $n = 50$ and between 22% and 73% for sample size n=100. In contrast to the ZMP case, for the sample size $n = 200$ the percent difference in the power for the score and Wald tests is still significant and lies between 2% and 56%. Thus the score test performs worse when an additional overdispersion parameter compared to the Poisson distribution is allowed. Moreover the LR test has significantly higher power than the score test which was not the case in ZMP regression. The percent difference in power for the score and LR tests is 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 observed that the LR test results in power loss up to 68% compared to the Wald test.

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 et al. (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 zero-inflated Poisson model is unsuitable for these data.

Here we consider two different ZMGP models for the entire data and one ZMGP model for its part that have been collected under 16–hour photoperiod. In the first

Figure 1: Estimated upper tail probabilities for Wald, LR and score statistics at $\chi^2_{1,1-\alpha}$ in the ZMGP regression based on 1000 samples from the ZMGP model with linear predictors $\eta_i(\beta) = 1 + 0.5x_i$



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 analogously to Model 2, i.e. μ takes different values for each four levels of Factor "H". This model is further referred as Model 3. Overdispersion parameter φ is taken to be constant in all models. Further we are interested in testing for zero-modification, i.e. the null hypothesis $H_0 : \omega = 0$ against the alternative $H_1 : \omega \neq 0$.

The values of the corresponding score, Wald and LR statistics for testing zero-modification are given in Table 2. Thus the Wald and LR tests clearly indicate that a simple GP regression without zero-modification is not sufficient for the whole apple propagation data as well as for its part with 16–hour photoperiod . The score test detects zero-modification only in the partial data and is not powerful enough to do it in the entire data. Moreover we see that for the partial data the Wald test gives much higher evidence for zero-modification than the LR and score tests which is due to the fact that the Wald test is much more powerful compared to them, as seen in the 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 $E(\widehat{Y|H=i})$ and variance $Var(\widehat{Y|H=i})$ $(i = 1, \dots 4)$ for the 4 different levels of Factor "H". Recall that the data contains replications for each level of Factor "H", therefore the $E(\widehat{Y|H=i})$ and $Var(\widehat{Y|H=i})$ $(i = 1, \dots 4)$ can be computed. Further the mean and variance in the GP and ZMGP regression models are given by

$$\begin{aligned} E(Y|H=i) &= \exp\left(\mathbf{x}_i^t \boldsymbol{\beta}^{\mathrm{GP}}\right), \\ Var(Y|H=i) &= \left(\varphi^{\mathrm{GP}}\right)^2 \exp\left(\mathbf{x}_i^t \boldsymbol{\beta}^{\mathrm{GP}}\right) \end{aligned}$$

Table 2: The values of the score, Wald and LR statistics for testing zero-modification in the apple propagation data. The corresponding p–values are given in parenthesis.

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

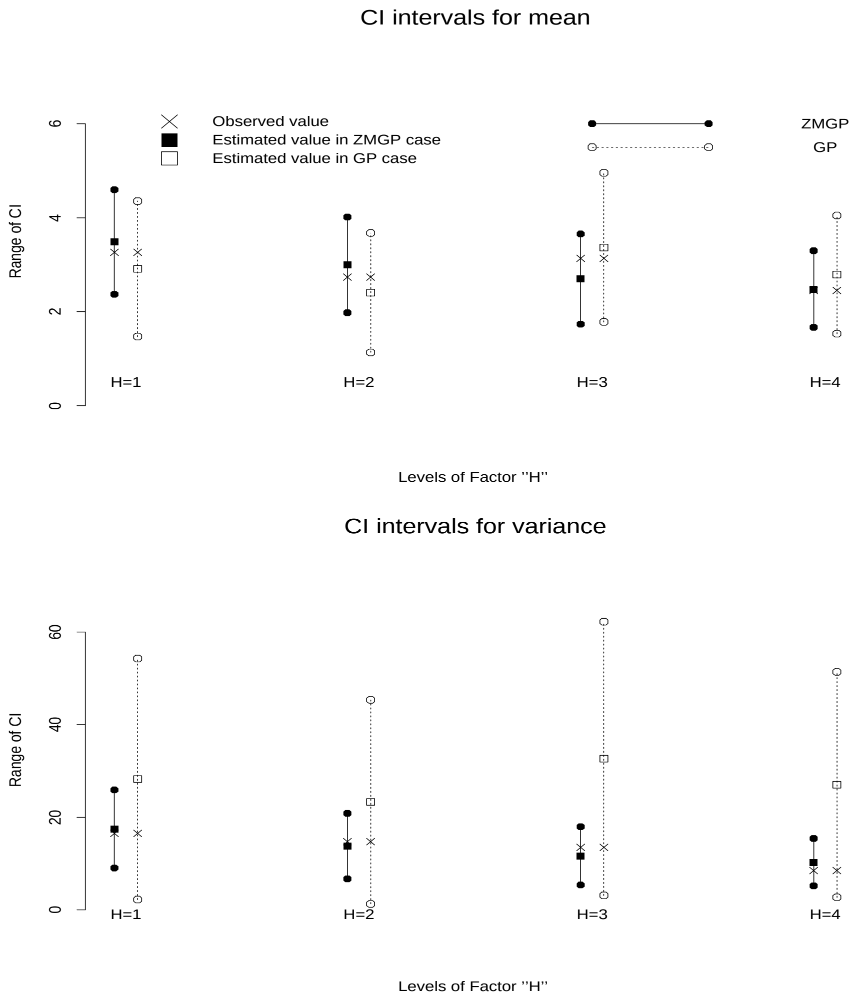
and

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

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 confidence intervals (CI) 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 the Delta method (van der Vaart (1998)) and asymptotic normality of the ML estimator $\hat{\boldsymbol{\delta}}$ in ZMGP and GP regression models (Theorem 1 and Remark (ii)).

From Figure 2 we see that CI in the ZMGP case are always shorter and predicted values for mean and variance are more closer to their empirical values than in the GP

Figure 2: Confidence intervals (CI) for the mean (top panel) and variance (bottom panel) of the partial apple propagation data for ZMGP and GP models



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 frequency of observed zero responses is here lower compared to other levels of Factor "H" (40% ($H = 3$) versus 50% ($H = 1$), 53.3% ($H = 2$) and 47.5% ($H = 4$)). The ML estimates and the corresponding asymptotic 95% confidence intervals for the zero-modification parameter ω and overdispersion parameter φ given in Table 3 also support the necessity of zero-modification in GP models for the apple propagation data.

Table 3: ML estimators 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	$(0.3749, 0.5527)$	$(1.1327, 1.6981)$

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 for the apple propagation data.

6 Conclusions and Discussions

This paper shows that the ML estimators 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 family. General results of Fahrmeir and Kaufmann (1985) for noncanonical links in GLM have been adopted for this purpose. The simulation study exhibits that the power of the score test for testing zero-modification 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 up to 87%. The effect of the poor performance of the score test in our simulation studies can be seen in the analysis of the entire apple propagation data. The score test does not detect any zero-modification despite the high proportion of zeros observed for one level of Factor "P". Note that zero-inflated count regression models are found to be appropriate for this data by Ridout et al. (2001) and Gupta et al. (2004). Therefore we conclude that score test for testing zero-modification in ZMP and ZMGP models can be highly misleading and the Wald and LR tests should be used instead.

The ZMGP regression model presented here can be generalized by allowing a regression formulation for the overdispersion parameter φ and zero-modification parameter ω. In this case nonnested testing situations with regard to the choice of covariates for the parameters μ, φ and ω will arise. A possible way to deal with this is to use the Vuong's test (Vuong (1989)). It should be noted that regression models associated with the RGP distribution will belong to this general class of regression models. Asymptotic theory for this general regression model as well as its application are under current investigation by the authors.

It is often of interest to test whether the GP regression is more appropriate for count regression data than the Poisson regression. This is the subject of our future work. The null hypothesis is here $\varphi = 1$ versus the alternative $\varphi > 1$. Note that in this testing problem the true parameter φ lies on the boundary of a parameter space and therefore we have to deal with a delicate boundary problem (see e.g. Vu and Zhou (1997)).

Appendix

The Hessian matrix $\mathcal{H}_n(\boldsymbol{\delta})$ in the ZIGP regression may be partitioned as

$$\mathcal{H}_n(\boldsymbol{\delta}) = \begin{pmatrix} \frac{\partial l_n(\boldsymbol{\delta})}{\partial\beta\beta^t} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial\beta\varphi} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial\beta\omega} \\ \frac{\partial l_n(\boldsymbol{\delta})}{\partial\varphi\beta^t} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial\varphi\varphi} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial\varphi\omega} \\ \frac{\partial l_n(\boldsymbol{\delta})}{\partial\omega\beta^t} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial\omega\varphi} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial\omega\omega} \end{pmatrix}, \tag{13}$$

where $\frac{\partial l_n(\boldsymbol{\delta})}{\partial\beta\beta^t}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial\beta\varphi}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial\beta\omega}$ are matrices of dimension $(p+1)\times(p+1)$, $(p+1)\times 1$, $(p+1)\times 1$, respectively, and $\frac{\partial l_n(\boldsymbol{\delta})}{\partial\varphi\varphi}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial\varphi\omega}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial\omega\omega}$ are scalars. Entries $h_{rs}(\boldsymbol{\delta})$'s of $\mathcal{H}_n(\boldsymbol{\delta})$ can be straightforwardly computed. For instance entries of the matrix $\frac{\partial l_n(\boldsymbol{\delta})}{\partial\beta\beta^t}$ are given by

$$\begin{aligned} h_{rs}(\boldsymbol{\delta}) &:= \frac{\partial l_n(\boldsymbol{\delta})}{\partial\beta_r\beta_s} \\ &= -\sum_{i=1}^{n} \mathbb{1}_{\{y_i=0\}} x_{ir}x_{is}(1-\omega)\mu_i(\boldsymbol{\beta})f_i(\boldsymbol{\beta},\varphi) \\ &\times \frac{[1-\mu_i(\boldsymbol{\beta})/\varphi]\, g_i(\boldsymbol{\delta}) + (1-\omega)f_i(\boldsymbol{\beta},\varphi)\mu_i(\boldsymbol{\beta})/\varphi}{\varphi\,[g_i(\boldsymbol{\delta})]^2} \\ &- \sum_{i=1}^{n} \mathbb{1}_{\{y_i>0\}} x_{ir}x_{is}\mu_i(\boldsymbol{\beta}) \left(\frac{1}{\varphi} - \frac{y_i(y_i-1)(\varphi-1)}{[\mu_i(\boldsymbol{\beta})+(\varphi-1)y_i]^2}\right) \end{aligned} \tag{14}$$

for $r, s = 0, \ldots, p$.

Now set $\mathbf{H}_n(\boldsymbol{\delta}) = -\mathcal{H}_n(\boldsymbol{\delta})$. It is well known (see for example Mardia et al. (1979), p.98) that under mild general regularity assumptions which are satisfied here that the Fisher information matrix $\mathbf{F}_n(\boldsymbol{\delta})$ is equal to $E_\delta \mathbf{H}_n(\boldsymbol{\delta})$. Thus entries of $\mathbf{F}_n(\boldsymbol{\delta})$ can be straightforwardly computed and are given by

$$\begin{aligned} f_{r,s}(\boldsymbol{\delta}) &= f_{s,r}(\boldsymbol{\delta}) = \sum_{i=1}^{n} x_{ir}x_{is}(1-\omega)\mu_i(\boldsymbol{\beta})f_i(\boldsymbol{\beta},\varphi) \\ &\times \frac{[1-\mu_i(\boldsymbol{\beta})/\varphi]\, g_i(\boldsymbol{\delta}) + (1-\omega)f_i(\boldsymbol{\beta},\varphi)\mu_i(\boldsymbol{\beta})/\varphi}{\varphi g_i(\boldsymbol{\delta})} \\ &+ \sum_{i=1}^{n}(1-\omega)x_{ir}x_{is}\mu_i(\boldsymbol{\beta})\left(\frac{\mu_i(\boldsymbol{\beta})-2\varphi+2\varphi^2}{\varphi^2(\mu_i(\boldsymbol{\beta})-2+2\varphi)} - \frac{1}{\varphi}f_i(\boldsymbol{\beta},\varphi)\right) \end{aligned}$$

for $r,s = 0,\ldots,p$;

$$\begin{aligned} f_{p+1,r}(\boldsymbol{\delta}) &= f_{r,p+1}(\boldsymbol{\delta}) = \sum_{i=1}^{n} x_{ir}(1-\omega)f_i(\boldsymbol{\beta},\varphi)\mu_i(\boldsymbol{\beta}) \\ &\times \frac{g_i(\boldsymbol{\delta})\,[\mu_i(\boldsymbol{\beta})/\varphi - 1] - (1-\omega)f_i(\boldsymbol{\beta},\varphi)\mu_i(\boldsymbol{\beta})/\varphi}{\varphi^2 g_i(\boldsymbol{\delta})} \\ &- \sum_{i=1}^{n}(1-\omega)x_{ir}\mu_i(\boldsymbol{\beta})\left(\frac{2(\varphi-1)}{\varphi^2(\mu_i(\boldsymbol{\beta})-2+2\varphi)} - \frac{f_i(\boldsymbol{\beta},\varphi)}{\varphi^2}\right) \end{aligned}$$

for $r = 0,\ldots,p$;

$$f_{p+2,r}(\boldsymbol{\delta}) = f_{r,p+2}(\boldsymbol{\delta}) = -\sum_{i=1}^{n}\frac{x_{ir}f_i(\boldsymbol{\beta},\varphi)\mu_i(\boldsymbol{\beta})}{\varphi g_i(\boldsymbol{\delta})}$$

for $r = 0,\ldots,p$;

$$\begin{aligned} f_{p+1,p+1}(\boldsymbol{\delta}) &= -\sum_{i=1}^{n}(1-\omega)f_i(\boldsymbol{\beta},\varphi)\mu_i(\boldsymbol{\beta}) \\ &\times \frac{g_i(\boldsymbol{\delta})\,(\mu_i(\boldsymbol{\beta})-2\varphi) - (1-\omega)f_i(\boldsymbol{\beta},\varphi)\mu_i(\boldsymbol{\beta})}{\varphi^4 g_i(\boldsymbol{\delta})} \\ &+ \sum_{i=1}^{n}2(1-\omega)\mu_i(\boldsymbol{\beta})\left(\frac{1}{\varphi^2(\mu_i(\boldsymbol{\beta})-2+2\varphi)} - \frac{f_i(\boldsymbol{\beta},\varphi)}{\varphi^3}\right); \end{aligned}$$

$$f_{p+2,p+1}(\boldsymbol{\delta}) = f_{p+1,p+2}(\boldsymbol{\delta}) = \sum_{i=1}^{n}\frac{f_i(\boldsymbol{\beta},\varphi)\mu_i(\boldsymbol{\beta})}{\varphi^2 g_i(\boldsymbol{\delta})}$$

and

$$f_{p+2,p+2}(\boldsymbol{\delta}) = \sum_{i=1}^{n}\left(\frac{[1-f_i(\boldsymbol{\beta},\varphi)]^2}{g_i(\boldsymbol{\delta})} + \frac{1-f_i(\boldsymbol{\beta},\varphi)}{1-\omega}\right).$$

The proof of Theorem 1 follows the proof of Theorem 4 given in Fahrmeir and Kaufmann (1985). In particular, we have to prove asymptotic normality of the normalized score vectors $\mathbf{F}_n^{t/2}\mathbf{s}_n$ (Lemma 3) and show (Lemma 4) that

$$\max_{\delta\in N_n(\varepsilon)} \|\mathbf{V}_n(\boldsymbol{\delta}) - \mathbf{I}_{p+3}\| \xrightarrow{P} 0 \qquad \text{for all} \quad \epsilon > 0,$$

where $\mathbf{V}_n(\boldsymbol{\delta}) := \mathbf{F}_n^{-1/2}\mathbf{H}_n(\boldsymbol{\delta})\mathbf{F}_n^{-t/2}$ for $n = 1, 2, \ldots$. The complex expression for the entries of the Fisher information matrix and Hessian matrix, respectively, requires more effort for proving Lemma 4 than in the case of the GLM.

First we proceed with two preliminary lemmas. Recall that we drop the dependency on $\boldsymbol{\delta}_0$,$\boldsymbol{\beta}_0$, φ_0 and use μ_i, $\mathbf{F}_n$, E, etc.

LEMMA 1. *Let $\tilde{Y}_i \sim GP(\mu_i, \varphi_0)$ for $i = 1, \ldots, n$ be a sequence of random variables. Then under assumptions* (A2) *and* (A3),

$$\max_{i=1,\ldots,n} E\left(\frac{1}{(\mu_i + (\varphi_0 - 1)\tilde{Y}_i)^k}\right) \le C_1$$

and

$$\max_{i=1,\ldots,n} E(\tilde{Y}_i^k) \le C_2$$

for any finite integer $k > 0$, where C_1 and C_2 are positive constants depending only on k and $\boldsymbol{\delta}_0$.

Proof. Let us show the first inequality of the Lemma. It is evident using (A3) that

$$E\left(\frac{1}{(\mu_i + (\varphi_0 - 1)\tilde{Y}_i)^k}\right) \le \frac{1}{\mu_i^k}. \tag{15}$$

Now it follows

$$\max_{i=1,\ldots,n} \frac{1}{\mu_i^k} = \max_{i=1,\ldots,n} \frac{1}{\exp\left(k\mathbf{x}_i^t\boldsymbol{\beta}_0\right)} \le \max_{\mathbf{x}\in K_x} \frac{1}{\exp\left(k\mathbf{x}^t\boldsymbol{\beta}_0\right)} \le C_1(\boldsymbol{\beta}_0, k),$$

since K_x is a compact and $\exp\left(k\mathbf{x}^t\boldsymbol{\beta}_0\right)$ is a continuous function of $\mathbf{x}$. It should be noted that $C_1(\boldsymbol{\beta}_0, k)$ is continuous with respect to $\boldsymbol{\beta}_0$ and well defined for all $\boldsymbol{\beta}_0 \in B$.

Now we show the second inequality of the lemma. First, we reparametrize the GP distribution by introducing new parameters $\theta_i := \mu_i/\varphi_0$ and $\lambda_0 := (\varphi_0 - 1)/\varphi_0$, $i = 1, \ldots, n$. Consul and Shenton (1974) gave the following recurrence formula for the noncentral moments of the $GP(\theta_i, \lambda_0)$ distribution:

$$(1 - \lambda_0)m_{i,k+1} = \theta_i m_{i,k} + \theta_i\frac{\partial m_{i,k}}{\partial\theta_i} + \lambda_0\frac{\partial m_{i,k}}{\partial\lambda_0}, \quad k = 0, 1, 2, \ldots,$$

where $m_{i,k} := E(\tilde{Y}_i^k)$.

Solving this recursion for fixed k shows that $m_{i,k}$ is a polynomial in θ_i, λ_0 and $1/(1 - \lambda_0)$. Thus, $m_{i,k}$ is a continuous function with respect to (θ_i, λ_0) and consequently, it is also continuous with respect to (μ_i, φ_0). It follows now that

$$\begin{aligned}
\max_{i=1,\ldots,n} E(\tilde{Y}_i^k) &= \max_{i=1,\ldots,n} m_{i,k}\left(\theta_i, \lambda_0\right) \\
&= \max_{i=1,\ldots,n} m_{i,k}\left(\mu_i/\varphi_0, (\varphi_0 - 1)/\varphi_0\right) \\
&\le \max_{\mathbf{x}\in K_x} m_k\left(e^{\mathbf{x}^t\beta_0}/\varphi_0, (\varphi_0 - 1)/\varphi_0\right) \\
&\le C_2(\boldsymbol{\delta}_0),
\end{aligned}$$

where $m_k := E(\tilde{Y}^k)$ and $\tilde{Y} \sim GP(\exp(\mathbf{x}^t\boldsymbol{\beta}_0), \varphi_0)$. It is not difficult to see that $C_2(\boldsymbol{\delta}_0)$ is continuous with respect to $\boldsymbol{\delta}_0$ and well defined for all $\boldsymbol{\delta}_0 \in K_\delta$. □

LEMMA 2. *Let $Q_k(y)$ be a polynomial of a finite order k ($k \in \mathbb{N}$) whose coefficients are positive continuous functions of $\mathbf{x}$, $\boldsymbol{\delta}$ and $\boldsymbol{\delta}_0$. Further, let $Y_i \sim ZIGP(\exp(\mathbf{x}_i^t\boldsymbol{\beta}_0), \varphi_0, \omega_0)$ for $i = 1, \ldots, n$. If* (A1)–(A3) *hold then*

$$\max_{\delta \in N_n(\varepsilon)} \max_{i=1,\ldots,n} E\left(\mathbb{1}_{\{Y_i>0\}} Q_k(Y_i)\right) < C,$$

where C is a positive constant depending on k and $\boldsymbol{\delta}_0$.

Proof. Note that under (A1) the neighborhood $N_n(\varepsilon)$ is a compact for any $n \in \mathbb{N}$ and shrinks to $\boldsymbol{\delta}_0$ for any $\varepsilon > 0$ as $n \to \infty$. Using Lemma 1 and the continuity of the coefficients of Q_k, it follows now that

$$\begin{aligned}
\max_{\delta \in N_n(\varepsilon)} \max_{i=1,\ldots,n} E\left(\mathbb{1}_{\{Y_i>0\}} Q_k(Y_i)\right) &\leq \max_{\delta \in N_n(\varepsilon)} \max_{i=1,\ldots,n} (1-\omega_0) E\left(Q_k(\tilde{Y}_i)\right) \\
&\leq \max_{\delta \in N_1(\varepsilon)} \max_{\mathbf{x} \in K_x} (1-\omega_0) E\left(Q_k(\tilde{Y})\right) \\
&\leq C,
\end{aligned}$$

where $\tilde{Y}_i \sim GP(\exp(\mathbf{x}_i^t\boldsymbol{\beta}_0), \varphi_0)$ and $\tilde{Y} \sim GP(\exp(\mathbf{x}^t\boldsymbol{\beta}_0), \varphi_0)$. □

LEMMA 3. *Under assumptions* (A1)–(A3), $\mathbf{F}_n^{-1/2}\mathbf{s}_n \overset{\mathcal{D}}{\Rightarrow} N_{p+3}(\mathbf{0}, \mathbf{I}_{p+3})$ *as $n \to \infty$, where $N_{p+3}(\mathbf{0}, \mathbf{I}_{p+3})$ is a $(p+3)$-dimensional normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_{p+3}$.*

Proof. According to the Cramer-Wald device, it is sufficient to show that a linear combination $\mathbf{a}^t\mathbf{F}_n^{-1/2}\mathbf{s}_n$ converges in distribution to $N(0, \mathbf{a}^t\mathbf{a})$ for any vector $\mathbf{a} \in \mathbb{R}^{p+3}$ ($\mathbf{a} \neq \mathbf{0}$). Without loss of generality, we set $\|\mathbf{a}\| = 1$.

Now observe that $\mathbf{s}_n$ can be written as a sum of independent random vectors, namely $\mathbf{s}_n = \sum_{i=1}^n \mathbf{s}_{ni}$, where $\mathbf{s}_{ni} = (s_{0,i}, \ldots, s_{p,i}, s_{p+1,i}, s_{p+2,i})^t$ with $s_{k,i} := s_{k,i}(\boldsymbol{\delta}_0)$ defined in (9), (10) and (11) for $k = 0, \ldots, p+2$ and $i = 1, \ldots, n$, respectively. Further, define independent random variables ξ_{in} by $\xi_{in} := \mathbf{a}^t\mathbf{F}_n^{-1/2}\mathbf{s}_{ni}$. Since $E(\xi_{in}) = 0$ and $Var\left(\sum_{i=1}^n \xi_{in}\right) = 1$, it is enough to show that the Lyapunov condition is satisfied, i.e.

$$L_s := \sum_{i=1}^n E|\xi_{in}|^s \overset{n\to\infty}{\longrightarrow} 0, \quad \text{for some } s > 2,$$

say $s = 3$ (see for example Hoffmann-Jørgensen (1994), p. 393). Noticing that $\|\mathbf{F}_n^{-1/2}\|^2 = 1/\lambda_{min}(\mathbf{F}_n)$, it follows from (A1) that

$$\begin{aligned}
L_3 &\leq \sum_{i=1}^n E\left(\left\|\mathbf{a}^t\right\|^3 \left\|\mathbf{F}_n^{-1/2}\right\|^3 \|\mathbf{s}_{ni}\|^3\right) \\
&\leq \frac{C}{n^{3/2}} \sum_{i=1}^n E\|\mathbf{s}_{ni}\|^3 \leq \frac{C}{\sqrt{n}} \max_{i=1,\ldots,n} E\|\mathbf{s}_{ni}\|^3.
\end{aligned}$$

Using an extension of the c_r-inequality given by

$$E\left|\sum_{i=1}^{m}\zeta_i\right|^k \leq m^{k-1}\sum_{i=1}^{m}E|\zeta_i|^k \qquad (\ k>1,\ k\in\mathbb{R}), \tag{16}$$

to m arbitrary random variables $\zeta_1,\ldots,\zeta_m$ (see, for example, Petrov (1995), p.58) yields that

$$E\left\|\mathbf{s}_{ni}\right\|^3 \leq C\left(E\left|s_{0,i}\right|^3+\ldots+E\left|s_{p,i}\right|^3+E\left|s_{p+1,i}\right|^3+E\left|s_{p+2,i}\right|^3\right).$$

Thus, it remains to establish that $\max_{i=1,\ldots,n}E\left|s_{r,i}\right|^3$ is uniformly bounded in n for $r=0,\ldots,p+2$. This will be shown for case $r=0,\ldots,p$. The remaining cases can be treated similarly. Without loss of generality, set $r=p$. Using now (16) with $m=2$, we have

$$\begin{aligned}\max_{i=1,\ldots,n}E\left|s_{p,i}\right|^3 &\leq 2^2\max_{i=1,\ldots,n}E\left|x_{ip}\mathbb{1}_{\{y_i=0\}}\frac{(1-\omega_0)f_i\mu_i}{\varphi_0 g_i}\right|^3\\ &+ 2^2\max_{i=1,\ldots,n}E\left(\left|x_{ip}\mathbb{1}_{\{y_i>0\}}\left(1+\frac{\mu_i(y_i-1)}{\mu_i+(\varphi_0-1)y_i}-\frac{\mu_i}{\varphi_0}\right)\right|^3\right)\\ &=: 4A_p(\boldsymbol{\delta}_0)+4B_p(\boldsymbol{\delta}_0).\end{aligned}$$

The last step in the proof is now to show that

$$A_p(\boldsymbol{\delta}_0)<C_1 \qquad \text{and} \qquad B_p(\boldsymbol{\delta}_0)<C_3, \tag{17}$$

where C_1 and C_3 are some constants depending on $\boldsymbol{\delta}_0$.

For proving (17) we note that

$$A_p(\boldsymbol{\delta}_0)\leq\max_{\mathbf{x}\in K_x}\|\mathbf{x}\|^3\left|\frac{(1-\omega_0)f_i\mu_i}{\varphi_0 g_i}\right|^3 g_i\leq C_1.$$

Let us now consider $B_p(\boldsymbol{\delta}_0)$. Simple arguments with Inequality (16), Cauchy-Schwarz inequality and Lemma 1, respectively, give

$$\begin{aligned}B_p(\boldsymbol{\delta}_0) &\leq \max_{i=1,\ldots,n}E\left((1-\omega_0)\left|x_{ir}\right|^3\cdot\left|1+\frac{\mu_i(\tilde{Y}_i-1)}{\mu_i+(\varphi_0-1)\tilde{Y}_i}-\frac{\mu_i}{\varphi_0}\right|^3\right)\\ &\leq C\max_{\mathbf{x}\in K_x}(1-\omega_0)\|\mathbf{x}\|^3\left(1^3+E\left|\frac{\mu_i(\tilde{Y}-1)}{\mu_i+(\varphi_0-1)\tilde{Y}}\right|^3+\left(\frac{\mu_i}{\varphi_0}\right)^3\right)\\ &\leq C_1(\boldsymbol{\delta}_0)+C_2(\boldsymbol{\delta}_0)\max_{\mathbf{x}\in K_x}E\left|\tilde{Y}-1\right|^3\\ &\leq C_1(\boldsymbol{\delta}_0)+C_2(\boldsymbol{\delta}_0)\max_{\mathbf{x}\in K_x}\sqrt{E\left(\tilde{Y}-1\right)^6}\\ &\leq C_3(\boldsymbol{\delta}_0),\end{aligned}$$

where $\tilde{Y}_i\sim GP\left(\mu_i,\varphi_0\right)$ for $i=1,\ldots,n$ and $\tilde{Y}\sim GP\left(\exp(\mathbf{x}^t\boldsymbol{\beta}_0),\varphi_0\right)$. □

LEMMA 4. *Under the assumptions* (A1)–(A3),

$$\max_{\delta \in N_n(\varepsilon)} \|\mathbf{V}_n(\boldsymbol{\delta}) - \mathbf{I}_{p+3}\| \xrightarrow{P} 0 \qquad \textit{for all} \quad \epsilon > 0. \tag{18}$$

Proof. It holds a.s. that

$$\begin{aligned} \|\mathbf{V}_n(\boldsymbol{\delta}) - \mathbf{I}_{p+3}\| &= \left\| \mathbf{F}_n^{-1/2} \left[\mathbf{H}_n(\boldsymbol{\delta}) - \mathbf{F}_n\right] \mathbf{F}_n^{-t/2} \right\| \\ &\leq \frac{1}{\lambda_{min}(\mathbf{F}_n)} \|\mathbf{H}_n(\boldsymbol{\delta}) - \mathbf{F}_n\| \\ &\leq \frac{C}{n} \|\mathbf{H}_n(\boldsymbol{\delta}) - \mathbf{F}_n\| \\ &\leq C \left\| \frac{1}{n} \left(\mathbf{H}_n(\boldsymbol{\delta}) - E\mathbf{H}_n(\boldsymbol{\delta})\right) \right\| + C \left\| \frac{1}{n} \left(E\mathbf{H}_n(\boldsymbol{\delta}) - \mathbf{F}_n\right) \right\|. \end{aligned}$$

Thus, conditions

$$\max_{\delta \in N_n(\varepsilon)} \left\| \frac{1}{n} \left(\mathbf{H}_n(\boldsymbol{\delta}) - E\mathbf{H}_n(\boldsymbol{\delta})\right) \right\| \xrightarrow{P} 0 \tag{19}$$

and

$$\max_{\delta \in N_n(\varepsilon)} \left\| \frac{1}{n} \left(E\mathbf{H}_n(\boldsymbol{\delta}) - \mathbf{F}_n\right) \right\| \longrightarrow 0 \tag{20}$$

imply (18).

In order to show (19) it is enough to establish that the maximum over $\boldsymbol{\delta} \in N_n(\varepsilon)$ of the absolute value of the (r, s)-element of the random matrix $[\mathbf{H}_n(\boldsymbol{\delta}) - E\mathbf{H}_n(\boldsymbol{\delta})]/n$ converges to zero in probability, i.e.

$$\max_{\delta \in N_n(\varepsilon)} \frac{|h_{rs}(\boldsymbol{\delta}) - Eh_{rs}(\boldsymbol{\delta})|}{n} \xrightarrow{P} 0.$$

Note that the Hessian matrix given in (13) has 6 different types of entries. We shall illustrate the above convergence for $h_{rs}(\boldsymbol{\delta})$'s defined in (14). The remaining cases can be treated similarly. Without loss of generality, we show

$$\max_{\delta \in N_n(\varepsilon)} \left| \frac{1}{n} \left(h_{p,p}(\boldsymbol{\delta}) - Eh_{p,p}(\boldsymbol{\delta})\right) \right| \xrightarrow{P} 0. \tag{21}$$

Let $Z_i := \mathbb{1}_{\{Y_i > 0\}} Y_i(Y_i - 1)$, $U_i(\boldsymbol{\beta}, \varphi) := \mu_i(\boldsymbol{\beta}) + (\varphi - 1)Y_i$, $q_{i,p}(\boldsymbol{\delta}) := x_{ip}^2 \mu_i(\boldsymbol{\beta})(\varphi - 1)$ and

$$v_{i,p}(\boldsymbol{\delta}) := x_{ip}^2 (1 - \omega) f_i(\boldsymbol{\beta}, \varphi) \mu_i(\boldsymbol{\beta}) \frac{\left[1 - \mu_i(\boldsymbol{\beta})/\varphi\right] g_i(\boldsymbol{\delta}) + (1 - \omega) f_i(\boldsymbol{\beta}, \varphi) \mu_i(\boldsymbol{\beta})/\varphi}{\varphi \left[g_i(\boldsymbol{\delta})\right]^2}$$

for $i = 1, \ldots, n$. It easy to see that (21) will now follow from the next three conditions:

$$\max_{\delta \in N_n(\varepsilon)} \left| \frac{1}{n} \sum_{i=1}^{n} v_{i,p}(\boldsymbol{\delta}) \left(\mathbb{1}_{\{Y_i=0\}} - E(\mathbb{1}_{\{Y_i=0\}}) \right) \right| \xrightarrow{P} 0,$$

$$\max_{\delta \in N_n(\varepsilon)} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{q_{i,p}(\boldsymbol{\delta})}{\varphi} \left(\mathbb{1}_{\{Y_i>0\}} - E(\mathbb{1}_{\{Y_i>0\}}) \right) \right| \xrightarrow{P} 0$$

$$\max_{\delta \in N_n(\varepsilon)} \left| \frac{1}{n} \sum_{i=1}^{n} q_{i,p}(\boldsymbol{\delta}) \left[\frac{Z_i}{[U_i(\boldsymbol{\beta}, \varphi)]^2} - E\left(\frac{Z_i}{[U_i(\boldsymbol{\beta}, \varphi)]^2} \right) \right] \right| \xrightarrow{P} 0. \tag{22}$$

Since they have a similar structure we only establish the validity of the last relation. It is worth to recall that the dependency on $\boldsymbol{\delta}_0$, $\boldsymbol{\beta}_0$ and φ_0 is always dropped.

Observe that the right hand side of (22) may be bounded by a sum of

$$\begin{aligned} A_n &= \max_{\delta \in N_n(\varepsilon)} \left| \frac{1}{n} \sum_{i=1}^{n} q_{i,p}(\boldsymbol{\delta}) \left(\frac{Z_i}{[U_i(\boldsymbol{\beta}, \varphi)]^2} - \frac{Z_i}{U_i^2} \right) \right|, \\ B_n &= \max_{\delta \in N_n(\varepsilon)} \left| \frac{1}{n} \sum_{i=1}^{n} q_{i,p}(\boldsymbol{\delta}) \left[E \frac{Z_i}{[U_i(\boldsymbol{\beta}, \varphi)]^2} - E\left(\frac{Z_i}{U_i^2} \right) \right] \right|, \\ D_n &= \max_{\delta \in N_n(\varepsilon)} \left| \frac{1}{n} \sum_{i=1}^{n} q_{i,p}(\boldsymbol{\delta}) \left[\frac{Z_i}{U_i^2} - E\left(\frac{Z_i}{U_i^2} \right) \right] \right|. \end{aligned}$$

For A_n we have the following bounds a.s.:

$$\begin{aligned} A_n &\le \max_{\delta \in N_n(\varepsilon)} \frac{1}{n} \sum_{i=1}^{n} \frac{|q_{i,p}(\boldsymbol{\delta}) Z_i|}{\mu_i^2(\boldsymbol{\beta}) \mu_i^2} \cdot |U_i(\boldsymbol{\beta}, \varphi) + U_i| \, |\mu_i(\boldsymbol{\beta}) - \mu_i + (\varphi - \varphi_0) Y_i| \\ &\le \max_{\delta \in N_n(\varepsilon)} \frac{1}{n} \sum_{i=1}^{n} \frac{|q_{i,p}(\boldsymbol{\delta}) Z_i|}{\mu_i^2(\boldsymbol{\beta}) \mu_i^2} \cdot |(Y_i + 1)(\mu_i(\boldsymbol{\beta}) + \mu_i + \varphi + \varphi_0 - 2)| \\ &\quad \times |\mu_i(\boldsymbol{\beta}) - \mu_i + (\varphi - \varphi_0) Y_i| \\ &\le \frac{C_1}{n} \left(\sum_{i=1}^{n} Z_i (Y_i + 1) \right) \max_{\delta \in N_n(\varepsilon)} \max_{\mathbf{x} \in K_x} \left| \exp(\mathbf{x}^t \boldsymbol{\beta}) - \exp(\mathbf{x}^t \boldsymbol{\beta}_0) \right| \\ &+ \frac{C_1}{n} \left(\sum_{i=1}^{n} Z_i Y_i (Y_i + 1) \right) \max_{\delta \in N_n(\varepsilon)} |\varphi - \varphi_0| \\ &=: AB_n + AC_n. \end{aligned} \tag{23}$$

It is not difficult to see that

$$\frac{1}{n} \sum_{i=1}^{n} Z_i (Y_i + 1)$$

converges in probability as $n \to \infty$ to

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E\left(Z_i (Y_i + 1) \right)$$

which is finite by Lemma 2.

These facts and the continuity in $\boldsymbol{\beta}$ of the function $\max_{\mathbf{x}\in K_x} \left|\exp(\mathbf{x}^t\boldsymbol{\beta}) - \exp(\mathbf{x}^t\boldsymbol{\beta}_0)\right|$ with value zero at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ yield that AB_n converges to 0 in probability as $n \to \infty$. Convergence of AC_n to 0 in probability may be proven in the same way.

Using similar arguments as above one can show that B_n converges to 0. To prove $D_n \to 0$ in probability, observe that the function $\max_{i=1,\ldots,n} |q_{i,p}(\boldsymbol{\delta}) - q_{i,p}(\boldsymbol{\delta}_0)|$ can be bounded from above by the following continuous function of $\boldsymbol{\delta}$

$$C \max_{\mathbf{x}\in K_x} \left|\exp(\mathbf{x}^t\boldsymbol{\beta})(\varphi - 1) - \exp(\mathbf{x}^t\boldsymbol{\beta}_0)(\varphi_0 - 1)\right|$$

with zero at $\boldsymbol{\delta} = \boldsymbol{\delta}_0$. The desired result now follows from the law of large numbers and standard arguments.

It remains to show (20). We will show

$$\max_{\delta\in N_n(\varepsilon)} \left|\frac{[E\mathbf{H}_n(\boldsymbol{\delta}) - \mathbf{F}_n]_{rs}}{n}\right| \to 0 \tag{24}$$

and again restrict our proof to the case $r = s = p$. It easy to see that condition (24) will follow from the next three conditions :

$$\max_{\delta\in N_n(\varepsilon)} \left|\frac{1}{n}\sum_{i=1}^{n} (v_{i,p}(\boldsymbol{\delta}) - v_{i,p})\, E(\mathbb{1}_{\{Y_i=0\}})\right| \to 0, \tag{25}$$

$$\max_{\delta\in N_n(\varepsilon)} \left|\frac{1}{n}\sum_{i=1}^{n} \left(\frac{q_{i,p}(\boldsymbol{\delta})}{\varphi} - \frac{q_{i,p}}{\varphi_0}\right) E(\mathbb{1}_{\{Y_i>0\}})\right| \to 0, \tag{26}$$

$$\max_{\delta\in N_n(\varepsilon)} \left|\frac{1}{n}\sum_{i=1}^{n} \left(q_{i,p}(\boldsymbol{\delta})E\left(\frac{Z_i}{[U_i(\boldsymbol{\beta},\varphi)]^2}\right) - q_{i,p}E\left(\frac{Z_i}{U_i^2}\right)\right)\right| \to 0. \tag{27}$$

Now we see that the same technique used for deriving (22) can be employed to establish the convergence results (25)–(27). □

Acknowledgement

Both authors gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (Cz 86/1-1). They also thank Marie Hušková and Axel Munk for helpful discussions.

References

Bae, S., F. Famoye, J. T. Wulu, A. A. Bartolucci, and K. P. Singh (2005). A rich family of generalized Poisson regression models. *Math. Comput. Simulation 69*(1-2), 4–11.

Consul, P. C. (1989). *Generalized Poisson distributions*, Volume 99 of *Statistics: Textbooks and Monographs*. New York: Marcel Dekker Inc. Properties and applications.

Consul, P. C. and F. Famoye (1992). Generalized Poisson regression model. *Comm. Statist. Theory Methods 21*(1), 89–109.

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

Consul, P. C. and L. R. Shenton (1974). On the probabilistic structure and properties of discrete Lagrangian distribution. In G. Patil, S. Kotz, and J. Ord (Eds.), *A modern course on statitical distributions in scientific work*, Volume 1, pp. 41–57. D.Reidel Publishing Company, Boston.

Consul, P. C. and M. M. Shoukri (1985). The generalized Poisson distribution when the sample mean is larger than the sample variance. *Comm. Statist. Simulation Comput. 14*(3), 667–681.

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

Dietz, E. and D. Böhning (2000). On estimation of the Poisson parameter in zero-modified Poisson models. *Comput. Statist. Data Anal. 34*(4), 441–459.

Fahrmeir, L. and H. Kaufmann (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist. 13*(1), 342–368.

Famoye, F. (1993). Restricted generalized Poisson regression model. *Comm. Statist. Theory Methods 22*(5), 1335–1354.

Famoye, F. and K. P. Singh (2003). On inflated generalized Poisson regression models. *Adv. Appl. Stat. 3*(2), 145–158.

Famoye, F. and K. P. Singh (2006). Zero-inflated generalized Poisson model with an application to domestic violence data. *Journal of Data Science 4*(1), 117–130.

Gschlößl, S. and C. Czado (2005). Modelling count data with overdispersion and spatial effects. *Discussion paper 412 of SFB 386* (*http://www.stat.uni-muenchen.de/sfb386/*).

Gupta, P. L., R. C. Gupta, and R. C. Tripathi (2004). Score test for zero inflated generalized Poisson regression model. *Comm. Statist. Theory Methods 33*(1), 47–64.

Hall, D. B. and K. S. Berenhaut (2002). Score tests for heterogeneity and overdispersion in zero-inflated Poisson and binomial regression models. *Canad. J. Statist. 30*(3), 415–430.

Hoffmann-Jørgensen, J. (1994). *Probability with a view toward statistics. Vol. I.* Chapman & Hall Probability Series. New York: Chapman & Hall.

Jansakul, N. and J. P. Hinde (2002). Score tests for zero-inflated Poisson models. *Comput. Statist. Data Anal. 40*(1), 75–96.

Joe, H. and R. Zhu (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biom. J. 47*(2), 219–229.

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

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *Canad. J. Statist. 15*(3), 209–225.

Mardia, K. V., J. T. Kent, and J. M. Bibby (1979). *Multivariate analysis.* London: Academic Press [Harcourt Brace Jovanovich Publishers]. Probability and Mathematical Statistics: A Series of Monographs and Textbooks.

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

Mullahy, J. (1986). Specification and testing of some modified count data models. *J. Econometrics 33*(3), 341–365.

Petrov, V. V. (1995). *Limit theorems of probability theory sequences of independent variables.* Oxford studeis in probability. Oxford: Clarendon Press.

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

Ridout, M., J. Hinde, and C. G. B. Demétrio (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics 57*(1), 219–223.

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

van den Broek, J. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics 51*(2), 738–743.

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

Verbeke, G. and G. Molenberghs (2003). The use of score tests for inference on variance components. *Biometrics 59*(2), 254–262.

Vu, H. T. V. and S. Zhou (1997). Generalization of likelihood ratio tests under nonstandard conditions. *Ann. Statist. 25*(2), 897–916.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica 57*(2), 307–333.