Zero-inflated generalized Poisson regression models: Asymptotic theory and applications

CLAUDIA CZADO and ALEKSEY MIN

Center for Mathematical Sciences

Munich University of Technology

Boltzmannstr. 3

85747 Garching, Germany

Abstract: Poisson regression models for count variables have been utilized in many applications. However, in many problems overdispersion and zero-inflation occur. In this paper we study regression models associated with the generalized Poisson distribution (Consul (1989)). These regression models which have been used for about 15 years do not belong to the class of generalized linear models considered by McCullagh and Nelder (1989) for which an established asymptotic theory is available. We prove consistency and asymptotic normality of the maximum likelihood estimators in zero-inflated generalized Poisson regression models. Further the accuracy of the asymptotic normality approximation is investigated through a simulation study. It is also shown that a Wald test for detecting zero-inflation or zero-deflation based on our results is considerable more powerful than the score test in zero-modified Poisson regression models. The usefulness of the considered models is demonstrated in two applications.

1

1. INTRODUCTION

Poisson regression models are often used for an analysis of count data. However count regression data often exhibit substantial overdispersion which is present when the data has higher variability as it is allowed by the model. In particular, equality of mean and variance for count data analyzed under a Poisson assumption is often violated. Various reasons, e.g. unobserved heterogeneity, missing covariates or correlation among the measurements, make counts overdispersed (see Cameron and Trivedi (1998) or Winkelmann (2003)). Consequently, a number of different regression models in the literature have been proposed, which commonly handle overdispersion in two general approaches:

1) inclusion of random effects;

2) extension of the parametric model by extra parameters to allow for a more general variance structure.

Using the first approach Dean and Lawless (1989) treated overdispersion in the Poisson regression and investigated score tests for its detection. There are also several papers using more general setups in this direction (see e.g. Dean (1992), Lin (1997), Hall and Præstgaard (2001), Hall and Berenhaut (2002), Deng and Paul (2000), Deng and Paul (2005)). However, it should be noted here that there exists a confusion with regard to the limiting distribution of the score test statistics in the literature. In particular the problem of testing parameters on the boundary of the parameter space needs to be addressed when one sided alternatives are considered. For insightful discussions on this problem we would like to refer to Verbeke and Molenberghs (2003).

The second approach, which we follow in this paper, consists of considering a distribution with a more flexible variance function. A negative binomial (NB) and a generalized Poisson (GP) distributions are standard count distributions used for this purpose. Lawless (1987) first systematically studied the NB

2

regression model and showed asymptotic normality of its maximum likelihood (ML) estimator. Consul and Famoye (1992) introduced the GP regression model and applied it to several data sets. However asymptotic properties of the ML estimator in the GP regression have not been investigated and this has remained an open problem.

It became popular over the past decade to model count data with a large frequency of zeros using a mixture of a count distribution with a degenerate distribution supported at zero. This is another way to handle overdispersion caused by a large amount of zeroes. Zero-inflated Poisson (ZIP) regression is one of frequently used models for such count data. Here Lambert (1992) first investigated the asymptotic properties of the ML estimator. Further Jansakul and Hinde (2002) derived score tests for testing zero-inflation in ZIP models and investigated their power in a simulation study.

Recently, several authors (see Famoye and Singh (2003), Gupta, Gupta, and Tripathi (2004), Stekeler (2004)) have independently introduced zero-inflated generalized Poisson (ZIGP) regression models which can now handle overdispersed count data with a high incidence of zero outcomes. Famoye and Singh (2003) and Gupta, Gupta, and Tripathi (2004) also discussed score tests for testing overdispersion or zero-inflation in this regression model. Again, asymptotic properties of the ML estimator in the ZIGP regression model have not been investigated.

The objective of this paper is to derive the appropriate asymptotic theory for ZIGP regression models and to examine the accuracy of the normal approximation for the ML estimator. Our results also remain valid for GP and ZIP regression models. The paper is organized as follows. In Section 2 we introduce a zero-inflated count distribution and the GP distribution and discuss their basic properties. The ZIGP regression model will be defined in Section 3. Section 4 gives the asymptotic existence, the consistency and the asymptotic normality of the ML estimator in ZIGP regression model. The

accuracy of the asymptotic distributional approximation in small samples is investigated in a simulation study presented in Section 5. In Section 6 two applications are given, while in Section 7 we compare the score test for detecting zero-inflation or zero-deflation in a zero-modified Poisson (ZMP) model (see Dietz and Böhning (2000)) to the Wald test based on our asymptotic results in a simulation study. The paper closes with a discussion section. The Fisher information matrix for ZIGP models and the proof of Theorem 1 are given in Appendices 1 and 2, respectively.

2. ZERO-INFLATED COUNT DISTRIBUTIONS AND THE GP DISTRIBUTION

Suppose that we observe realizations of a count random variable Y and we believe that Y has a specified discrete count distribution. Further suppose that the observed data exhibits an excess of zeros which can not be modelled by the assumed model. This means that we cannot rely anymore on our hypothesis. But an assumption, that zeros arise from a mixture of a Bernoulli distribution and the conjectured distribution, makes it possible for us to investigate our conjecture. More precisely, we assume that the probability mass function of the observed response Y is given by

$$P(Y = y) = \begin{cases} \omega + (1 - \omega)P(\tilde{Y} = 0) & y = 0, \\ (1 - \omega)P(\tilde{Y} = y) & y = 1, 2, \ldots, \end{cases} \qquad 0 \leq \omega \leq 1,$$

where $\tilde{Y}$ is distributed according to the conjectured distribution with finite second moment. Simple calculations show that mean and variance of the zero-inflated random variable Y are given by

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

4

and

$$Var(Y) = (1 - \omega)Var(\tilde{Y}) + \omega(1 - \omega)\left(E(\tilde{Y})\right)^2. \tag{2}$$

Throughout this paper, we assume that the conjectured distribution of the response variable Y, i.e. the distribution of $\tilde{Y}$, is a GP distribution with two parameters μ and φ denoted by $GP(\mu, \varphi)$. This distribution was first introduced by Consul and Jain (1970) and subsequently studied in detail by Consul (1989). The probability mass function of the GP distribution is given by

$$P_{\mu,\varphi}(y) := \begin{cases} \mu(\mu + y(\varphi - 1))^{y-1}\varphi^{-y}e^{-(\mu+y(\varphi-1))/\varphi}/y! & \text{for } y = 0, 1, \ldots \\ 0 & \text{for } y > m, \quad \text{when } \varphi < 1 \end{cases} \tag{3}$$

and its real-valued parameters μ and φ satisfy the following constraints:

- $\mu > 0$;

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

If $\varphi < 1$ then (3) 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.

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 $Y \sim GP(\mu, \varphi)$ then mean and variance of Y are given by

$$E(Y) = \mu \tag{4}$$

and

$$Var(Y) \;=\; \varphi^2 \mu. \tag{5}$$

In contrast to the GP distribution, we have for the NB distribution with mean μ and overdispersion parameter a (see Lawless (1987) for notation) that its variance is equal to $\mu(1 + a\mu)$. Thus the overdispersion in the GP case is independent of the mean while this is not the case for the NB case. This implies for the NB case that overdispersion might be present even if the overdispersion parameter a is small. Already Lawless (1987) remarked this fact. 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 visual comparison of GP and NB probability functions is given for example in Gschlößl and Czado (2005).

3. ZIGP REGRESSION

A random variable Y is said to be distributed according to a ZIGP distribution with parameters μ, φ and ω, denoted by $ZIGP(\mu, \varphi, \omega)$, if its probability mass function is given by

$$
\begin{aligned}
P_{\mu,\varphi,\omega}(y) \;:=\; & P(Y = y) \\
=\; & \begin{cases} \omega + (1 - \omega)P_{\mu,\varphi}(0), & \text{if } y = 0 \\ (1 - \omega)P_{\mu,\varphi}(y), & \text{if } y = 1, 2, \ldots, \\ 0 & \text{for } y > m \quad \text{when } \varphi < 1, \end{cases}
\end{aligned} \tag{6}
$$

and zero otherwise, where $0 \le \omega \le 1$, $\mu > 0$, $\varphi \ge \max\{1/2, 1 - \mu/m\}$ and m $(m \ge 4)$ is a largest natural number for which $\mu + m(\varphi - 1) > 0$ when

$\varphi < 1$. Thus, the ZIGP distribution is a mixture of a Bernoulli distribution with parameter $1 - \omega$ and the GP distribution with parameters μ and φ. Equations (1), (2), (4) and (5) imply that mean and variance of the ZIGP distribution are connected with its parameters as follows

$$E(Y) \;=\; (1 - \omega)\mu \tag{7}$$

and

$$Var(Y) \;=\; E(Y)\left(\varphi^2 + \mu\omega\right). \tag{8}$$

One of the main benefits of considering a regression model based on the ZIGP 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 the zero-inflated Poisson regression when $\varphi = 1$. Moreover, by virtue of (7) and (8) this regression can be used to fit count regression data exhibiting overdispersion or underdispersion.

Analogously to GLM, we now introduce a regression model with response Y_i and (known) explanatory variables $\mathbf{x}_i = (x_{i0}, x_{i1}, \ldots, x_{ip})^t$ with $x_{i0} = 1$ for $i = 1, \ldots, n$:

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

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

3. *Parametric link component:*
   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 and in the sequel, $\mathbf{A}^t$ and $\mathbf{a}^t$ denote the transpose of the matrix $\mathbf{A}$ and the 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 ZIGP regression model. Further, we denote the joint vector of the regression parameters $\boldsymbol{\beta}$ and the parameters φ and ω of the ZIGP 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 $i = 1, \ldots, n$ 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 $y_1, \ldots, y_n$, the log-likelihood $l(\boldsymbol{\delta})$ derived from the ZIGP regression can be written as

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

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

$$
\mathbf{s}_n(\boldsymbol{\delta}) = (s_0(\boldsymbol{\delta}), \ldots, s_p(\boldsymbol{\delta}), s_{p+1}(\boldsymbol{\delta}), s_{p+2}(\boldsymbol{\delta}))^t, \tag{9}
$$

where

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

with

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

8

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

$$
\begin{aligned}
s_{p+1,i}(\boldsymbol{\delta}) \ := \ & \mathbb{1}_{\{y_i=0\}} \frac{(1-\omega) f_i(\boldsymbol{\beta}, \varphi) \mu_i(\boldsymbol{\beta})}{\varphi^2 g_i(\boldsymbol{\delta})} \\
& + \ \mathbb{1}_{\{y_i>0\}} \left(\frac{y_i(y_i-1)}{\mu_i(\boldsymbol{\beta}) + (\varphi-1)y_i} - \frac{y_i}{\varphi} + \frac{\mu_i(\boldsymbol{\beta}) - y_i}{\varphi^2} \right),
\end{aligned}
\tag{11}
$$

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

with

$$s_{p+2,i}(\boldsymbol{\delta}) \ := \ \mathbb{1}_{\{y_i=0\}} \frac{1 - f_i(\boldsymbol{\beta}, \varphi)}{g_i(\boldsymbol{\delta})} - \mathbb{1}_{\{y_i>0\}} \frac{1}{1 - \omega}, \tag{12}$$

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

To compute the ML estimator $\hat{\boldsymbol{\delta}}$, we solve simultaneously the equations obtained by equating the score vector (9) to zero.

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 ZIGP regression.

Analogously to 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 ZIGP regression will be denoted as $\boldsymbol{\delta}_0$. Further $\mathbf{F}_n(\boldsymbol{\delta})$ will stand for the Fisher information matrix in a ZIGP regression evaluated at $\boldsymbol{\delta}$.

Now denote by

$$N_n(\varepsilon) = \{\boldsymbol{\delta} : \|\mathbf{F}_n^{t/2}(\boldsymbol{\delta}_0)(\boldsymbol{\delta} - \boldsymbol{\delta}_0)\| \leq \varepsilon\} \tag{13}$$

a neighborhood of $\boldsymbol{\delta}_0$ 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_{δ_0}, E_{δ_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 k-dimensional unit matrix will be denoted by $\mathbf{I}_k$ and an admissible set for a regression parameter $\boldsymbol{\beta}$ will be denoted by B.

In the paper 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 $\Omega := [0, 1]$.

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)*.

The proof is given in Appendix 2.

REMARKS

(i) Assumption (A1) is more restrictive than the corresponding condition (D) of Fahrmeir and Kaufmann (1985).

(ii) Assumption (A2) simply means that we deal with compact regressors.

(iii) 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. However, one may investigate asymptotic properties of the ML estimator $\hat{\boldsymbol{\delta}}$ using results of Self and Liang (1987) and Moran (1971).

(iv) It is not difficult to see that the asymptotic results of Theorem 1 remain valid in GP or ZIP regression models subject to appropriate changes are performed in the log-likelihood, the ML equations and the Fisher information matrix as well as in Assumption (A3).

(v) A close look at the proof of Theorem 1 reveals that $\Omega := [0, 1]$ in Assumption (A3) can, in fact, be replaced by $\Omega_1 := [-c_\omega, 1]$, where $c_\omega > 0$ is a constant depending on K_x and $\boldsymbol{\delta}_0$ such that $P_{\mu,\varphi,\omega}(0)$ from (6) is a still probability. But then we can not interpret ω as a probability anymore (see e.g. Dietz and Böhning (2000)).

(vi) Looking at inequality (15) we see that an analogous upper bound can be obtained when $\varphi < 1$ if the admissible set B for regression parameters $\boldsymbol{\beta}$ is a compact set. This implies that the resulting μ_i's are bounded from below by a positive constant c_μ for $i = 1, \ldots, n$. Then an extension of the parameter space for φ to $\Phi = [c_\varphi, \infty]$ with $c_\varphi < 1$ may be also accomplished.

5. SIMULATION STUDY

We investigated the accuracy of the normal approximation based on Theorem 1 by performing a small simulation study in S-PLUS for samples of size $n = 50, 100$ and 200. We used a similar simulation setup as Stekeler (2004). It should be noted here that the first maximization routine has been written by Stekeler (2004) and we further updated it. A simple model with intercept and single covariate x was considered for the linear predictors $\eta_i(\boldsymbol{\beta})$'s, i.e. $\eta_i(\boldsymbol{\beta}) = \beta_0 + \beta_1 x_i$ for $i = 1, \ldots, n$. The values of the covariate x were chosen equally spaced between -1 and 1. Further we examined two choices for β_1 and set $\beta_0 = -1$. In the first case we put $\beta_1 = 2$ while $\beta_1 = 3$ was set in the second case, which will be in the sequel called Setting-1 and Setting-2, respectively. This allows us to compare models with a small (Setting-1) and large (Setting-2) range of the parameter μ of the ZIGP distribution. Since we are mostly interested in the case when Poisson regression does not satisfactorily fit the count regression data, the following values of ω and φ were considered: $\omega = 0.1, 0.25$ and $\varphi = 1.25, 3$. For each combination of sample size n, setting, ω and φ we simulated 100 samples of responses Y_i's, i.e. $Y_i \sim ZIGP(\exp(\beta_0 +$

$\beta_1 x_i), \varphi, \omega)$ for $i = 1, \ldots, n$.

We computed the average estimate and the estimated mean squared error (MSE) of the ML estimators $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\varphi}$ and $\hat{\omega}$ in 100 replications for each considered case. Simulation results for Setting-1 are a bit more accurate than for Setting-2 but they demonstrate similar patterns. This is natural to expect since μ has an influence on the range the of data. Here we present the results only for Setting-2 given in Table 1. Standard errors of the average estimate and estimated MSE are given in parentheses. From Table 1 we see as expected that the bias and MSE always decrease as the sample size n increases. An opposite pattern is observed with respect to φ. If φ increases, while n and ω remain fixed, the accuracy of the estimates becomes worse. This is explained by allowing for more dispersed data for larger φ. A similar pattern holds for ω. If ω increases, while n and φ remain fixed, then the accuracy becomes worse. Since a larger ω increases the overdispersion in the data this is to be expected. Note that in our simulation study φ has a larger influence on the accuracy of ML estimators than ω. This can be seen from the estimated MSE's. For instance, the estimated MSE of $\hat{\beta}_1$ is equal to 0.09 when $\varphi = 1.25$, $\omega = 0.1$ and $n = 200$. Now if ω is increased by 2.5 times then the estimated MSE approximately increases 20% while if φ is increased by 2.4 times then the estimated MSE approximately increases 110%.

To draw a normal quantile-quantile (QQ) plot for the empirical distribution of each component of the random vector $\mathbf{F}_n^{t/2}(\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0)$ considered in Theorem 1 and the standard normal distribution, the ML estimators $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\varphi}$ and $\hat{\omega}$ were centered by the corresponding true value and normalized by the corresponding square root of diagonal element of the inverse of the Fisher information matrix evaluated at the vector of true parameter values $(\varphi, \omega, \beta_0, \beta_1)$. The normalized and centered ML estimators are further denoted by $\hat{\beta}_0^{st}$, $\hat{\beta}_1^{st}$, $\hat{\varphi}^{st}$ and $\hat{\omega}^{st}$. Figures 1 ($\omega = 0.1$) and 2 ($\omega = 0.25$) display the QQ–plots for Setting-2. For a better visualization we connected points of QQ–plots with

13

Para-meter	True value	n	Estimate		MSE		Para-meter	True value	n	Estimate		MSE	
φ	1.25	50	1.222	(0.214)	0.047	(0.009)	φ	1.25	50	1.208	(0.213)	0.047	(0.007)
		100	1.219	(0.148)	0.023	(0.003)			100	1.190	(0.151)	0.026	(0.004)
		200	1.238	(0.122)	0.015	(0.002)			200	1.207	(0.126)	0.018	(0.002)
ω	0.1	50	0.112	(0.110)	0.012	(0.002)	ω	0.25	50	0.259	(0.147)	0.022	(0.003)
		100	0.104	(0.081)	0.007	(0.001)			100	0.264	(0.095)	0.009	(0.001)
		200	0.096	(0.056)	0.003	$(4\cdot10^{-4})$			200	0.250	(0.072)	0.005	(0.001)
β_0	-1	50	-1.018	(0.438)	0.192	(0.029)	β_0	-1	50	-1.206	(0.591)	0.391	(0.058)
		100	-1.044	(0.310)	0.098	(0.011)			100	-1.095	(0.357)	0.136	(0.023)
		200	-1.064	(0.237)	0.060	(0.010)			200	-1.027	(0.264)	0.070	(0.013)
β_1	3	50	2.976	(0.512)	0.263	(0.034)	β_1	3	50	3.228	(0.689)	0.526	(0.077)
		100	3.083	(0.406)	0.172	(0.019)			100	3.122	(0.464)	0.230	(0.037)
		200	3.065	(0.293)	0.090	(0.013)			200	3.022	(0.327)	0.107	(0.019)
Para-meter	**True value**	**n**	**Estimate**		**MSE**		**Para-meter**	**True value**	**n**	**Estimate**		**MSE**	
φ	3	50	2.672	(1.404)	2.079	(0.324)	φ	3	50	2.563	(1.255)	1.765	(0.214)
		100	2.865	(0.914)	0.853	(0.143)			100	2.936	(0.957)	0.921	(0.113)
		200	2.915	(0.514)	0.272	(0.037)			200	2.933	(0.648)	0.424	(0.061)
ω	0.1	50	0.165	(0.184)	0.038	(0.006)	ω	0.25	50	0.258	(0.220)	0.048	(0.005)
		100	0.117	(0.137)	0.019	(0.004)			100	0.227	(0.181)	0.033	(0.003)
		200	0.090	(0.095)	0.009	(0.001)			200	0.251	(0.129)	0.017	(0.002)
β_0	-1	50	1.382	(1.107)	1.372	(0.372)	β_0	-1	50	-1.423	(1.110)	1.412	(0.319)
		100	1.137	(0.606)	0.386	(0.075)			100	-1.241	(0.674)	0.512	(0.082)
		200	1.127	(0.388)	0.166	(0.029)			200	-1.095	(0.440)	0.203	(0.046)
β_1	3	50	3.430	(1.256)	1.762	(0.534)	β_1	3	50	3.399	(1.275)	1.784	(0.430)
		100	3.207	(0.720)	0.561	(0.116)			100	3.259	(0.748)	0.626	(0.103)
		200	3.120	(0.414)	0.186	(0.031)			200	3.141	(0.537)	0.308	(0.086)

Table 1: Average estimate and estimated MSE of $\hat{\varphi}$, $\hat{\omega}$, $\hat{\beta}_0$, $\hat{\beta}_1$ in Setting-2 for a *ZIGP* regression model on the basis of 100 replications replications (estimated standard errors are given in parentheses).
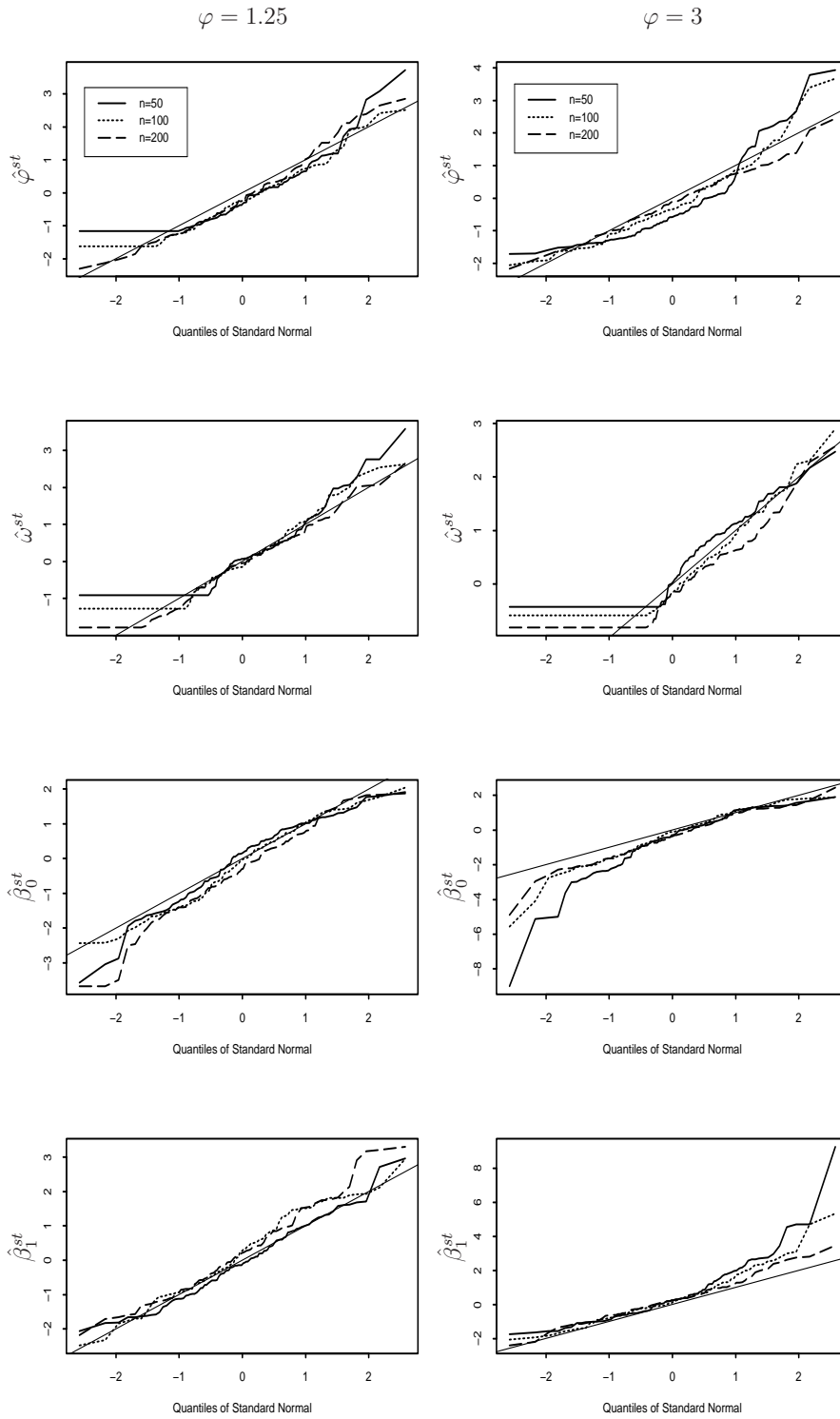
Figure 1: Normal QQ-plots of centered and normalized ML estimators in Setting-2 for a $ZIGP$ regression model with $\omega = 0.1$ based on 100 replications
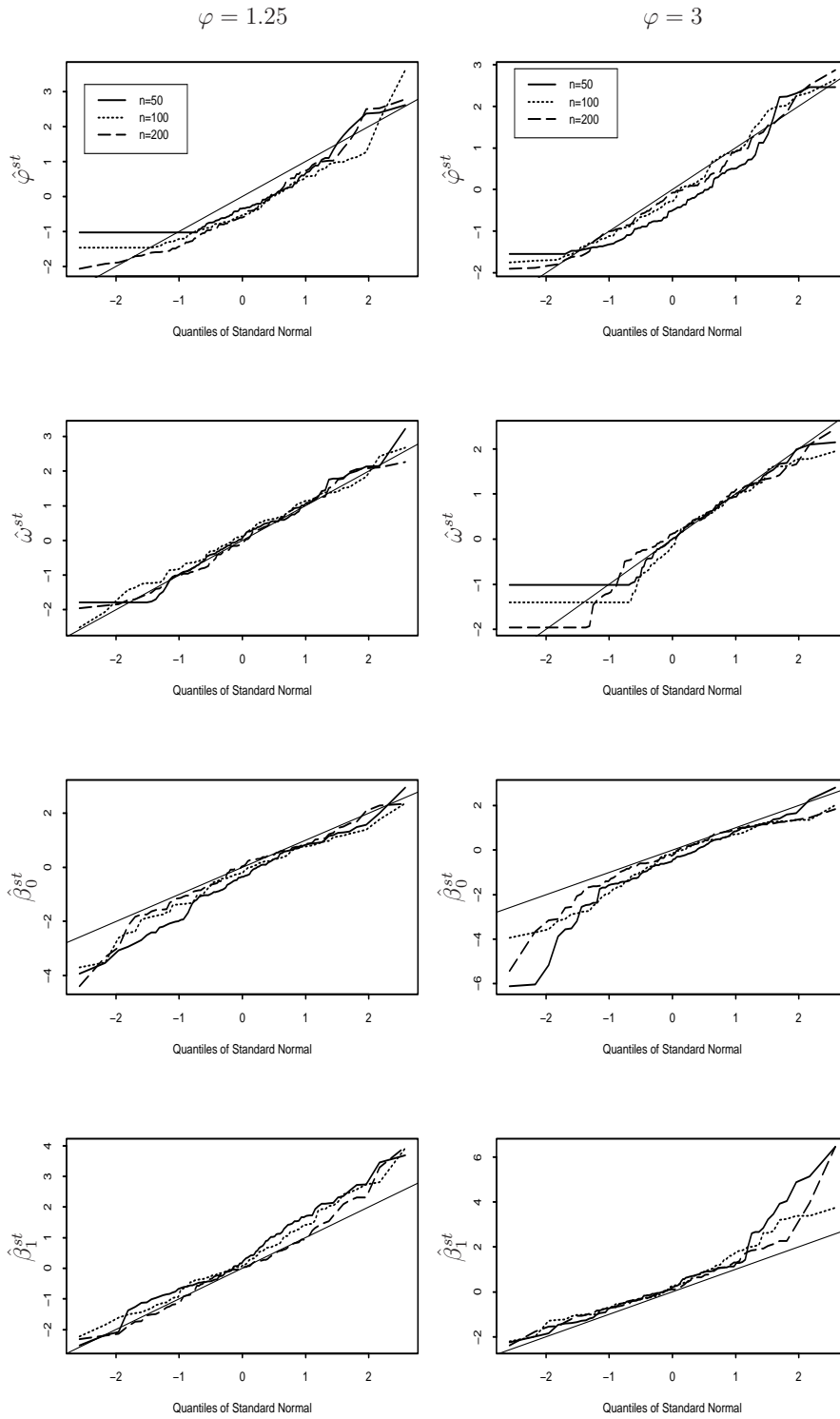
15

Figure 2: Normal QQ-plots of centered and normalized ML estimators in Setting-2 for a $ZIGP$ regression model with $\omega = 0.25$ based on 100 replications

16

different type of lines. The solid, dotted and dashed broken lines correspond to sample sizes $n = 50$, $n = 100$ and $n = 200$, respectively. The straight line corresponds to 45^0 degree line and indicates where the points of a standard normal distribution in a normal QQ–plot would fall.

From these plots we see that the normal approximation for $\hat{\beta}_0^{st}$ and $\hat{\beta}_1^{st}$ is quite satisfactory. This is only partially true for $\hat{\varphi}^{st}$ and $\hat{\omega}^{st}$ since we observe horizontal segments in the left bottom corner of the corresponding QQ-plot. A reason of the above anomaly is the closeness of the true values of φ and ω to their left boundary values 1 and 0, respectively. Therefore the log-likelihood reaches its maximum at $\varphi = 1 + 10^{-99}$ and $\omega = 10^{-99}$ which are the lower bound for φ and ω in the maximization routine. The standard normal QQ-plots for $\hat{\varphi}^{st}$ in Figure 1 and $\hat{\omega}^{st}$ in Figure 2 illustrate this fact. Note that the normal approximation for $\hat{\omega}^{st}$ in Figure 2 is worse for $\varphi = 3$ than for $\varphi = 1.25$. This occurs since data becomes more dispersed for large φ. The above anomaly is resolved when a higher sample size is used in these cases. This can be seen by comparing the first column of QQ-plots in Figure 1 with the corresponding QQ-plots in Figure 3 for $n = 500$. It should be noted here that the ML
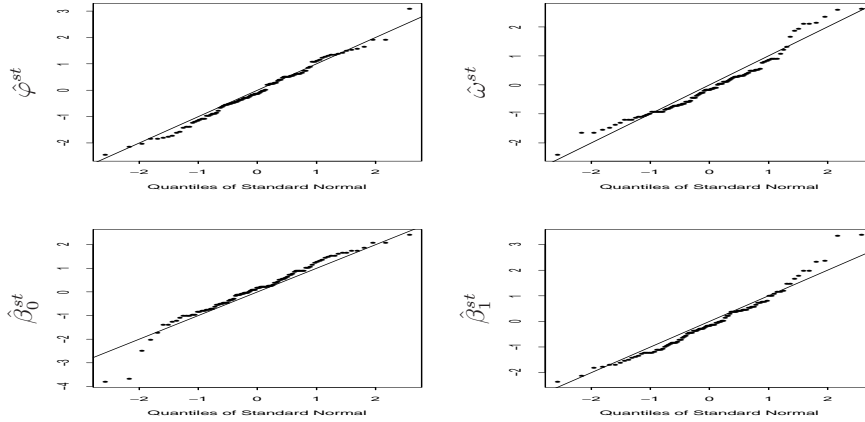

Figure 3: Normal QQ-plots of centered and normalized ML estimators in Setting-2 for a $ZIGP$ regression model with $n = 500$, $\omega = 0.1$ and $\varphi = 1.25$ based on 100 replications

17

estimators in GP and ZIP regressions have exhibited analogous asymptotic properties.

We also investigated the coverage of the true values of the parameters φ, ω, β_0 and β_1 of asymptotic confidence intervals based on Theorem 1 for sample size $n = 500$. Results of simulations showed good agreement with true values of φ, ω, β_0 and β_1.

6. EXAMPLES

6.1. PATENT DATA

Using a negative binomial regression model Czado and Sikora (2002) analyzed data on the number of patents of US high-tech firms in 1976 from Wang, Cockburn, and Puterman (1998). Czado and Sikora (2002) rejected the Poisson regression model in favor of negative binomial model using a p–value curve approach. We applied a ZIGP regression model to their model setup for the patent data. Since $[2.098, 3.263]$ and $[-0.03, 0.06]$ were obtained as asymptotic 95% confidence intervals for φ and ω, respectively, we see that zero-inflation is not present in this data set while there is overdispersion. Thus, the GP regression may be more appropriate than the ZIGP regression. An application of a GP regression to the patent data also confirms this decision since it produces $[2.152, 3.342]$ as an asymptotic 95% confidence interval for φ. In order to compare the NB model and the GP model we used Vuong's test which is applicable to nonnested models (see Vuong (1989)). It is based on the Kullback-Leibler distance (see Kullback and Leibler (1951)) and uses the asymptotic theory of the maximum likelihood estimator in misspecified models developed by White (1982).

Let $P_{\mu_i,\varphi}(y_i)$ denote the estimated probability that a random variables Y equals y_i under assumption that its distribution is $GP(\mu_i, \varphi)$. Analogously, $Q_{\mu_i,a}(y_i)$ is the estimated probability under assumptions that a random variable Y distributed according to the NB distribution with parameters (μ_i, a) (cf.

Lawless (1987)). Then Vuong's statistic V for testing the hypothesis of the NB model versus the GP model is constructed as follows:

$$V = \frac{\sqrt{n}\overline{m}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \overline{m})^2}},$$

where

$$m_i = \log\left(\frac{Q_{\mu_i,a}(y_i)}{P_{\mu_i,a}(y_i)}\right) \quad \text{and} \quad \overline{m} = \frac{1}{n}\sum_{i=1}^{n} m_i.$$

The test statistic V for testing the NB regression model versus the GP regression model in the patent data is equal to 0.426 with asymptotic p-value 0.670. Therefore we prefer neither the NB model nor the GP model, say, even at 10%-level. It should be noted that Vuong's statistic is bidirectional (see Vuong (1989)). Here a large positive value of the statistic would favor the NB model while a small negative value would favor the GP model.

6.2. APPLE SHOOT PROPAGATION DATA

Ridout, Demétrio, and Hinde (1998) analyzed data on the number of roots produced by 270 shoots of a certain apple cultivar. The shoots had been produced under an 8– and 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 Marin, Jones, and Hadlow (1993)). Since the data contains a large number of zero responses for the 16 hour photoperiod the extension to allow for zero-inflation is natural to consider. Ridout, Demétrio, and Hinde (1998) fitted Poisson and NB regression models as well as their zero-inflated copies with various combinations of covariate specifications to this data set. Here we consider specifically two ZIGP models which will be compared with corresponding ZINB models fitted by the the later authors. In the first model μ may take different values only for two levels of Factor "P", while in the second model μ may take different values for each of the eight treatment combination ("P∗H"). The choice of

19

these models is dictated by the fact that there is a large effect of Factor "P" and significant interaction between Factors "P" and "H". This follows from considering ZIGP and ZINB regression specifications using a likelihood ratio test. Overdispersion and zero-inflation parameters are taken to be a constant in the both models.

For one factor case asymptotic 95% confidence intervals for φ and ω are $(1.143, 1.414)$ and $(0.171, 0.273)$, respectively. Thus overdispersion and zero-inflation are present and ZIGP regression approach would be appropriate here. In the second case, asymptotic 95% confidence intervals for φ and ω are $(1.112, 1.373)$ and $(0.172, 0.274)$, respectively. Again, we see that ZIGP regression approach is suitable for these data. Moreover, Table 2 shows that Vuong's test would slightly favor the ZIGP model over the ZINB model for two factor "P∗H" mean specification.

Table 2: Vuong's statistic and its p–value for data on shoots of a apple cultivar.

Compared models	Vuong's statistic V	p-value
One Factor "P": ZINB vs. ZIGP	-1.391	0.164
Two Factors "P∗H": ZINB vs. ZIGP	-1.800	0.072

7. POWER COMPARISON OF SCORE AND WALD TESTS IN ZMP MODELS

Recently Jansakul and Hinde (2002) investigated performance of the score test for zero-inflation in small and moderate sample sizes within the ZIP regression model. As they noted, the score test, in fact, gives here a test of Poisson model against ZMP model (see Dietz and Böhning (2000)) avoiding the problem of testing on the boundary of zero-inflation.

By virtue of Theorem 1 and Remark (v), we can construct the Wald test for testing zero-inflation or zero-deflation and then compare its perfor-

mance with the performance of the corresponding score test simulated by Jansakul and Hinde (2002). At the moment, this can be done only for models with constant zero-inflation parameter. In particular, they 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 $(0, 1)$. Note that for each combination of sample size and model they simulated 1000 sets of responses from the working model.

The Wald statistic for testing $H_0 : \omega = 0$ versus $H_1 : \omega \neq 0$ has the following form

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

where $\hat{\omega}$ is the ML estimator of ω and $\hat{\sigma}_\omega^2$ is the estimated variance of $\hat{\omega}$ which is nothing else as the corresponding diagonal element of the Fisher information matrix evaluated at $(\hat{\omega}, \hat{\boldsymbol{\beta}})$. Estimated upper tail probabilities for an α size test are computed by calculating the proportion of times when W_ω is greater than or equal to the critical value $\chi^2_{1,1-\alpha}$, i.e.

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

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 estimated upper tail probabilities correspond to the estimated level of the test. For ZIP samples with zero-inflation $\omega > 0$ they give the estimated power function at ω. These values are given in Table 3 for the Wald test and the score test for linear predictors $\eta_i(\boldsymbol{\beta})$'s given by $\eta_i(\boldsymbol{\beta}) = 0.75 - 1.45x_i$, $i = 1, \ldots, n$. The results for the score test are reproduced from Jansakul and Hinde (2002). In general, the Wald test showed a considerable better performance than the score test especially in smaller samples $n = 50, 100$ and small α ($\alpha = 0.01$).

21

Table 3: Estimated upper tail probabilities for Wald (W_ω) and score (S_ω) statistics at $\chi^2_{1,1-\alpha}$ based on 1000 samples.

Level of the tests		$\alpha = 0.05$		$\alpha = 0.01$	
		W_ω	S_ω	W_ω	S_ω
$n = 50$	$\omega = 0.00$	0.036	0.042	0.014	0.008
	$\omega = 0.25$	0.452	0.292	0.279	0.129
	$\omega = 0.45$	0.780	0.583	0.661	0.372
$n = 100$	$\omega = 0.00$	0.035	0.056	0.010	0.012
	$\omega = 0.25$	0.786	0.551	0.608	0.324
	$\omega = 0.45$	0.964	0.911	0.918	0.782
$n = 200$	$\omega = 0.00$	0.036	0.054	0.011	0.006
	$\omega = 0.25$	0.936	0.900	0.861	0.752
	$\omega = 0.45$	1.000	0.999	0.999	0.993

8. DISCUSSION

This paper shows that the ML estimators in ZIGP (GP, ZIP) regression models possess analogous asymptotic properties as they do in GLM. General results of Fahrmeir and Kaufmann (1985) for noncanonical links in GLM have been adopted for this purpose. Simulation study illustrates that the normal approximation is satisfactory for moderate and large sample sizes. In particular for moderate overdispersion ($\varphi = 1.25$) and moderate zero-inflation ($\omega = 0.25$) sample sizes of $n = 200$ are sufficient.

In general score tests have a computational advantage over Wald tests since they require fitting the model only under the null hypothesis. However nowadays, given modern computing power, this point has lost importance in many problems. Our investigations show that for small and moderate sample sizes the Wald test for detecting zero-inflation or zero-deflation deflation in a ZMP regression model is considerable more powerful than the score test.

22

ACKNOWLEDGEMENTS

APPENDIX 1: FISHER INFORMATION MATRIX FOR ZIGP REGRESSION

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

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

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 straightforward computed. For instance entries of the matrix $\frac{\partial l_n(\boldsymbol{\delta})}{\partial\beta\beta^t}$ are given by

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

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, Kent, and Bibby (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 straightforward computed and are

given by

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

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

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

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 $r = 0, \ldots, p$;

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

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

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

APPENDIX 2: PROOF OF THEOREM 1

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 $n = 1, 2, \ldots$

We start the appendix 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),*

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

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

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

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

Now it follows

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

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 $\boldsymbol{\beta}_0 \in B$.

25

Now we show the second inequality of the lemma. First, we reparameterize 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

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

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 $\tilde{Y} \sim GP(\exp(\mathbf{x}^t \boldsymbol{\beta}_0), \varphi_0)$. $\qquad\qquad\square$

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 (10), (11) and (12) 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

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

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

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

27

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

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

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

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

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

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

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

Proof. We have a.s.

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

Thus, conditions

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

and

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

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.

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

Recall that the Hessian matrix 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

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

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

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

29

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

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

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

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

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

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

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

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

For A_n we have the following bounds:

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

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 $\max_{\mathbf{x}\in K_x} |\exp(\mathbf{x}^t\boldsymbol{\beta}) - \exp(\mathbf{x}^t\boldsymbol{\beta}_0)|$ 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 in probability. 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 ourself to the case $r = s = p$. It easy to see that condition (24) will follow from the next three conditions :

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

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

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

31

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

References

Cameron, A. and P. Trivedi (1998). *Regression analysis of count data.* Cambridge University Press.

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 propertis of discrete Lagrangian distribution. In *A modern course on statistical distribution in scientific work*, Volume 1. Boston.

Consul, P. C. and M. M. Shoukri (1985). The generalized Poisson distribution when the sample mean is large than the sample variance. *Comm. Statist. Theory Methods 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/).*

Dean, C. (1992). Testing for overdispersion in Poisson and binomial regression models. *JASA 87*(418), 451–457.

Dean, C. and J. F. Lawless (1989). Tests for detecting overdispersion in Poisson regression models. *J. Amer. Statist. Assoc. 84*(406), 467–472.

Deng, D. and S. R. Paul (2000). Score tests for zero inflation in generalized linear models. *Canad. J. Statist. 28*(3), 563–570.

Deng, D. and S. R. Paul (2005). Score tests for zero-inflation and over-dispersion in generalized linear models. *Statist. Sinica 15*(1), 257–276.

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

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.

Hall, D. B. and J. T. Præstgaard (2001). Order-restricted score tests for homogeneity in generalised linear and nonlinear mixed models. *Biometrika 88*(3), 739–751.

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.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *Ann. Math. Statistics 22*, 79–86.

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

Lawless, J. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics 15*, 209–225.

Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika 84*(2), 309–326.

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.

McCullagh, P. and J. Nelder (1989). *Generalized linear models*. Chapman & Hall.

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

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

Ridout, M., C. G. B. Demétrio, and J. Hinde (1998). Models for count data with many zeroes. In *Proceedings of the XIXth International Biometrics Conference*, pp. 179–192. Cape Town.

Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc. 82*(398), 605–610.

Stekeler, D. (2004). Verallgemeinerte Poissonregression und daraus abgeleitete Zero-Inflated und Zero-Hurdle Regressionsmodelle. Master's thesis, Technische Universität München (*www-m4.ma.tum.de/Diplarb/*).

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

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

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

Wang, P., I. Cockburn, and M. Puterman (1998). Analysis of patent data – a mixed Poisson regression model approach. *J. Bus. Econom. Statist. 16*, 27–41.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*(1), 1–25.

Winkelmann, R. (2003). *Econometric analysis of count data* (fourth ed.). Berlin: Springer-Verlag.