On Estimation and Influence Diagnostics for Zero-Inflated Negative Binomial Regression Models

A. M. Garay[a], **E. M. Hashimoto**[b], **E. M. M. Ortega**[b] **and V. H. Lachos**[a][1]

[a] *Departmento de Estatistica, Universidade Estadual de Campinas, Brazil*
[b] *Departmento de Ciências Exatas, Universidade de São Paulo, Brazil*

Abstract

The zero-inflated negative binomial model is used to account for overdispersion detected in data that are initially analyzed under the zero-inflated Poisson model. We consider a frequentist analysis, a jackknife estimator and non-parametric bootstrap for parameter estimation of zero-inflated negative binomial regression models. In addition, an EM-type algorithm is developed to perform maximum likelihood estimation. Then, we derive the appropriate matrices for assessing local influence on the parameter estimates under different perturbation schemes and present some ways to perform global influence analysis. In order to study departures from the error assumption as well as the presence of outliers, we perform residual analysis based on the standardized Pearson residuals. The relevance of the approach is illustrated with a real data set, where it is shown that, by removing the most influential observations, the decision about which model best fits the data changes.

Keywords: binomial negative distribution, EM-algorithm, bootstrap, global influence, local influence, zero-inflated models.

1 Introduction

Count data with many zeros (or zero-inflation) are commonly encountered in many disciplines, including medicine (Bohning et al., 1999), public health (Zhou and Tu, 2000), environmental sciences (Agarwal et al., 2002), agriculture (Hall, 2000) and manufacturing applications (Lambert, 1992). Zero-inflation, a frequent manifestation of overdispersion, means that the incidence of zero counts is greater than expected. This is of interest because zero counts frequently have special status. For example, in counting the number of responses to an exposure, an individual may have no disease response because of his/her immunity or resistance to the disease. If overdispersion in raw data is caused by the zero inflation, then the zero-inflated Poisson (ZIP) model, described in Lambert (1992) seminal work, provides a standard framework to fit the data. The basic idea behind the derivation of the ZIP model is to mix a distribution degenerate at zero with a Poisson distribution. Since one could theoretically mix the degenerate distribution with any count distribution, we refer to the latter (nondegenerate) distribution/model as the baseline model.

Having accounted for zero inflation, if the data continue to suggest additional overdispersion, we should consider the zero inflated negative binomial (ZINB) model, mixing a distribution degenerate at zero with a baseline negative binomial distribution, over the zero-inflated Poisson model. Without confusion, overdispersion can be the result of excess zeros or some other cause. In any case, the result is excess variability. In some cases, the ZIP model may be not appropriate for such data, since the baseline (Poisson) model does not accommodate the remaining overdispersion not accounted for through zero

[1]Corresponding author: Departamento de Estatística, Universidade Estatual de Campinas, Rua Sérgio Buarque de Holanda, 651, Cidade Universitária Zeferino Vaz, Campinas, São Paulo, Brazil. CEP 13083-859.

inflation and it is well known that negative binomial (NB) models are more flexible than their simpler Poisson counterparts in accommodation of overdispersion (Lawless, 1987). The ZINB model was discussed in Ridout et al. (2001), where a a score test is provided for testing ZIP regression models against ZINB alternatives. More recently, Mwalili et al. (2008) illustrated how the ZINB regression model can be corrected for misclassification.

Influence diagnostics is an important aspect in the analysis of a data set following parameter estimation, as it provides an indication of lack of fit or influential observations. Cook (1986) proposed a diagnostic approach, named local influence, to assess the effect of small perturbations in the model and/or data on the parameter estimates. Several authors have applied the local influence methodology in more general regression models than the normal regression model (see for instance, Galea et al., 2004; Zeller et al., 2009). Moreover, some authors have investigated the assessment of local influence in survival analysis models: for instance, Carrasco et al. (2008) derived the appropriate matrices for assessing local influence in log-modified Weibull regression models and Silva et al. (2008) adapted global and local influence methods in log-Burr XII regression models with censored data. Influence diagnostics for NB models can be found, for instance, in Svetliza and Paula (2003) and for ZIP models, the recent works by Xie et al. (2008) and Xie and Wei (2009) can be cited. However, to the best of our knowledge there are neither studies on ZINB models related to influence diagnostics nor on local influence. Thus, the main objective of this work is to develop estimation methods and diagnostics analysis, based on case-deletion and the local influence approach, for ZINB regression models. In the presence of zero inflation, we expect that the techniques developed here will enable practitioners to make correct conclusions and valid inferences from zero-inflated regression models.

The paper is organized as follows. In Section 2 we give a brief sketch of ZINB regression models, and consider maximum likelihood (ML) estimation, jackknife and non-parametric bootstrap estimators for the model parameters, including an EM-type algorithm for ML estimation. In addition, the observed information matrix is derived analytically. In Section 3, we study the local influence and illustrate the curvature calculations for four perturbation schemes. Furthermore, we present some ways to perform global influence and residual analysis based on the standardized Pearson residuals. The methodology is illustrated in Section 4, in which we compare ZINB and ZIP models according to the robustness aspects of the ML estimates. Finally, we make some concluding remarks and suggestions for further research in Section 5.

2 The ZINB regression model

The NB and Poisson regression model is a popular tool for modeling count data and is applied in a wide range of applications in the social and physical sciences. Real data, however, are often overdispersed (zero-inflated), and are thus not appropriate for NB and Poisson regression. We study a regression model based on the NB distribution to address this problem. We consider a ZINB regression model in which the response variable $Y_i, (i = 1, \ldots, n)$ has a probability mass function (p.m.f.) given by

$$Pr(Y_i = y_i) = \begin{cases} p_i + (1-p_i)\left(\frac{\phi}{\mu_i+\phi}\right)^{\phi}, & y_i = 0; \\ (1-p_i)\dfrac{\Gamma(\phi+y_i)}{\Gamma(y_i+1)\Gamma(\phi)}\left(\dfrac{\mu_i}{\mu_i+\phi}\right)^{y_i}\left(\dfrac{\phi}{\mu_i+\phi}\right)^{\phi}, & y_i = 1, 2, \ldots, \end{cases} \quad (1)$$

where $0 \le p_i \le 1$, $\mu_i \ge 0$, ϕ^{-1} is the dispersion parameter with $\phi > 0$ and $\Gamma(.)$ is the gamma function. The mean and the variance of the model defined in (1) are $E(Y_i) = (1-p_i)\mu_i$, $Var(Y_i) = (1-p_i)\mu_i(1+\mu_i\phi^{-1}+p_i\mu_i)$. When $p_i = 0$, the random variable Y_i has a binomial negative distribution with mean μ_i and dispersion parameter ϕ, i.e, $Y_i \sim NB(\mu_i, \phi)$ in the usual notation.

In many practical applications it is common to assume that the parameters p_i and μ_i depend on vectors of explanatory variables $\mathbf{x}_i$ and $\mathbf{z}_i$, respectively. In this work, we assume the specific models

$$\log(\mu_i) = \mathbf{x}_i^T\boldsymbol{\beta} \quad \text{and} \quad \log\left(\frac{p_i}{1-p_i}\right) = \mathbf{z}_i^T\boldsymbol{\gamma},\ i = 1,\ldots,n, \tag{2}$$

where $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^T$ and $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_q)^T$ are unknown parameters. However, similar results could be derived for other link functions.

Now, consider an observed sample $(y_1,\mathbf{x}_1,\mathbf{z}_1),\ldots,(y_n,\mathbf{x}_n,\mathbf{z}_n)$ of n independent observations, where each observed response is denoted by y_i. Then, the log-likelihood function for the vector of parameters $\boldsymbol{\theta} = (\phi,\boldsymbol{\beta}^T,\boldsymbol{\gamma}^T)^T$, given the observed sample, has the form

$$l(\boldsymbol{\theta}) = \sum_{i:y_i=0} l_1(\phi,\mathbf{x}_i^T\boldsymbol{\beta},\mathbf{z}_i^T\boldsymbol{\gamma}) + \sum_{i:y_i>0} l_2(\phi,\mathbf{x}_i^T\boldsymbol{\beta},\mathbf{z}_i^T\boldsymbol{\gamma}), \tag{3}$$

where

$$\begin{aligned} l_1(\phi,\mathbf{x}_i^T\boldsymbol{\beta},\mathbf{z}_i^T\boldsymbol{\gamma}) &= -\log\left[1+\exp(\mathbf{z}_i^T\boldsymbol{\gamma})\right] + \log\left\{\exp(\mathbf{x}_i^T\boldsymbol{\beta}) + \left[\frac{\phi}{\phi+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}\right]^{\phi}\right\} \quad \text{and} \\ l_2(\phi,\mathbf{x}_i^T\boldsymbol{\beta},\mathbf{z}_i^T\boldsymbol{\gamma}) &= -\log\left[1+\exp(\mathbf{z}_i^T\boldsymbol{\gamma})\right] + \log[\Gamma(\phi+y_i)] - \log[\Gamma(y_i+1)] - \log[\Gamma(\phi)] + \\ &\quad + y_i\log\left[\frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{\phi+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}\right] + \phi\log\left[\frac{\phi}{\phi+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}\right]. \end{aligned}$$

The ML estimate $\widehat{\boldsymbol{\theta}}$ of the vector of unknown parameters can be calculated by maximizing the log-likelihood given in (3). There are many optimization procedures available in standard programs, such as the MaxBFGS routine in the matrix programming language Ox (see, Doornik, 2007), which need only the original estimator function rather than their derivatives. Hypothesis testing and standard errors of the resulting estimators $\widehat{\boldsymbol{\theta}}$ can be based on the asymptotic normal approximation

$$(\widehat{\phi},\widehat{\boldsymbol{\beta}}^T,\widehat{\boldsymbol{\gamma}}^T)^T \quad \sim \quad \mathrm{N}_{(p+q+1)}\Big\{(\phi,\boldsymbol{\beta}^T,\boldsymbol{\gamma}^T)^T, -\ddot{\mathbf{L}}^{-1}(\boldsymbol{\theta})\Big\},$$

where $-\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \left\{\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right\}$ is the $(2p+1)\times(2p+1)$ observed information matrix, which has the form

$$\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{L}_{\phi\phi} & \mathbf{L}_{\phi\beta_j} & \mathbf{L}_{\phi\gamma_j} \\ . & \mathbf{L}_{\beta_j\beta_k} & \mathbf{L}_{\beta_j\gamma_k} \\ . & . & \mathbf{L}_{\gamma_j\gamma_k} \end{pmatrix}$$

(the corresponding sub-matrices are given in Appendix A).

A disadvantage of direct maximization of the log-likelihood function is that it may not converge unless good starting values are used. Thus, we also use the EM algorithm, which is stable and straightforward to implement since the iterations converge monotonically and no second derivatives are required. Moreover, the EM-estimates are quite insensitive to the stating values.

2.1 The EM-algorithm

As in Hall (2000), the missing element/factor in this problem is a vector of indicator variables $\mathbf{w} = (w_1,\ldots,w_n)^\top$, where $w_i = 1$, when Y_i is from the zero state and $w_i = 0$ when Y_i is from the NB state. The complete-data log-likelihood associate with $\mathbf{Y}_c = (\mathbf{y},\mathbf{w})$ is then

$$\ell_c(\boldsymbol{\theta}|\mathbf{Y}_c) \quad = \quad \sum_{i=1}^{n}\Big\{w_i\mathbf{z}_i\boldsymbol{\gamma} - \log[1+\exp(\mathbf{z}_i\boldsymbol{\gamma})] + (1-w_i)\log[g(y_i;\boldsymbol{\beta},\phi)]\Big\},$$

where $g(y_i;\boldsymbol{\beta},\phi) = \dfrac{\Gamma(\phi+y_i)}{\Gamma(y_i+1)\Gamma(\phi)}\Big(\dfrac{\mu_i}{\mu_i+\phi}\Big)^{y_i}\Big(\dfrac{\phi}{\mu_i+\phi}\Big)^{\phi}$, with $\mu_i = \exp(\mathbf{x}_i^\top\boldsymbol{\beta}\ )$. Notice that $\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{z})$ has a particularly convenient form for the EM algorithm; i.e., $\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{z})$ is linear in $\mathbf{w}$, so that at iteration

(k) of the algorithm, the E- step consists of replacing $\mathbf{w}$ by its conditional expectation given $\mathbf{y}$, $\widehat{\boldsymbol{\beta}}^{(k)}$ and $\widehat{\boldsymbol{\gamma}}^{(k)}$. This conditional expectation is easily calculated as

$$\hat{w}_i^{(k)} = \begin{cases} \left(1+\exp(-\mathbf{z}_i\widehat{\boldsymbol{\gamma}}^{(k)})\left[\dfrac{\widehat{\phi}^{(k)}}{\exp(\mathbf{x}_i\widehat{\boldsymbol{\beta}}^{(k)})+\widehat{\phi}^{(k)}}\right]^{\widehat{\phi}^{(k)}}\right)^{-1} & , \quad \text{if } y_i = 0, \\ 0 & , \quad \text{if } y_i > 0, \end{cases}$$

so that

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = E(\ell_c(\boldsymbol{\theta}|\mathbf{Y}_c)|\mathbf{y},\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\beta},\phi|\widehat{\boldsymbol{\theta}}^{(k)}),$$

where

$$Q_{1i}(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}}^{(k)}) = \widehat{w}_i^{(k)}\mathbf{z}_i\boldsymbol{\gamma} - \log[1+\exp(\mathbf{z}_i\boldsymbol{\gamma})]$$

and

$$Q_{2i}(\boldsymbol{\beta},\phi|\widehat{\boldsymbol{\theta}}^{(k)}) = (1-\widehat{w}_i^{(k)})\log\left\{\frac{\Gamma(\phi+y_i)}{\Gamma(\phi)\Gamma(1+y_i)}\left[\frac{\exp(\mathbf{x}_i^\top\boldsymbol{\beta}\ )}{\exp(\mathbf{x}_i^\top\boldsymbol{\beta}\ )+\phi}\right]^{y_i}\left[\frac{\phi}{\exp(\mathbf{x}_i^\top\boldsymbol{\beta}\ )+\phi}\right]^{\phi}\right\}.$$

Note that $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ is easily maximized (M-step) with respect a $\boldsymbol{\gamma}$ and $(\boldsymbol{\beta},\phi)$, because $\sum_{i=1}^{n} Q_{1i}(\boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}}^{(k)})$ it is equal to the log-likelihood for an unweighted binomial logistic regression of $\widehat{\mathbf{w}}^{(k)}$ on $\mathbf{z}$ (a term not involving $\boldsymbol{\beta}$) and $\sum_{i=1}^{n} Q_{2i}(\boldsymbol{\beta},\phi|\widehat{\boldsymbol{\theta}}^{(k)})$ is the log-likelihood for a weighted NB log-linear regression of $\mathbf{y}$ on $\boldsymbol{\beta}$ and ϕ (a term not involving $\boldsymbol{\gamma}$). These optimization procedures can be easily accomplished, for instance, in the R software through the efficient routine *glm*().

The iterations are repeated until a suitable convergence rule is satisfied, e.g., if $||\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)}||$ is sufficiently small, $k = 0, 1, ...$

2.2 Jackknife estimator

Jackknifing involves transforming the problem of estimating a population parameter into the problem of estimating a population mean. According to this method, a mean value is first estimated, although the approach estimation is unusual. A framework for implementing the jackknife method is given by Lipsitz et al. (1990), who suggest an alternative robust estimator of the covariance matrix based on jackknifing in order to analyze data from repeated measures studies. In this paper, we use this method as an alternative to estimate the population parameters.

Suppose that $Y_1, \ldots, Y_n$ is a random sample of n values and that $\bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n}$ is the sample mean used to estimate the mean of the population. The sample mean calculated with the lth observation missing is

$$\bar{Y}_{-l} = \frac{\sum_{i=1}^{n} Y_i - Y_l}{n-1},$$

for which

$$Y_l = n\hat{Y} - (n-1)\bar{T}_{-l}. \tag{4}$$

Using a general example, let θ be a parameter estimated by $\hat{E}(Y_1, \ldots, Y_n)$. For ease of notation, we drop $(Y_1, \ldots, Y_n)$. Finally, $\hat{E}_{-l}$ is calculated, which is obtained with the Y_l observation missing. It follows from equation (4) that pseudo-values can be calculated as follows:

$$\hat{E}_l^* = n\hat{E} - (n-1)\hat{E}_{-l}, \quad l = 1, \ldots, n.$$

The average of the pseudo-values is the jackknife estimate of θ, given by

$$\hat{E}^* = \frac{\sum_{l=1}^{n} \hat{E}_l^*}{n}.$$

Manly (2006) suggested that an approximate $100(1-\alpha)\%$ confidence interval for θ is given by $\hat{E}^* \pm t_{\alpha/2,n-1} s/\sqrt{n}$, where s is the standard deviation of the pseudo-values; and $t_{\alpha/2,n-1}$ is the upper $(1-\alpha/2)$ point of the t-distribution with $(n-1)$ degrees of freedom, which has the effect of removing bias of order $1/n$. The jackknife estimation calculations for the ZINB regression model are performed for ϕ, β_j and γ_j $(j = 1, \ldots, p)$, and confidence intervals are calculated separately for each parameter.

2.3 Bootstrap re-sampling method

The bootstrap re-sampling method was proposed by Efron (1979). The method treats the observed sample as if it represented the population. From the information obtained from such a sample, B bootstrap samples of similar size to that of the observed sample are generated, from which it is possible to estimate various characteristics of the population, such as mean, variance, percentiles and so on.

According to the literature, the re-sampling method may be non-parametric or parametric. In this study, the non-parametric bootstrap method is addressed, according to which the distribution function F can be estimated by empirical distribution $\hat{F}$.

Let $\mathbf{T}=(T_1, \ldots, T_n)$ be an observed random sample and $\hat{F}$ be the empirical distribution of $\mathbf{T}$. Thus, a bootstrap sample $\mathbf{T}^*$ is constructed by re-sampling with replacement of n elements of the sample $\mathbf{T}$. For the B bootstrap samples generated, $T_1^*, \ldots, T_B^*$, the bootstrap replication of the parameter of interest for the b-th sample is given by:

$$\hat{\boldsymbol{\theta}}_b^* = s(T_b^*),$$

that is, the value of $\hat{\boldsymbol{\theta}}$ for sample T_b^*, $b = 1, \ldots, B$.

The bootstrap estimator of the standard error (Efron and Tibshirani, 1993) is the standard deviation of these bootstrap samples; it is denoted by $\hat{EP}_B$ and obtained by the following expression:

$$\hat{EP}_B = \left[\frac{1}{(B-1)} \sum_{b=1}^{B} \left(\hat{\theta}_b^* - \bar{\theta}_B \right)^2 \right]^{1/2},$$

in which $\bar{\theta}_B = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b^*$. Note that B is the number of bootstrap samples generated. According to Efron and Tibshirani (1993) , assuming $B \geq 200$, it is generally sufficient to present good results to determine the bootstrap estimations. However, to achieve greater accuracy, a reasonably high B value must be considered. In this study, we consider $B = 3000$ bootstrap samples. We describe the bias corrected and accelerated (BCa) method for constructing approximated confidence intervals based on the bootstrap re-sampling method. For further details on bootstrap intervals, see for example, Efron and Tibshirani (1993), DiCiccio and Efron (1996) and Davison and Hinkley (1997).

BCa bootstrap interval

The bootstrap interval based on the BCa method assumes that the percentiles used in delimitating the bootstrap confidence intervals depend on the corrections for tendency $\hat{a}$ and acceleration $\hat{z}_0$.

The bias correction value $\hat{z}_0$ is generated based on the proportion of estimations of bootstrap samples that are smaller than the original estimation $\hat{\boldsymbol{\theta}}$. The expression of $\hat{z}_0$ is given by

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\sharp(\hat{\boldsymbol{\theta}}_b^* < \hat{\boldsymbol{\theta}})}{B} \right), \quad b = 1, \ldots, B.$$

Note that $\Phi^{-1}(\cdot)$ is the inverse of the accumulated standard normal distribution; B is the number of generated bootstrap samples; $\hat{\boldsymbol{\theta}}$ is the MLE of the observed sample; and $\hat{\boldsymbol{\theta}}_b^*$ is the MLE of the b-th bootstrap sample.

Let $\hat{\boldsymbol{\theta}}_{(i)}$ be the MLE of the sample without the i-th observation. Then $\hat{a}$ is given by

$$\hat{a} = \frac{\sum_{i=1}^{n} \left[\hat{\boldsymbol{\theta}}_{(\cdot)} - \hat{\boldsymbol{\theta}}_{(i)}\right]^3}{6\left\{\sum_{i=1}^{n} \left[\hat{\boldsymbol{\theta}}_{(\cdot)} - \hat{\boldsymbol{\theta}}_{(i)}\right]^2\right\}^{3/2}}.$$

Note that $\hat{\boldsymbol{\theta}}_{(\cdot)} = \sum_{i=1}^{n} \hat{\boldsymbol{\theta}}_{(i)}/n$ and n is the sample size.

Hence, the BCa bootstrap interval of coverage $100(1-2\alpha)\%$ is given by

$$\left[\hat{\boldsymbol{\theta}}^*_{(B\alpha_1)}, \hat{\boldsymbol{\theta}}^*_{(B\alpha_2)}\right],$$

in which

$$\alpha_1 = \Phi\left\{\hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(\alpha)}{1 - \hat{a}\left[\hat{z}_0 + \Phi^{-1}(\alpha)\right]}\right\} \quad \text{and} \quad \alpha_2 = \Phi\left\{\hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(1-\alpha)}{1 - \hat{a}\left[\hat{z}_0 + \Phi^{-1}(1-\alpha)\right]}\right\}.$$

Note that α_1 and α_2 are corrections to the bootstrap percentiles; $\Phi(\cdot)$ is an accumulated distribution function of the standard normal distribution; and $\Phi^{-1}(\cdot)$ is the inverse of the accumulated distribution function of the standard normal distribution (Efron and Tibshirani, 1993).

3 Sensitivity analysis

3.1 Global influence

A first tool to perform sensitivity analysis, as stated before, is by means of global influence starting from case deletion. Case deletion is a common approach to study the effect of dropping the i-th case from the data set. The case deletion model for (1) is given by

$$Pr(Y_{(i)} = y_{(i)}) = \begin{cases} p_{(i)} + (1-p_{(i)})\left(\frac{\phi}{\mu_{(i)}+\phi}\right)^{\phi}, & y_{(i)} = 0; \\ (1-p_{(i)})\frac{\Gamma(\phi+y_{(i)})}{\Gamma(y_{(i)}+1)\Gamma(\phi)}\left(\frac{\mu_{(i)}}{\mu_{(i)}+\phi}\right)^{y_{(i)}}\left(\frac{\phi}{\mu_{(i)}+\phi}\right)^{\phi}, & y_{(i)} = 1, 2, \ldots, \end{cases} \tag{5}$$

where $\mu_{(i)} = \exp(\mathbf{x}_{(i)}^T\boldsymbol{\beta})$, $p_{(i)} = \frac{\exp(\mathbf{z}_{(i)}^T\boldsymbol{\gamma})}{1+\exp(\mathbf{z}_{(i)}^T\boldsymbol{\gamma})}$ and $i = 1, \ldots, n$. In the following, a quantity with subscript "(i)" means the original quantity with the i-th case deleted. For model (5), the log-likelihood function of $\boldsymbol{\theta}$ is denoted by $l_{(i)}(\boldsymbol{\theta})$.

Let $\hat{\boldsymbol{\theta}}_{(i)} = (\hat{\phi}_{(i)}, \hat{\boldsymbol{\beta}}_{(i)}^T, \hat{\boldsymbol{\gamma}}_{(i)}^T)^T$ be the ML estimate of $\boldsymbol{\theta}$ without the ith observation in the sample. To assess the influence of the i-th case on the ML estimate $\hat{\boldsymbol{\theta}} = (\hat{\phi}, \hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\gamma}}^T)^T$, the basic idea is to compare the difference between $\hat{\boldsymbol{\theta}}_{(i)}$ and $\hat{\boldsymbol{\theta}}$. If deletion of a case seriously influences the estimates, more attention should be paid to that case. Hence, if $\hat{\boldsymbol{\theta}}_{(i)}$ is far from $\hat{\boldsymbol{\theta}}$, then i-th case is regarded as an influential observation. A first measure of global influence is defined as the standardized norm of $\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}$ (generalized Cook's distance)

$$GD_i(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^T\left[-\ddot{\mathbf{L}}(\boldsymbol{\theta})\right](\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}).$$

Another alternative is to assess $GD_i(\phi)$, $GD_i(\boldsymbol{\beta})$ or $GD_i(\boldsymbol{\gamma})$, whose values reveal the impact of the i-th case on the estimates of ϕ, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. Another popular measure of the difference between $\hat{\boldsymbol{\theta}}_{(i)}$ and $\hat{\boldsymbol{\theta}}$ is the likelihood distance

$$LD_i(\boldsymbol{\theta}) = 2\Big\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{(i)})\Big\}.$$

Besides, this one can also compute $\hat{\beta}_j - \hat{\beta}_{j(i)}(j = 1, \ldots, p)$ to assess the difference between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{(i)}$. Alternative global influence measures are possible. One could think of the behavior of test statistics, such as those of the Wald test for explanatory variables or censoring effect, under a case-deletion scheme.

As $\hat{\boldsymbol{\theta}}_{(i)}$ is needed for every case, a very heavy total computational burden may be involved. In this case, the following one-step approximation for $\hat{\boldsymbol{\theta}}_{(i)}$ can be used to reduce the burden

$$\hat{\boldsymbol{\theta}}_{(i)} \cong \hat{\boldsymbol{\theta}} + \ddot{\mathbf{L}}(\hat{\boldsymbol{\theta}})^{-1}\dot{l}_i(\hat{\boldsymbol{\theta}}),$$

where $\dot{l}_i(\hat{\boldsymbol{\theta}}) = \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ (Cook and Weisberg, 1982, for instance)

We can also apply the techniques developed by Wang et al. (1996) to evaluate how the i-th observation affects a set of parameter estimates. We define the following quantity as the influential estimate (IE) for individual i and for parameters vector $\boldsymbol{\theta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, which has the form

$$IE(\boldsymbol{\theta})_i = \frac{1}{(2p+1)}\sum_{k=1}^{2p+1}\frac{|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{(i)k}|}{SE(\hat{\boldsymbol{\theta}}_k)}, \qquad IE(\boldsymbol{\gamma})_i = \frac{1}{p}\sum_{k=1}^{p}\frac{|\hat{\boldsymbol{\gamma}}_k - \hat{\boldsymbol{\gamma}}_{(i)k}|}{SE(\hat{\boldsymbol{\gamma}}_k)},$$
$$IE(\boldsymbol{\beta})_i = \frac{1}{p}\sum_{k=1}^{p}\frac{|\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_{(i)k}|}{SE(\hat{\boldsymbol{\beta}}_k)}, \tag{6}$$

where $\hat{\boldsymbol{\theta}}_k$, $\hat{\boldsymbol{\theta}}_{(i)k}$, $\hat{\boldsymbol{\gamma}}_k$, $\hat{\boldsymbol{\gamma}}_{(i)k}$, $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\beta}}_{(i)k}$ are the MLEs of the ZINB regression model. The $IE(.)_i$ calculated for individual i can be interpreted as the average relative coefficient changes for a set of estimates. It is useful for assessing the effect of parameter estimates by exclusion of the ith observation. Therefore, a relatively large value of $IE(.)_i$ indicates a potentially influential observation that might cause instability in model fitting.

3.2 Local influence

Our interest focuses on the influence of the subjects on the parameter estimates in the ZINB regression model. Another approach is suggested by Cook (1986), where instead of removing observations, weights are given to them. Local influence calculation can be carried out for model (1). If likelihood displacement $LD(\boldsymbol{\omega}) = 2\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$ is used, where $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denotes MLE under the perturbed model, the normal curvature for $\boldsymbol{\theta}$ in the direction $\mathbf{d}, \| \mathbf{d} \| = 1$, is given by $C_{\mathbf{d}}(\boldsymbol{\theta}) = 2|\mathbf{d}^T\boldsymbol{\Delta}^T\big[\ddot{\mathbf{L}}(\boldsymbol{\theta})\big]^{-1}\boldsymbol{\Delta}\mathbf{d}|$, where $\boldsymbol{\Delta}$ is a $(p+q+1)\times n$ matrix that depends on the perturbation scheme, and whose elements are given by $\Delta_{vi} = \partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\alpha_v\partial\omega_i$, $i = 1, \ldots, n$ and $v = 1, \ldots, p+q+1$ evaluated at $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$, where $\boldsymbol{\omega}_0$ is the no perturbation vector (see Cook, 1986). For the ZINB regression model, the elements of $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ are given in Appendix A. We can also calculate normal curvatures $C_{\mathbf{d}}(\phi)$, $C_{\mathbf{d}}(\boldsymbol{\beta})$ and $C_{\mathbf{d}}(\boldsymbol{\gamma})$ to perform various index plots, for instance, the index plot of $\mathbf{d}_{max}$, the eigenvector corresponding to $C_{\mathbf{d}_{max}}$, the largest eigenvalue of the matrix $\mathbf{B} = -\boldsymbol{\Delta}^T\big[\ddot{\mathbf{L}}(\boldsymbol{\theta})\big]^{-1}\boldsymbol{\Delta}$ and the index plots of $C_{\mathbf{d}_i}(\phi)$, $C_{\mathbf{d}_i}(\boldsymbol{\beta})$ and $C_{\mathbf{d}_i}(\boldsymbol{\gamma})$, named total local influence, where $\mathbf{d}_i$ denotes an $n \times 1$ vector of zeros with one at the i-th position. Thus, the curvature in direction $\mathbf{d}_i$ assumes the form $C_i = 2|\boldsymbol{\Delta}_i^T\big[\ddot{\mathbf{L}}(\boldsymbol{\theta})\big]^{-1}\boldsymbol{\Delta}_i|$, where $\boldsymbol{\Delta}_i^T$ denotes the i-th row of $\boldsymbol{\Delta}$. It is usual to point out those cases such that $C_i \geq 2\bar{C}$, where $\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$.

3.3 Curvature calculations

Next, we calculate for three perturbation schemes the matrix

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{vi})_{\big[(p+q+1)\times n\big]} = \left(\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\alpha_v\boldsymbol{\omega}_i}\right)_{\big[(p+q+1)\times n\big]}, \tag{7}$$

where $v = 1, \ldots, p+q+1$ and $i = 1, \ldots, n$. We will consider the model defined in (1) and its log-likelihood function given by (3).

3.3.1 Case-weight perturbation

First, we consider an arbitrary attribution of weights for the log-likelihood function, which may capture departures in general directions, represented by writing

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i:y_i=0} \omega_i l_1(\phi, \mathbf{x}_i^T\boldsymbol{\beta}, \mathbf{z}_i^T\boldsymbol{\gamma}) + \sum_{i:y_i>0} \omega_i l_2(\phi, \mathbf{x}_i^T\boldsymbol{\beta}, \mathbf{z}_i^T\boldsymbol{\gamma}), \tag{8}$$

where $0 \leq \omega_i \leq 1$, $\boldsymbol{\omega}_0 = (1, \ldots, 1)^T$ and $l_u(.)$ is defined in equation (3), $u = 1, 2$. The elements of the matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\phi^T, \boldsymbol{\Delta}_{\boldsymbol{\beta}}^T, \boldsymbol{\Delta}_{\boldsymbol{\gamma}}^T)^T$ are given in Appendix B.

3.3.2 Explanatory variables perturbation

In this section, we consider the influence that perturbation in the explanatory variables may produce on the parameter estimates for three cases (i) zero-inflation portion, (ii) NB portion, and (iii) zero-inflation and NB portion in the ZINB model.

- **zero-inflation portion**

  Consider an additive perturbation on the explanatory variable related to the zero-inflation portion, namely $\mathbf{z}_i$, by making $z_{it\omega} = z_{it} + \omega_i S_z$, where S_z is a scaled factor, $\omega_i \in \mathbf{R}$. This perturbation scheme leads to the following expressions for the perturbed log-likelihood function:

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i:y_i=0} l_1(\phi, \mathbf{x}_i^T\boldsymbol{\beta}, \mathbf{z}_i^{*T}\boldsymbol{\gamma}) + \sum_{i:y_i>0} l_2(\phi, \mathbf{x}_i^T\boldsymbol{\beta}, \mathbf{z}_i^{*T}\boldsymbol{\gamma}) \tag{9}$$

  where $\mathbf{z}_i^{*T}\boldsymbol{\gamma} = \gamma_0 + \gamma_1 z_{i1} + \ldots + \gamma_t\big(z_{it} + \omega_i S_z\big) + \ldots + \gamma_q z_{iq}$, $\boldsymbol{\omega}_0 = (0, \ldots, 0)^T$, and $l_u(.)$ is defined in equation (3), $u = 1, 2$. Matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\phi^T, \boldsymbol{\Delta}_{\boldsymbol{\beta}}^T, \boldsymbol{\Delta}_{\boldsymbol{\gamma}}^T)^T$ is given in Appendix C.

- **NB portion**

  Now the additive perturbation is introduced in the explanatory variables related to the baseline distribution, namely $\mathbf{x}_i$, by making $x_{it\omega} = x_{it} + \omega_i S_x$, where S_x is a scaled factor, $\omega_i \in \mathbf{R}$. This perturbation scheme leads to the following expressions for the perturbed log-likelihood function:

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i:y_i=0} l_1(\phi, \mathbf{x}_i^{*T}\boldsymbol{\beta}, \mathbf{z}_i^T\boldsymbol{\gamma}) + \sum_{i:y_i>0} l_2(\phi, \mathbf{x}_i^{*T}\boldsymbol{\beta}, \mathbf{z}_i^T\boldsymbol{\gamma}), \tag{10}$$

  where $\mathbf{x}_i^{*T}\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_t(x_{it} + \omega_i S_x) + \cdots + \beta_p x_{ip}$, and $l_u(.)$ is defined in equation (3), $u = 1, 2$. Matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\phi^T, \boldsymbol{\Delta}_{\boldsymbol{\beta}}^T, \boldsymbol{\Delta}_{\boldsymbol{\gamma}}^T)^T$ is given in Appendix D.

- **Zero-inflation and NB portion**

  Finally, additive perturbations are introduced simultaneously in the explanatory variables related to the baseline and degenerated distributions by making $z_{it\omega} = z_{it} + \omega_i S_z$ e $x_{it\omega} = x_{it} + \omega_i S_x$, where S_x e S_z are scaled factors. This perturbation scheme leads to the following expressions for the perturbed log-likelihood function:

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i:y_i=0} l_1(\phi, \mathbf{x}_i^{*T}\boldsymbol{\beta}, \mathbf{z}_i^{*T}\boldsymbol{\gamma}) + \sum_{i:y_i>0} l_2(\phi, \mathbf{x}_i^{*T}\boldsymbol{\beta}, \mathbf{z}_i^{*T}\boldsymbol{\gamma}), \tag{11}$$

  where $\mathbf{z}_i^{*T}\boldsymbol{\gamma} = \gamma_0 + \gamma_1 z_{i1} + \ldots + \gamma_t\big(z_{it} + \omega_i S_z\big) + \ldots + \gamma_q z_{iq}$, $\mathbf{x}_i^{*T}\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_t(x_{it} + \omega_i S_x) + \cdots + \beta_p x_{ip}$, $\boldsymbol{\omega}_0 = (0, \ldots, 0)^T$ and $l_u(.)$ is defined in equation (3), $u = 1, 2$. Matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\phi^T, \boldsymbol{\Delta}_{\boldsymbol{\beta}}^T, \boldsymbol{\Delta}_{\boldsymbol{\gamma}}^T)^T$ is given in Appendix E.

3.4 Goodness-of-fit and analysis residuals

The assessment of the fitted model is an important part of the data analysis, particularly in regression models, and residual analysis is a helpful tool for validation of the fitted model. Examination of residuals may be used, for instance, to detect the presence of outlying observations, model misspecification and/or the absence of components in the systematic part of the model and departures from the error and variance assumptions. However, finding appropriate residuals in non-normal regression models has been an important topic of research, particularly in overdispersed (zero-inflated) models. Most residuals are based on the differences between the observed responses and the fitted conditional mean $(y_i - \widehat{E(Y_i)})$. We consider the standardized ordinary residual (Pearson residual) to perform residual analysis, which is defined as:

$$\widehat{r}_i = \frac{y_i - \widehat{E(Y_i)}}{\sqrt{\widehat{Var(Y_i)}}}, \ i = 1, \ldots, n,$$

where $\widehat{E(Y_i)} = (1-\widehat{p}_i)\widehat{\mu}_i$, $\widehat{Var(Y_i)} = (1-\widehat{p}_i)\widehat{\mu}_i(1+\widehat{\mu}_i\widehat{\phi}^{-1}+\widehat{p}_i\widehat{\mu}_i)$, $\widehat{\mu}_i = \exp(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}})$ and $\widehat{p}_i = \dfrac{\exp(\mathbf{z}_i^\top \widehat{\boldsymbol{\gamma}})}{1+\exp(\mathbf{z}_i^\top \widehat{\boldsymbol{\gamma}})}$, with $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\phi}$ denoting the ML estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and ϕ, respectively. We also generate envelopes, as suggested by Atkinson (1981), to detect incorrect specification of the error distribution as well as the presence of outlying observations.

4 Application

To demonstrate the proposed methodology, we use the data set of apple cultivar reported by Ridout et al. (2001), referring to the number of roots produced by 270 micropropagated shoots of the columnar apple cultivar Trajan. The shoots had been produced under an 8- or 16-hour photoperiod in culture systems that utilized one of four different concentrations of the cytokinin BAP in the culture medium. There were 30 or 40 shoots of each of these eight treatment combinations. Of the 140 shoots produced under the 8-hour photoperiod, only 2 failed to produce roots, but 62 of the 130 shoots produced under the 16-hour photoperiod failed to root. The study of the influence of BAP concentration on two photoperiods on rooting of apple cultivars. The sample size was of $n = 270$ and the percentage of zeros observed was 23.7%. Thus, the explanatory variables are:

- y_i : count of roots;
- x_{i1}: photoperiod(0=8-hour, 1=16-hour);
- x_{i2}: concentrations of the cytokinin BAP,

where $i = 1, 2, \ldots, 270$.

Following Ridout et al. (2001) we fit a ZINB model as defined in equation (1), with

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) \quad \text{and} \quad p_i = \frac{\exp(\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2})}{1+\exp(\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2})}, \quad i = 1, \ldots, 270.$$

4.1 Estimation

To obtain the MLEs of the parameters in the ZINB regression model, we used the subroutine MAXBFGS in Ox, whose results are given in Table 1. Additionally, in Table 1 we report the EM estimates.

Table 1: Maximum likelihood estimates and EM estimates from the ZINB regression model fitted to apple shoots data.

	MLEs				EM		
$\boldsymbol{\theta}$	Estimate	S.E.	p-value	95% C.I.	Estimate	S.E.	95% C.I.
ϕ	12.350	3.980	-	(4.550, 20.149)	12.349	4.122	(4.270, 20.428)
β_0	1.977	0.065	0.000	(1.850, 2.104)	1.977	0.065	(1.850, 2.104)
β_1	-0.283	0.075	0.000	(-0.430, -0.135)	-0.282	0.075	(-0.430, -0.135)
β_2	-0.001	0.006	0.903	(-0.012, 0.011)	-0.001	0.005	(-0.012, 0.011)
γ_0	-4.523	0.975	<0.001	(-6.433, -2.612)	-4.523	0.975	(-6.433, -2.612)
γ_1	4.407	0.981	<0.001	(2.484, 6.331)	4.407	0.981	(2.484, 6.330)
γ_2	-0.000	0.029	0.998	(-0.056, 0.056)	-0.000	0.029	(-0.056, 0.056)

From Table 1, we can observe that the explanatory variable x_1 is significant at the 5% level, while the variable x_2, related to concentrations of the cytokinin BAP, is statistically non-significant for the count of roots ($p-value \approx 1$). These results indicate that the proposed EM-algorithm works very well and can be used reliably for ML estimation in ZINB models.

Now, by using the non-parametric bootstrap method with $B = 3000$, we find the bootstrap estimated and the BCa confidence intervals as described in Section 2.3. The results are presented in Table 2, along with the jackknife estimates. Note that estimatives from the four methods taken for illustration are very similar, as expected. However, since the developed methods are based on the likelihood, and asymptotic normality is expected for this sample size ($n = 270$), we will continue the analysis by using the ML estimates based on the EM-algorithm.

Table 2: Non-parametric bootstrap estimates and jackknife estimates from the ZINB regression model fitted to apple shoots data.

	Non-parametric bootstrap			Jackknife estimates		
$\boldsymbol{\theta}$	Estimate	S.E.	95% C.I.	Estimate	S.E.	95% C.I.
ϕ	15.108	6.615	(7.319, 19.948)	10.691	4.224	(2.375, 19.007)
β_0	1.975	0.065	(1.870, 2.082)	1.980	0.067	(1.848, 2.112)
β_1	-0.288	0.083	(-0.417, -0.145)	-0.281	0.083	(-0.444, -0.118)
β_2	-0.000	0.005	(-0.010, 0.008)	-0.001	0.006	(-0.013, 0.011)
γ_0	-4.500	0.650	(-5.670, -3.575)	-3.507	1.581	(-6.620, -0.394)
γ_1	4.329	0.687	(2.353, 5.561)	3.385	1.674	(0.089, 6.681)
γ_2	-0.005	0.028	(-0.049, 0.043)	0.001	0.031	(-0.060, 0.062)

4.2 Sensitivity analysis

In this section, we use the matrix programming language Ox to compute the case-deletion measures $GD_i(\boldsymbol{\theta})$, $LD_i(\boldsymbol{\theta})$, $IE(\boldsymbol{\theta})_i$, $IE(\boldsymbol{\gamma})_i$ and $IE(\boldsymbol{\beta})_i$ presented in Section 3.1. The results of such influence measures on the index plots are displayed in Figure 1 and Figure 2. From this figure we note that cases ♯191 and ♯192 are possibly influential observations, specifically in the ML estimates of $\boldsymbol{\gamma}$ related with the zero-inflation portion.

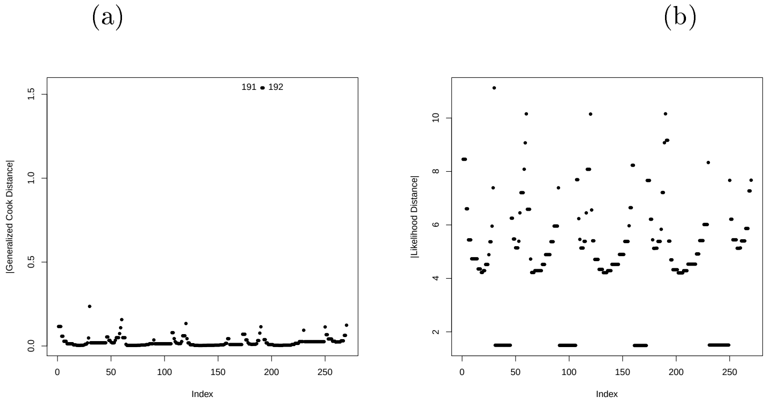

Figure 1: Index plot of global influence from the ZINB regression model fitted to apple shoots data. (a) Generalized Cook's distance. (b) Likelihood distance.

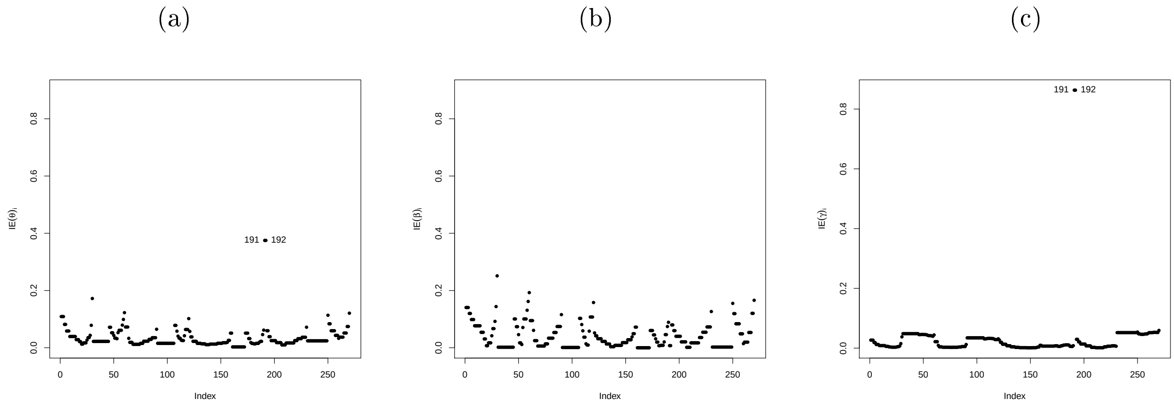

Figure 2: (a) Index plot of $IE(\boldsymbol{\theta})_i$ (b) Index plot of $IE(\boldsymbol{\beta})_i$ (c) Index plot of $IE(\boldsymbol{\gamma})_i$ from the ZINB regression model fitted to apple shoots data.

By applying the local influence theory developed in Section 3.2, where case-weight perturbation is used, the value $C_{\mathbf{d}_{max}} = 1.527$ was obtained as a maximum curvature. Now, considering the explanatory variables, the value for the maximum curvature calculated is $C_{\mathbf{d}_{max}} = 0.248$, $C_{\mathbf{d}_{max}} = 0.256$ and $C_{\mathbf{d}_{max}} = 0.258$ for the zero-inflation portion, NB portion and zero-inflation and NB portion, respectively. Once again, we note that cases ♯191 and ♯192 are possibly influential observations under case-weight perturbation.

(a) (b) (c) (d)

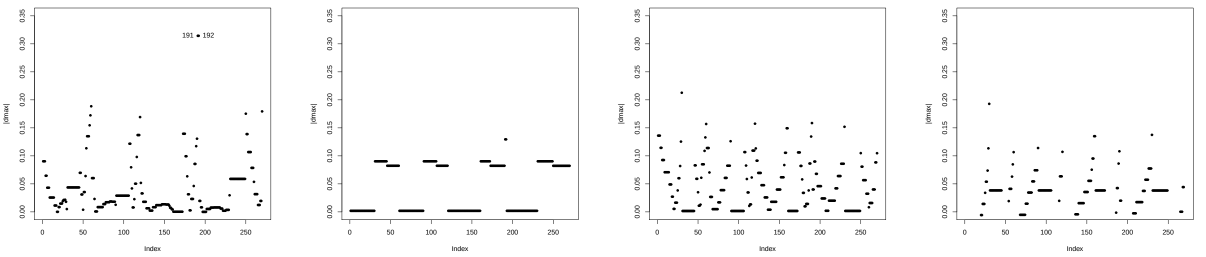

Figure 3: Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\theta}$ from the ZINB regression model fitted to apple shoots data. (a) Case-weight perturbation. (b) Explanatory variables perturbation: zero-inflation portion. (c) Explanatory variables perturbation: NB portion. (d) Explanatory variables perturbation: zero-inflation and NB portion.

(a) (b) (c) (d)

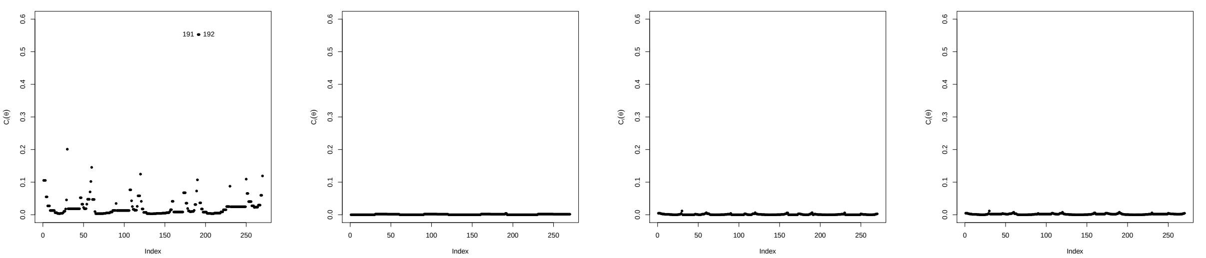

Figure 4: Index plot of local total C_i for $\boldsymbol{\theta}$ from the ZINB regression model fitted to apple shoots data. (a) Case-weight perturbation. (b) Explanatory variables perturbation: zero-inflation portion. (c) Explanatory variables perturbation: NB portion. (d) Explanatory variables perturbation: zero-inflation and NB portion.

4.3 Impact of the detected influential observations

Hence, the diagnostic analysis (global influence, local influence and residual analysis) indicated the two observations (♯191 and ♯192) as potentially influential. These observations represent the only apple shoots that were exposed to sunlight within 8 hours, in which there were observed no roots. In order to reveal the impact of these two observations on the parameter estimates, we refitted the model under some situations. First, we individually eliminated each of these two cases. Next, we removed all of potentially influential observations from the set “A" (original data set). In Table 3 we show the relative changes (in percentage) of each parameter estimate, defined by $\mathbf{RC}_{\theta_j} = \left[(\hat{\theta}_j - \hat{\theta}_{j(I)})/\hat{\theta}_j\right] \times 100$, parameter estimates and the corresponding p-values, where $\hat{\alpha}_{j(I)}$ denotes the MLE of θ_j after the set “I” of observations was removed. From Table 3, we note that the relative changes after the set I of observations was removed were computed as well as the new p-values (see Table 3). Note that large variations are only present for the parameters γ_0 and γ_1. This result agrees with the graphical analysis depicted in Figure 2.

Table 3: Relative changes [-**RC**-in %], estimates and the corresponding p-values in parentheses for the regression coefficients to explain the number of apple root shoots produced.

Set$\{I\}$	ϕ	β_0	β_1	β_2	γ_0	γ_1	γ_2
A	-	-	-	-	-	-	-
	12.350	1.977	-0.283	-0.001	-4.523	4.407	0.000
	(-)	(0.000)	(0.000)	(0.903)	(0.000)	(0.000)	(0.998)
A-{♯191}	[0]	[0]	[0]	[0]	[-24]	[-27]	[-24]
	12.298	1.978	-0.283	-0.001	-5.632	5.582	-0.007
	(-)	(0.000)	(0.000)	(0.893)	(0.012)	(0.014)	(0.801)
A-{♯192}	[0]	[0]	[0]	[0]	[-24]	[-27]	[-24]
	12.298	1.978	-0.283	-0.001	-5.632	5.582	-0.007
	(-)	(0.000)	(0.000)	(0.893)	(0.012)	(0.014)	(0.801)
A-{♯191 and ♯192}	[-6]	[0]	[-1]	[100]	[-445]	[-459]	[-24]
	13.055	1.975	-0.286	0.000	-24.672	24.653	-0.011
	(-)	(0.000)	(0.000)	(0.987)	(0.967)	(0.967)	(0.709)

Table 4: Comparison between ZINB and ZIP models by using different information criteria.

Model	AIC	BIC	CAIC
ZINB	1257.9	1283.1	1290.1
ZIP	1273.0	1295.0	1301.0

4.4 Model checking

Ridout et al. (2001) fitted various models to these data, based on the Poisson and negative binomial distributions and their zero-inflated counterparts. Based on a score test, they note that the ZINB regression model performed better than the ZIP model. Here we compute some information criteria and the standardized Pearson residuals to model comparison and/or check whether there is evidence of model misspecification.

The QQ-plots and simulated envelopes for the Pearson residuals are shown in Figure 5. The lines in these figures represent the 5th percentile, the mean, and the 95th percentile of 100 simulated points for each observation. These figures show that the ZINB regression model provides a better fit to the data set than the ZIP regression model. Moreover, there is no evidence of lack of fit for the ZINB model.

For model selection, we considered the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the consistent Akaike information criterion (CAIC). A summary of these values is given in Table 4. Clearly, the ZINB regression model outperforms the ZIP model in all the criteria and thus, the ZINB model can be used effectively in the analysis of these data. This conclusion agrees with Ridout et al. (2001), where also the ZIP model as found unsuitable for these data.

Finally, we fitted a ZINB model without the covariate x_{i2} (non-significant), i.e.,

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1}) \quad \text{and} \quad p_i = \frac{\exp(\gamma_0 + \gamma_1 x_{i1})}{1 + \exp(\gamma_0 + \gamma_1 x_{i1})},$$

for $i = 1, \ldots, 270$. The ML estimates for the parameters of this model are given in Table 5. Notice that the photoperiod has a negative influence in the root count, i.e., we could say that when the shoots are produced under a 16-hour photoperiod, the predicted root count is reduced by 25.6%.

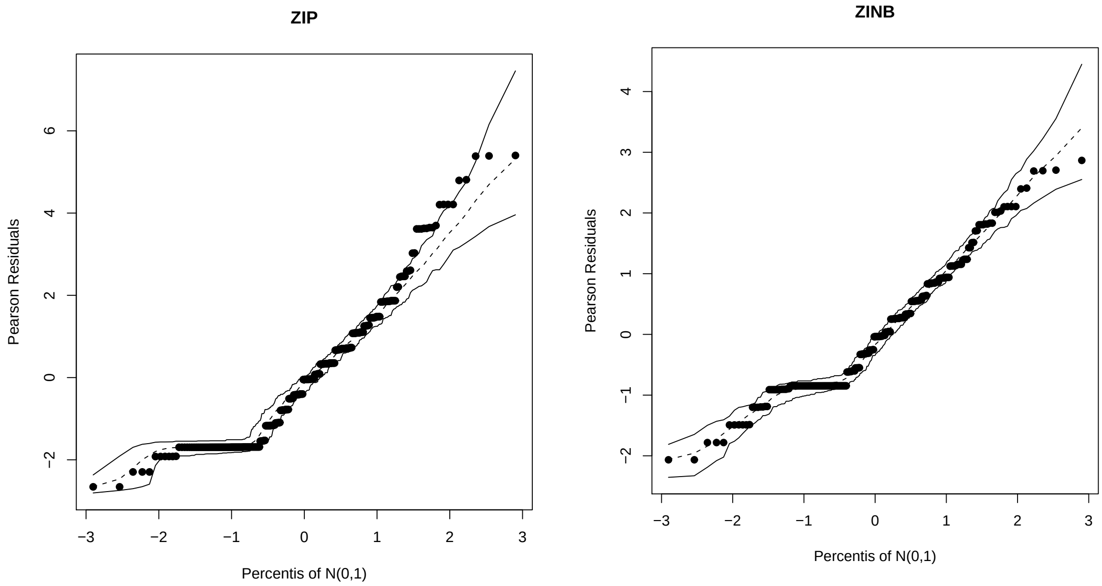

Figure 5: Plots of the Pearson residuals against the order statistics of the normal distribution from the ZIP and ZINB regression models fitted to apple shoots data.

On the other hand, related with the estimative of proportion of zeros (p), we obtain

$$\hat{p} = \frac{1}{270}\sum_{i=1}^{270}\hat{p}_i = 0.232.$$

Table 5: Maximum likelihood estimates from the ZINB regression model fitted to the final apple shoot data.

	MLEs			
$\boldsymbol{\theta}$	Estimate	S.E.	p-value	95% C.I.
ϕ	12.377	3.981	-	(4.547,20.179)
β_0	1.971	0.040	0.000	(1.892,2.050)
β_1	-0.283	0.075	0.000	(-0.430,-0.136)
γ_0	-4.513	0.947	<0.001	(-6.370,-2.656)
γ_1	4.397	0.973	<0.001	(2.509,6.285)

5 Concluding remarks

In this paper, we presented extensions of some estimation and influence diagnostics methods to ZINB regression models. We used the Quasi-Newton and EM-algorithm to obtain the maximum likelihood estimates and performed asymptotic tests for the parameters based on the asymptotic distribution of the ML estimates. The EM algorithm for fitting ZINB regression models is straightforward to implement in any statistical package that includes facilities for fitting weighted generalized linear models for Poisson and binomial data. On the other hand, as an alternative analysis, we discussed use of the jackknife estimator and parametric bootstrap for the ZINB regression model. We also discussed the influence diagnostics and model checking analysis in the ZINB regression models and the sensitivity of the maximum likelihood estimates via Pearson residuals and sensitivity analysis. We demonstrated through one application that the ZINB regression model can produce better fit than the ZIP regression models. The results derived in this work agree with the considerations that in this respect are presented in Ridout et al. (2001).

Due to recent advances in computational technology, it is worthwhile to carry out Bayesian treatments via Markov chain Monte Carlo (MCMC) sampling methods in the context of ZINB models. Bayesian influence diagnostics can be treated via the Kullback-Leibler divergence, as proposed by Cho et al. (2009). Other extensions of the current work include, for example, diagnostics analysis in zero-inflated negative binomial mixed models.

Appendix A: Matrix of second derivatives $\ddot{\mathbf{L}}(\boldsymbol{\theta})$

Here we derive the necessary formulas to obtain the second-order partial derivatives of the log-likelihood function. After some algebraic manipulations, we obtain

$$
\begin{aligned}
\mathbf{L}_{\phi\phi} &= \sum_{i:y_i>0}\Bigg\{[g_1(\mathbf{x}_i)]^{\phi}\left[\log[g_1(\mathbf{x}_i)]+g_2(\mathbf{x}_i)\right]^2+\frac{[g_1(\mathbf{x}_i)]g_2(\mathbf{x}_i)^{\phi}}{h(\mathbf{z}_i,\mathbf{x}_i)}\left[\phi^{-1}-\frac{1}{\phi+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}\right]\\
&\quad -\left\{\frac{[g_1(\mathbf{x}_i)]^{\phi}[\log[g_1(\mathbf{x}_i)]+g_2(\mathbf{x}_i)]}{h(\mathbf{z}_i,\mathbf{x}_i)}\right\}^2\Bigg\}\\
&\quad +\sum_{i:y_i>0}\left\{\psi'(y_i+\phi)-\psi'(\phi)+\frac{y_i}{[\phi+\exp(\mathbf{x}_i^T\boldsymbol{\beta})]^2}+g_2(\mathbf{x}_i)[\phi^{-1}-\frac{1}{\phi+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}]\right\},\\
\mathbf{L}_{\phi\beta_j} &= \sum_{i:y_i=0}\Bigg\{\frac{\phi x_{ij}[g_1(\mathbf{x}_i)]^{\phi}g_2(\mathbf{x}_i)[\log(g_1(\mathbf{x}_i))+g_2(\mathbf{x}_i)]}{h(\mathbf{z}_i,\mathbf{x}_i)}\left\{\frac{[g_1(\mathbf{x}_i)]^{\phi}}{h(\mathbf{z}_i,\mathbf{x}_i)}-1\right\}\\
&\quad -\frac{x_{ij}[g_1(\mathbf{x}_i)]^{\phi}g_2(\mathbf{x}_i)}{h(\mathbf{z}_i,\mathbf{x}_i)}\{2[1-g_1(\mathbf{x}_i)]+g_2(\mathbf{x}_i)\}\Bigg\}\\
&\quad +\sum_{i:y_i>0}\left\{\frac{y_ix_{ij}g_2(\mathbf{x}_i)}{\phi+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}-x_{ij}g_2(\mathbf{x}_i)\{2[1-g_1(\mathbf{x}_i)]+g_2(\mathbf{x}_i)\}\right\},\\
\mathbf{L}_{\phi\gamma_j} &= -\sum_{i:y_i=0}\left\{\frac{z_{ij}\exp(\mathbf{z}_i^T\boldsymbol{\gamma})[g_1(\mathbf{x}_i)]^{\phi}}{[h(\mathbf{z}_i,\mathbf{x}_i)]^2}\{\log[g_1(\mathbf{x}_i)]-g_2(\mathbf{x}_i)\}\right\},\\
\mathbf{L}_{\beta_j\beta_k} &= \sum_{i:y_i=0}\left\{\frac{\phi x_{ij}x_{ik}[g_1(\mathbf{x}_i)]^{\phi}[g_2(\mathbf{x}_i)]^2}{h(\mathbf{z}_i,\mathbf{x}_i)}\left[\phi+1-\frac{1}{h(\mathbf{z}_i,\mathbf{x}_i)}\right]-\frac{\phi x_{ij x_{ik}}[g_1(\mathbf{x}_i)]^{\phi}g_2(\mathbf{x}_i)}{h(\mathbf{z}_i,\mathbf{x}_i)}\right\}\\
&\quad +\sum_{i:y_i>0}\{y_ix_{ij}x_{ik}\{1-g_1(\mathbf{x}_i)-g_2(\mathbf{x}_i)[3-2g_2(\mathbf{x}_i)]\}+x_{ij}x_{ik}g_1(\mathbf{x}_i)g_2(\mathbf{x}_i)(y_i-\phi)\},\\
\mathbf{L}_{\beta_j\gamma_k} &= \sum_{i:y_i=0}\left\{\frac{\phi x_{ij}z_{ik}[g_1(\mathbf{x}_i)]^{\phi}g_2(\mathbf{x}_i)\exp(\mathbf{z}_i^T\boldsymbol{\gamma}}{[h(\mathbf{z}_i,\mathbf{x}_i)]^2}\right\}\\
\mathbf{L}_{\gamma_j\gamma_k} &= \sum_{i:y_i=0}\left\{\frac{z_{ij}z_{ik}\exp(\mathbf{z}_i^T\boldsymbol{\gamma})}{h(\mathbf{z}_i,\mathbf{x}_i)}\left[1-\exp(\mathbf{z}_i^T\boldsymbol{\gamma})\right]\right\}-\sum_{i=1}^{n}\{z_{ij}z_{ik}v(\mathbf{z}_i)[1-v(\mathbf{z}_i)]\}
\end{aligned}
$$

where $\psi(k)=\frac{\partial\log[\Gamma(k)]}{\partial k}$, $\psi'(k)=\frac{\partial\psi(k)}{\partial k}$, $g_1(\mathbf{x}_i)=\frac{\phi}{\phi+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}$, $g_2(\mathbf{x}_i)=\frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{\phi+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}$, $h(\mathbf{z}_i,\mathbf{x}_i)=\exp(\mathbf{z}_i^T\boldsymbol{\gamma})+\left[\frac{\phi}{\phi+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}\right]^{\phi}$, $v(\mathbf{z}_i)=\frac{\exp(\mathbf{z}_i^T\boldsymbol{\gamma})}{1+\exp(\mathbf{z}_i^T\boldsymbol{\gamma})}$ and $j=k=1,2,\ldots,p$.

Appendix B: Case-weight perturbation scheme

Here we provide the elements considering the case-weight perturbation scheme. The elements of the matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\phi, \boldsymbol{\Delta}_{\boldsymbol{\beta}}, \boldsymbol{\Delta}_{\boldsymbol{\gamma}})^T$ are expressed as

$$\Delta_i = \begin{cases} \frac{[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\{\log[\hat{g}_1(\mathbf{x}_i)]+\hat{g}_2(\mathbf{x}_i)\}}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} & \text{if} \quad y_i = 0, \\ \psi(\hat{\phi}+y_i) - \psi(\hat{\phi}) - \left[\frac{y_i}{\hat{\phi}+\exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})}\right] + \log[\hat{g}_1(\mathbf{x}_i)] + \hat{g}_2(\mathbf{x}_i) & \text{if} \quad y_i > 0. \end{cases}$$

For $j = 1, \ldots, p$.

$$\Delta_{ji} = \begin{cases} -\frac{\hat{\phi}x_{ij}[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\hat{g}_2(\mathbf{x}_i)}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} & \text{if} \quad y_i = 0, \\ y_i x_{ij}\hat{g}_1(\mathbf{x}_i) - \hat{\phi}x_{ij}\hat{g}_2(\mathbf{x}_i) & \text{if} \quad y_i > 0. \end{cases}$$

For $j = 1, \ldots, p$.

$$\Delta_{ji} = \begin{cases} -z_{ij}\hat{v}(\mathbf{z}_i) + \frac{z_{ij}\exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} & \text{if} \quad y_i = 0, \\ -z_{ij}\hat{v}(\mathbf{z}_i) & \text{if} \quad y_i > 0. \end{cases}$$

Appendix C: Explanatory variable perturbation (zero-inflation portions)

Here we provide the elements considering the explanatory variable perturbation scheme (zero-inflation portion). The elements of the matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\phi, \boldsymbol{\Delta}_{\boldsymbol{\beta}}, \boldsymbol{\Delta}_{\boldsymbol{\gamma}})^T$ are expressed as

$$\Delta_i = \begin{cases} -\frac{[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\{\log[\hat{g}_1(\mathbf{x}_i)]+\hat{g}_2(\mathbf{x}_i)\}\hat{\gamma}_t S_z \exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}{[\hat{h}(\mathbf{z}_i,\mathbf{x}_i)]^2} & \text{if} \quad y_i = 0, \\ 0 & \text{if} \quad y_i > 0. \end{cases}$$

For $j = 1, \ldots, p$.

$$\Delta_{ji} = \begin{cases} \frac{x_{ij}\hat{\phi}\hat{\gamma}_t S_z \exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\hat{g}_2(\mathbf{x}_i)}{[\hat{h}(\mathbf{z}_i,\mathbf{x}_i)]^2} & \text{if} \quad y_i = 0, \\ 0 & \text{if} \quad y_i > 0. \end{cases}$$

For $t = j$.

$$\Delta_{ti} = \begin{cases} -S_z\left[\hat{v}(\mathbf{z}_i) - \frac{\exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)}\right] + z_{it}\hat{\gamma}_t S_z\hat{v}(\mathbf{z}_i)[\hat{v}(\mathbf{z}_i)-1] + \frac{z_{it}\hat{\gamma}_t S_z\exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)}\left[1 - \frac{\exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)}\right] & \text{if} \quad y_i = 0, \\ -\hat{v}(\mathbf{z}_i)\{S_z - z_{it}\hat{\gamma}_t[\hat{v}(\mathbf{z}_i)-1]\} & \text{if} \quad y_i > 0. \end{cases}$$

For $t \neq j$ and $j = 1, \ldots, p$.

$$\Delta_{ji} = \begin{cases} z_{ij}\hat{\gamma}_t S_z\hat{v}(\mathbf{z}_i)[\hat{v}(\mathbf{z}_i)-1] + \frac{z_{ij}\hat{\gamma}_t S_z\exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)}\left[1 - \frac{\exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)}\right] & \text{if} \quad y_i = 0, \\ z_{ij}\hat{\gamma}_t S_z\hat{v}(\mathbf{z}_i)[\hat{v}(\mathbf{z}_i)-1] & \text{if} \quad y_i > 0. \end{cases}$$

Appendix D: Explanatory variable perturbation (NB portion)

Here we provide the elements considering the explanatory variable perturbation scheme (NB portion). The elements of the matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\phi, \boldsymbol{\Delta}_{\boldsymbol{\beta}}, \boldsymbol{\Delta}_{\boldsymbol{\gamma}})^T$ are expressed as

$$\Delta_i = \begin{cases} \hat{\phi}\hat{q}(\mathbf{z}_i,\mathbf{x}_i)\{\log[g_1\hat{(}\mathbf{x}_i) + \hat{g}_2(\mathbf{x}_i)]\}\left[\frac{\hat{g}_1(\mathbf{x}_i)}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)}\right] + \hat{q}(\mathbf{z}_i,\mathbf{x}_i)\{2[\hat{g}_1(\mathbf{x}_i)-1] + \hat{g}_2(\mathbf{x}_i)\} & \text{if} \quad y_i = 0, \\ \hat{\beta}_t S_x\hat{g}_2(\mathbf{x}_i)\left\{\hat{g}_2(\mathbf{x}_i) + \frac{y_i}{\hat{\phi}+\exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})} + 2[\hat{g}_1(\mathbf{x}_i)-1]\right\} & \text{if} \quad y_i > 0. \end{cases}$$

For $t = j$.

$$\Delta_{ti} = \begin{cases} \frac{\hat{\phi}S_x[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\hat{g}_2(\mathbf{x}_i)}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)}\left\{x_{it}\hat{\phi}\hat{\beta}_t\hat{g}_2(\mathbf{x}_i)\left[1 - \frac{[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)}\right] + \hat{\beta}_t[\hat{g}_2(\mathbf{x}_i) - x_{it}] - 1\right\} & \text{if} \quad y_i = 0, \\ x_{it}\hat{\beta}_t S_x\hat{g}_2(\mathbf{x}_i)\hat{u}_i - y_i x_{it}\hat{\beta}_t S_x[\hat{g}_2(\mathbf{x}_i)-1] - S_x[\hat{g}_2(\mathbf{x}_i)(\hat{\phi}+y_i) - y_i] & \text{if} \quad y_i > 0. \end{cases}$$

For $t \neq j$ and $j = 1, \ldots, p$.

$$\Delta_{ji} = \begin{cases} \frac{x_{ij}\hat{\phi}\hat{\beta}_t S_x[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\hat{g}_2(\mathbf{x}_i)}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} \left\{ \hat{g}_2(\mathbf{x}_i) \left[\hat{\phi} + 1 - \frac{\hat{\phi}[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} \right] - 1 \right\} & \text{if} \quad y_i = 0, \\ y_i x_{ij}\hat{\beta}_t S_x \left\{1 - \hat{g}_2(\mathbf{x}_i)[3 - 2\hat{g}_2(\mathbf{x}_i)]\right\} + [\hat{g}_2(\mathbf{x}_i) - 1][y_i x_{ij}\hat{\beta}_t S_x \hat{g}_1(\mathbf{x}_i) + x_{ij}\hat{\phi}\hat{\beta}_t S_x \hat{g}_2(\mathbf{x}_i)] & \text{if} \quad y_i > 0. \end{cases}$$

For $j = 1, \ldots, p$.

$$\Delta_{ji} = \begin{cases} \frac{z_{ij}\hat{\phi}\hat{\beta}_t S_x \exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\hat{g}_2(\mathbf{x}_i)}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} & \text{if} \quad y_i = 0, \\ 0 & \text{if} \quad y_i > 0. \end{cases}$$

Appendix E:Simultaneous explanatory variable perturbation (Zero-inflation and NB portion)

Here we provide the elements considering the simultaneous explanatory variable perturbation scheme (Zero-inflation and NB portion). The elements of the matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\phi, \boldsymbol{\Delta}_{\boldsymbol{\beta}}, \boldsymbol{\Delta}_{\boldsymbol{\gamma}})^T$ are expressed as

$$\Delta_i = \begin{cases} -\hat{\phi}\hat{q}_i\{\log[\hat{g}_1(\mathbf{x}_i)] + \hat{g}_2(\mathbf{x}_i)\} + \hat{q}_i\{2[\hat{g}_1(\mathbf{x}_i) - 1] + \hat{g}_2(\mathbf{x}_i)\} - \frac{[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\{\log[\hat{g}_1(\mathbf{x}_i)] + \hat{g}_2(\mathbf{x}_i)\}}{[\hat{h}(\mathbf{z}_i,\mathbf{x}_i)]^2}\hat{b}_i & \text{if} \quad y_i = 0, \\ \hat{\beta}_t S_x \hat{g}_2(\mathbf{x}_i) \left\{ \frac{y_i}{\hat{\phi} + \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})} + \hat{g}_2(\mathbf{x}_i) + 2[\hat{g}_1(\mathbf{x}_i) - 1] \right\} & \text{if} \quad y_i > 0. \end{cases}$$

For $t = j$.

$$\Delta_{ti} = \begin{cases} \frac{x_{it}\hat{\phi}\hat{\beta}_t S_x[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\hat{g}_2(\mathbf{x}_i)}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} \left\{ \hat{g}_2(\mathbf{x}_i)[\hat{\phi} + 1] - 1 \right\} + \frac{\hat{\phi}[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\hat{g}_2(\mathbf{x}_i)}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} \left[\frac{x_{it}}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} - S_x \right] & \text{if} \quad y_i = 0, \\ x_{it}\hat{\beta}_t S_x \hat{g}_2(\mathbf{x}_i)\hat{a}_i - x_{it}\hat{\beta}_t S_x\{y_i[\hat{g}_2(\mathbf{x}_i) - 1] - \hat{\phi}\hat{g}_2(\mathbf{x}_i)\} - S_x[\hat{g}_2(\mathbf{x}_i)(y_i + \hat{\phi}) - y_i] & \text{if} \quad y_i > 0. \end{cases}$$

For $t \neq j$ and $j = 1, \ldots, p$.

$$\Delta_{ji} = \begin{cases} x_{ij}\hat{\phi}\hat{q}_i[\hat{g}_2(\mathbf{x}_i)(\hat{\phi} + 1) - 1] + \frac{x_{ij}\hat{\phi}[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\hat{g}_2(\mathbf{x}_i)}{[\hat{h}(\mathbf{z}_i,\mathbf{x}_i)]^2}\hat{b}_i & \text{if} \quad y_i = 0, \\ y_i x_{ij}\hat{\beta}_t S_x\{1 + \hat{g}_2(\mathbf{x}_i)[2\hat{g}_2(\mathbf{x}_i) - 3]\} + [\hat{g}_2(\mathbf{x}_i) - 1]\{x_{ij}\hat{\beta}_t S_x[y_i\hat{g}_1(\mathbf{x}_i) + \hat{\phi}\hat{g}_2(\mathbf{x}_i)]\} & \text{if} \quad y_i > 0. \end{cases}$$

For $t = j$.

$$\Delta_{ti} = \begin{cases} -S_z\hat{v}(\mathbf{z}_i)\{1 - z_{it}\hat{\gamma}_t[\hat{v}(\mathbf{z}_i) - 1]\} + \frac{S_z\exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)}[1 + z_{it}\hat{\gamma}_t] - \frac{z_{it}\exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}{[\hat{h}(\mathbf{z}_i,\mathbf{x}_i)]^2}\hat{b}_i & \text{if} \quad y_i = 0, \\ S_z\hat{v}(\mathbf{z}_i)\{z_{it}\hat{\gamma}_t[\hat{v}(\mathbf{z}_i) - 1] - 1\} & \text{if} \quad y_i > 0. \end{cases}$$

For $t \neq j$ and $j = 1, \ldots, p$.

$$\Delta_{ji} = \begin{cases} z_{ij}\hat{\gamma}_t S_z\hat{v}(\mathbf{z}_i)[\hat{v}(\mathbf{z}_i) - 1] + \frac{z_{ij}\exp(\mathbf{z}_i^t\hat{\boldsymbol{\gamma}})}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} \left[\hat{\gamma}_t S_z - \frac{\hat{b}_i}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)} \right] & \text{if} \quad y_i = 0, \\ z_{ij}\hat{\gamma}_t S_z\hat{v}(\mathbf{z}_i)[\hat{v}(\mathbf{z}_i) - 1] & \text{if} \quad y_i > 0, \end{cases}$$

where $\hat{g}_1(\mathbf{x}_i) = \frac{\hat{\phi}}{\hat{\phi} + \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})}$, $\hat{g}_2(\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})}{\hat{\phi} + \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})}$, $\hat{h}(\mathbf{z}_i, \mathbf{x}_i) = \exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}}) + \left[\frac{\hat{\phi}}{\hat{\phi} + \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})}\right]^{\hat{\phi}}$, $\hat{v}(\mathbf{z}_i) = \frac{\exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}{1 + \exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}})}$, $\hat{q}_i = \frac{\hat{\beta}_t S_x[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}}{\hat{h}(\mathbf{z}_i,\mathbf{x}_i)}$, $\hat{u}_i = \hat{g}_1(\mathbf{x}_i) + \hat{\phi}[\hat{g}_2(\mathbf{x}_i) - 1] + y_i[2\hat{g}_2(\mathbf{x}_i) - 3]$,
$\hat{b}_i = \hat{\gamma}_t S_z \exp(\mathbf{z}_i^T\hat{\boldsymbol{\gamma}}) - \hat{\phi}\hat{\beta}_t S_x[\hat{g}_1(\mathbf{x}_i)]^{\hat{\phi}}\hat{g}_2(\mathbf{x}_i)$ and $\hat{a}_i = \hat{g}_2(\mathbf{x}_i)(2y_i + \hat{\phi}) + y_i[\hat{g}_1(\mathbf{x}_i) - 3]$.

References

Agarwal, D., Gelfand, A., Citron-Pousty, S., 2002. Zero-inflated models with application to spatial count data. Environmental and Ecological Statistics 9 (4), 341–355.

Atkinson, A., 1981. Two graphical displays for outlying and influential observations in regression. Biometrika 68 (1), 13.

Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L., Kirchner, U., 1999. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. Journal of the Royal Statistical Society. Series A, 195–209.

Carrasco, J., Ortega, E., Paula, G., 2008. Log-modified Weibull regression models with censored data: Sensitivity and residual analysis. Computational Statistics and Data Analysis 52 (8), 4021–4039.

Cho, H., Ibrahim, J. G., Sinha, D., Zhu, H., 2009. Bayesian Case Influence Diagnostics for Survival Models. Biometrics 65 (1), 116.

Cook, R., 1986. Assessment of local influence. Journal of the Royal Statistical Society. Series B (Methodological) 48 (2), 133–169.

Cook, R., Weisberg, S., 1982. Residuals and influence in regression. Chapman and Hall New York.

Davison, A., Hinkley, D., 1997. Bootstrap methods and their application. Cambridge Univ Pr.

DiCiccio, T., Efron, B., 1996. Bootstrap confidence intervals. Statistical Science, 189–212.

Doornik, J., 2007. Object-Oriented Matrix Programming using Ox 3.0, 4th edn, Timberlake Consultants Ltd, London. See http://www. nuff. ox. ac. uk/Users/Doornik, accessed April.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. The Annals of Statistics 7 (1), 1–26.

Efron, B., Tibshirani, R., 1993. An introduction to the bootstrap. Chapman & Hill.

Galea, M., Leiva-Sanchez, V., Paula, G., 2004. Influence diagnostics in log-Birnbaum-Saunders regression models. Journal of Applied Statistics 31 (9), 1049–1064.

Hall, D., 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56 (4), 1030–1039.

Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 1–14.

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

Lipsitz, S., Laird, N., Harrington, D., 1990. Using the jackknife to estimate the variance of regression estimators from repeated measures studies. Communications in Statistics: Theory and Methods 19 (3), 821–845.

Manly, B., 2006. Randomization, bootstrap and Monte Carlo methods in biology. Chapman & Hall/CRC.

Mwalili, S., Lesaffre, E., Declerck, D., 2008. The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. Statistical Methods in Medical Research 17 (2), 123.

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

Silva, G., Ortega, E., Cancho, V., Barreto, M., 2008. Log-Burr XII regression models with censored data. Computational Statistics and Data Analysis.

Svetliza, C., Paula, G., 2003. Diagnostics in nonlinear negative binomial models. Communications in Statistics: Theory and Methods 32 (6), 1227–1250.

Wang, P., Puterman, M., Cockburn, I., Le, N., 1996. Mixed Poisson regression models with covariate dependent rates. Biometrics 52 (2), 381–400.

Xie, F., Wei, B., 2009. Influence analysis for count data based on generalized Poisson regression models. Statistics doi: 10.1080/02331880903138494.

Xie, F., Wei, B., Lin, J., 2008. Assessing influence for pharmaceutical data in zero-inflated generalized Poisson mixed models. Statistics in medicine 27 (18), 3656–3673.

Zeller, C., Labra, F., Lachos, V., Balakrishnan, N., 2009. Influence analyses of skew-normal/independent linear mixed models. Computational Statistics and Data Analysis doi:10.1016/j.csda.2009.11.008.

Zhou, X., Tu, W., 2000. Confidence intervals for the mean of diagnostic test charge data containing zeros. Biometrics 56 (4), 1118–1125.