Revista Colombiana de Estadística - Applied Statistics
July 2021, volume 44, issue 2, pp. 253–278
http://doi.org/10.15446/rce.v44n2.85606

Influence Diagnostics for Correlated Binomial Regression Models: An Application to a Data Set on High-Cost Health Services Occurrence

Diagnósticos de influencia para modelos de regresión binomial correlacionada: una aplicación a un conjunto de datos sobre la ocurrencia de servicios de salud de alto costo

Carlos Diniz[1,a], Rubiane Pires[1,b], Carolina Paraíba[2,c], Paulo Ferreira[2,d]

[1]Department of Statistics, Federal University of São Carlos, São Carlos, Brazil
[2]Department of Statistics, Federal University of Bahia, Salvador, Brazil

Abstract

This paper considers a frequentist perspective to deal with the class of correlated binomial regression models (Pires & Diniz, 2012), thus providing a new approach to analyze correlated binary response variables. Model parameters are estimated by direct maximization of the log-likelihood function. We also consider a diagnostic analysis under the correlated binomial regression model setup, which is performed considering residuals based on predictive values and deviance residuals (Cook & Weisberg, 1982) to check for model assumptions, and global influence measure based on case-deletion (Cook, 1977) to detect influential observations. Moreover, a sensitivity analysis is carried out to detect possible influential observations that could affect the inferential results. This is done using local influence metrics (Cook, 1986) with case-weight, response, and covariate perturbation schemes. A simulation study is conducted to assess the frequentist properties of model parameter estimates and check the performance of the considered diagnostic metrics under the correlated binomial regression model. A data set on high-cost claims made to a private health care provider in Brazil is analyzed to illustrate the proposed methodology.

***Key words*:** Generalized binomial distribution; Health care provider; Influence; Overdispersion; Regression; Residuals.

[a]Ph.D. E-mail: dcad@ufscar.br
[b]Ph.D. E-mail: rubianemariapires@gmail.com
[c]Ph.D. E-mail: carolina.paraiba@ufba.br
[d]Ph.D. E-mail: paulohenri@ufba.br

Resumen

Este artículo considera una perspectiva frecuentista para tratar con la clase de modelos de regresión binomial correlacionada (Pires & Diniz, 2012), proporcionando así un nuevo enfoque para analizar variables de respuesta binaria correlacionadas. Los parámetros del modelo se estiman mediante la maximización directa de la función de log-verosimilitud. También consideramos un análisis de diagnóstico bajo la configuración del modelo de regresión binomial correlacionada, que se realiza considerando los residuos basados en valores predictivos y los residuos de desviación (Cook & Weisberg, 1982) para verificar los supuestos del modelo y la medida de influencia global basada en la eliminación de casos (Cook, 1977) para detectar observaciones influyentes. Además, se realiza un análisis de sensibilidad para detectar posibles observaciones influyentes que podrían afectar los resultados inferenciales. Esto se hace utilizando métricas de influencia local (Cook, 1986) con esquemas de perturbación de covariable, variable respuesta y ponderación de casos. Se realiza un estudio de simulación para evaluar las propiedades frecuentistas de los estimadores de parámetros del modelo y verificar el rendimiento de las métricas de diagnóstico consideradas bajo el modelo de regresión binomial correlacionada. Se analiza un conjunto de datos sobre un plan de salud de un operador brasileño para ilustrar la metodología propuesta.

***Palabras clave*:** Distribución binomial generalizada; Plan de salud; Influencia; Sobredispersión; Regresión; Residuos.

1. Introduction

In real practical situations, the observed data may feature a response variable representing the sum of dependent Bernoulli random variables. McCullagh & Nelder (1989, p. 125) argue that, unless there are good reasons for relying on the binomial assumption, a more prudent approach would be to assume overdispersion to be present in this type of data. Overdispersion is a phenomenon that occurs when a higher variability than that assigned to the usual binomial model is observed in the data and it can be attributed to several causes, such as correlation between the binary responses, absence of relevant explanatory variables, and others.

An alternative to overcome extra-binomial variation is to consider a distribution that generalizes the usual binomial distribution. Among a number of distributions that have been proposed as alternatives to model binary data subject to overdispersion (e.g., the well-known beta-binomial distribution of Skellam (1948), the additive and multiplicative binomial distributions of Altham (1978), and the double-binomial distribution of Efron (1986), among others), we center our attention on the generalized binomial distribution proposed by Luceño (1995). For a detailed discussion on this distribution, see Diniz et al. (2010).

Pires & Diniz (2012) derived a new class of correlated binomial regression models based on the generalized binomial distribution (Luceño, 1995). The authors used a data augmentation scheme to overcome the complexity of the mixture likelihood and a full Bayesian methodology was proposed for model parameter estimation, as well as for model diagnostics.

In this paper, we propose a frequentist model formulation as an alternative approach to the class of correlated binomial regression models with parameters estimated by direct maximization of the log-likelihood function. The methodology presented in this paper circumvents the latent variable setup of Pires & Diniz (2012). The frequentist formulation does not require specification or derivation of prior distribution. Therefore, inference and diagnostics can be performed by means of a well-known, easy-to-implement, and quickly computable optimal procedure when prior information is not available. Moreover, for the data set considered in this paper, if a statistical analysis were to be carried out to study the occurrence of high-cost health services claims by workers made to private health care providers, regulatory agencies would demand objective and standardized procedures where the effect of "ones' opinion" on the outcome is nonexistent, i.e., the analyst's prior (subjective) information on the matter would not be accepted. In this setting, the results provided by the frequentist formulation are much more likely to be preferable, as it is both objective and optimal.

Besides, it is well known that statistical modeling procedures are usually based on initial model assumptions and can be misleading if the fitted model is not plausible enough. Therefore, this paper aims not only to propose a frequentist estimation method for the class of correlated binomial regression models, but also to consider frequentist metrics of model diagnostics. Specifically, we use residuals based on predictive values and deviance residuals (Cook & Weisberg, 1982) to check model assumptions. Furthermore, case-deletion influence diagnostic metrics (Cook, 1977; Cook & Weisberg, 1982), namely the Cook's generalized distance and the likelihood distance, are considered to detect influential observations on parameter estimates. We also perform a sensitivity study to detect influential cases affecting the obtained inferential results by means of local influence measures (Cook, 1986) based on case-weight, response, and covariate perturbation schemes. Two predictive model selection criteria, the Akaike's information criterion (AIC; Akaike 1974) and the Bayesian information criterion (BIC; Schwarz 1978), are used for model selection.

The remainder of the paper is organized as follows. In Section 2, we describe the real data set used in our work. In Section 3, we review the class of correlated binomial regression models proposed by Pires & Diniz (2012). In Section 3.1, we present a discussion on the link function considered for modeling the probability of success parameter and the dependence parameter, and we also discuss the inferential procedure used to conduct parameter estimation for the class of correlated binomial regression models. In Section 4, we develop some diagnostic methods, which consist of two types of residuals (Section 4.1), two metrics of global influence based on case-deletion (Section 4.2), and a local influence metric (Section 4.3). In Section 5, we present results based on simulated data sets to assess the frequentist properties of the estimation procedure and check the performance of the considered diagnostic metrics. Section 6 deals with the illustration of the proposed frequentist methodology for correlated binomial regression models, by considering a data set from a private health care provider in Brazil. Finally, in Section 7, some conclusions are drawn.

2. Health Care Provider Data Set

As a motivational example of overdispersed binomial data analysis using the correlated binomial regression model with parameter estimation and diagnostics both performed under the frequentist perspective, we consider a data set from a private health care provider in Brazil. This data set comprises a portfolio of companies (clusters) for which the occurrence or not of high-cost health services - such as oncological surgery, prosthesis, chemotherapy and hemodialysis - is observed for each employee. The data set is available at `http://www.ufscar.br/~des/docente/carlos/Dados/Dados2.txt`. Information on private health care providers in Brazil are available at `http://www.ans.gov.br/`, the site [1] of the Brazilian *Agência Nacional de Saúde Suplementar* (ANS, freely translated as National Supplementary Health Agency), which is the regulatory agency responsible for the regulation and oversight of privately run health care providers in Brazil.

The available data for the i-th company with n_i employees, $i = 1, 2, \dots, 160$, consist of $W_{i1}, W_{i2}, \dots, W_{in_i}$, each one assuming value 0 or 1, depending on the status of the employee (0 = not occurrence; 1 = occurrence). Thus, the response variable for the i-th company, $Y_i = \sum_{j=1}^{n_i} W_{ij}$, assumes values in $\{0, 1, \dots, n_i\}$ according to the number of employees who have used high-cost health services. For this particular data set, a dependence structure between the Bernoulli variables inside the same company (cluster) could be assumed and explained by the fact that it is reasonable to consider employees within the same company to be exposed to the same environment conditions. This data set also contains the following covariates: average number of medical appointments per employee; average cost of a medical test; occurrence of surgical procedure; number of therapies; number of emergency procedures; number of days between the beginning of the plan period and the first high-cost health service occurrence per each employee; and specific information about the companies (size, number of employees, business activity). From the private health care provider's point of view, the main interest while analyzing this data set would be to fit a regression model able to precisely determine the probability of a high-cost health service occurrence in a company, which would be taken into consideration at the time of renewing - or not - the contract with the company.

3. Correlated Binomial Regression Model

Assume $Y_1, Y_2, \dots, Y_m$ are independent random variables such that each Y_i follows a correlated binomial distribution, denoted by $Y_i \sim \mathrm{CB}(n_i, p_i, \rho_i)$, for $i = 1, 2, \dots, m$. The correlated binomial distribution (Luceño, 1995) provides

[1] The site is in Portuguese with no translation option.

a suitable way to represent the distribution of sums of equicorrelated Bernoulli random variables. This distribution is given by the mixture of the distributions of two random variables, with one of them following a binomial distribution, $\mathrm{B}(n_i, p_i)$, with a mixing probability $(1-\rho_i)$, and the other one following a modified Bernoulli distribution, $\mathrm{MBern}(p_i)$, taking values 0 or n_i (Fu & Sproule, 1995) (rather than the conventional values 0 or 1), with a mixing probability ρ_i. Taking this information into account, Y_i, the number of successes in n_i trials of Bernoulli, $i = 1, 2, \ldots, m$, is the sum of equicorrelated binary responses with a constant probability of success p_i and a common correlation coefficient equal to ρ_i. Thus, $Y_i = \sum_{j=1}^{n_i} W_{ij}$, where W_{ij} is a binary $(0, 1)$ variable with $E(W_{ij}) = p_i$, $\mathrm{Var}(W_{is}) = \mathrm{Var}(W_{it}) = p_i(1-p_i)$ and $\mathrm{Corr}(W_{is}, W_{it}) = \rho_i$, for all s and t, $s \neq t$. The probability distribution of Y_i, given n_i, p_i and ρ_i, is then given by

$$\begin{aligned} P(Y_i = y_i \mid n_i, p_i, \rho_i) = &\binom{n_i}{y_i} p_i^{y_i} (1-p_i)^{n_i - y_i} (1-\rho_i) I_{A_{1i}}(y_i) \\ &+ p_i^{\frac{y_i}{n_i}} (1-p_i)^{\frac{n_i - y_i}{n_i}} \rho_i I_{A_{2i}}(y_i), \end{aligned}$$

where $A_{1_i} = \{0, 1, \ldots, n_i\}$, $A_{2_i} = \{0, n_i\}$, $n_i \in \mathbb{N} - \{0\}$, $0 < p_i < 1$ and $0 \leq \rho_i \leq 1$.

The mean and variance of Y_i are $n_i p_i$ and $p_i(1-p_i)\{n_i + \rho_i n_i (n_i - 1)\}$, respectively. Note that the binomial model is a particular case of the CB (n_i, p_i, ρ_i) model when $\rho_i = 0$. This distribution can be interpreted as a zero-n_i inflated distribution (Lambert, 1992). The zero and n_i values, which occur with greater frequency than expected under the binomial distribution, are captured by the modified Bernoulli distribution. The occurrence of many zero and n_i values can be explained by the positive correlation between the individuals inside the cluster.

We note that observations $Y_1, Y_2, \ldots, Y_m$ are mutually independent and that correlation takes place among the W_{ij}'s, $j = 1, 2, \ldots, n_i$, inside the cluster i, $i = 1, 2, \ldots, m$.

3.1. Inference

Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_m)^\top$ be a set of observed values of response variables $\boldsymbol{Y} = (Y_1, Y_2, \ldots,\ Y_m)^\top$, and $\boldsymbol{n} = (n_1, n_2, \ldots, n_m)^\top$ a vector with the cluster sizes. Then, the likelihood function of $\boldsymbol{p} = (p_1, p_2, \ldots,\ p_m)^\top$, the vector of success probabilities for each cluster, and $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_m)^\top$, the vector of the correlation between any two individuals within the cluster, may be written as

$$\begin{aligned} \mathcal{L}(\boldsymbol{p}, \boldsymbol{\rho} \mid m, \boldsymbol{n}, \boldsymbol{y}) = \prod_{i=1}^{m} \Bigg\{ & a_i \Big((1-p_i)^{n_i} (1-\rho_i) + (1-p_i)\rho_i \Big) \\ & + b_i \Big(p_i^{n_i} (1-\rho_i) + p_i \rho_i \Big) \\ & + (1 - a_i - b_i) \left(\binom{n_i}{y_i} p_i^{y_i} (1-p_i)^{n_i - y_i} (1-\rho_i) \right) \Bigg\}, \end{aligned} \tag{1}$$

where $a_i = 1$ if $y_i = 0$, and $a_i = 0$ otherwise; $b_i = 1$ if $y_i = n_i$, and $b_i = 0$ otherwise. Note that a_i and b_i are known values, $i = 1, 2, \ldots, m$.

In order to define the class of correlated binomial regression models, the success probability, p_i, and the correlation parameter, ρ_i, are jointly modeled using the sets of covariates available for the clusters and for the individuals inside the clusters. Thus, the p_i's are modeled using the link functions Q_i's specified in Table 1, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, and $\eta_i = \sum_{r=0}^{k} \beta_r x_{ir}$. The coefficients $\beta_0, \beta_1, \ldots, \beta_k$ are unknown regression parameters to be estimated; $x_{i0} = 1$, for all i; and $x_{i1}, x_{i2}, \ldots, x_{ik}$ represent the values of the k covariates for the i-th cluster.

Table 1: Some link functions used to model p_i.

Link function	Q_i
Logit	$\exp\{\eta_i\} / (1 + \exp\{\eta_i\})$
Log-log	$\exp\{-\exp\{-\eta_i\}\}$
Complementary log-log	$1 - \exp\{-\exp\{\eta_i\}\}$
Probit	$\Phi(\eta_i)$

The correlation structure is parameterized considering a specific function of available covariates, which is able to relate the dependence between individuals inside the cluster. In general, the correlation structure can be written as

$$R_i = h\left(v\left(\boldsymbol{r}_i\right), \gamma\right),$$

where $h(v(\boldsymbol{r}_i), \gamma)$ is a suitable nonlinear, monotonic and differentiable function, representing the correlation between any two individuals inside the i-th cluster; $v(\boldsymbol{r}_i)$ is a function of the individual covariate values, assuming positive values; $\boldsymbol{r}_i = (r_{i11}, \ldots, r_{i1n_i}, r_{i21}, \ldots, r_{i2n_i}, r_{iq1}, \ldots, r_{iqn_i})^\top$, with r_{ilj} representing the value of the l-th covariate for the j-th individual within the i-th cluster, $i = 1, 2, \ldots, m$, $l = 1, 2, \ldots, q$ and $j = 1, 2, \ldots, n_i$; γ is the parameter determining the rate of decay of the correlation as a function of $v(\boldsymbol{r}_i)$ (Sherman, 2011). Using spatial ideas of correlation structures, the possible choices for $v(\boldsymbol{r}_i)$ can be made considering, for instance, continuous functions of some distance between position vectors or between other available vectors which enable us to characterize the relationship among the individuals within the cluster (Sherman, 2011). Hence, candidates for $v(\boldsymbol{r}_i)$, using only the covariates r_{i1} and r_{i2}, could be the Euclidean distance measure, defined as $\sqrt{\sum_{l=1,2} \sum_s \sum_{s<t} (r_{ils} - r_{ilt})^2}$; the Manhattan distance, defined as $\sum_{l=1,2} \sum_s \sum_{s<t} |r_{ils} - r_{ilt}|$; the maximum distance, defined as $\max_{s,t} |r_{i1s} - r_{i1t}|$; and the minimum distance, defined as $\min_{s,t} |r_{i2s} - r_{i2t}|$, $s, t = 1, 2, \ldots, n_i$.

It is worth pointing out that the dependence between the Bernoulli trials is aggregated into the function of individual covariates of the correlation structure. In other words, after determining a specific function which summarizes the dependence between the individuals, then the observations within the cluster are mutually independent.

Therefore, the likelihood for the correlated binomial regression model can be rewritten as a function of the regression coefficients $\beta_0, \beta_1, \ldots, \beta_k$, associated with the covariates, and of the coefficient γ, associated with the correlation structure. Let the observed data set be $\mathcal{D} = (m, \boldsymbol{n}, \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{r})^\top$, where $\boldsymbol{n} = (n_1, n_2, \ldots, n_m)^\top$, $\boldsymbol{y} = (y_1, y_2, \ldots, y_m)^\top$, $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)^\top$, $\boldsymbol{x}_i = (x_{i0}, x_{i1}, x_{i2}, \ldots, x_{ik})^\top$, $\boldsymbol{r} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_m)^\top$, $\boldsymbol{r}_i = (r_{i11}, \ldots, r_{i1n_i}, r_{i21}, \ldots, r_{i2n_i}, r_{iq1}, \ldots, r_{iqn_i})^\top$. Using a link function Q_i and a correlated structure R_i, the likelihood function (1) can be expressed as a function of $\boldsymbol{\theta} = (\beta_0, \beta_1, \ldots, \beta_k, \gamma)^\top$. Thus,

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta} \mid \mathcal{D}) = \prod_{i=1}^{m} \Bigg\{ & a_i \Bigg((1-Q_i)^{n_i} (1-R_i) + (1-Q_i) R_i \Bigg) \\ & + b_i \Bigg(Q_i^{n_i} (1-R_i) + Q_i R_i \Bigg) \\ & + (1-a_i-b_i) \Bigg(\binom{n_i}{y_i} Q_i^{y_i} (1-Q_i)^{n_i-y_i} (1-R_i) \Bigg) \Bigg\},\end{aligned} \tag{2}$$

where $a_i = 1$ if $y_i = 0$, and $a_i = 0$ otherwise; $b_i = 1$ if $y_i = n_i$, and $b_i = 0$ otherwise, with $i = 1, 2, \ldots, m$. When R_i assumes zero value, we need to consider $R_i = \zeta$, where ζ is a fixed value very close to zero.

Maximum likelihood estimates (MLEs) can be obtained by direct maximization of the log-likelihood function ($\ell(\boldsymbol{\theta} \mid \mathcal{D}) = \log \mathcal{L}(\boldsymbol{\theta} \mid \mathcal{D})$) using, for instance, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Nocedal & Wright, 2006). The advantage of this procedure is that it runs easily from statistical packages such as R (R Development Core Team, 2007). The code implemented in R and used in this paper is available upon request or at `http://www.ufscar.br/~des/docente/carlos/Dados/MRBC_EMV.txt`.

Under some regularity conditions (Lehmann & Casella 1998, Theorem 5.1, p.463), the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is a multivariate normal distribution $\mathrm{N}_{k+2}\left(\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta})\right)$, where $I(\boldsymbol{\theta})$ is the Fisher information matrix which can be approximated by the $(k+2) \times (k+2)$ observed information matrix, $J(\hat{\boldsymbol{\theta}})$, defined as

$$J(\hat{\boldsymbol{\theta}}) = - \left. \frac{\partial^2 \ell(\boldsymbol{\theta} \mid \mathcal{D})}{\partial \boldsymbol{\theta} \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}. \tag{3}$$

The first-order partial derivatives (score functions) and the second-order partial derivatives (Hessian) of the log-likelihood function are given in the appendix.

The asymptotic confidence interval (ACI), at a confidence level $100(1-\alpha)\%$, for the w-th component of the parameter vector $\boldsymbol{\theta}$, θ_w, $w = 1, 2, \ldots, k+2$, can be computed by

$$\hat{\theta}_w \pm \mathcal{Z}_{\alpha/2} \sqrt{J^{-1}_{(w)}(\hat{\boldsymbol{\theta}})}\ ,$$

where $\mathcal{Z}_{\alpha/2}$ is the $(\alpha/2)$-th superior quantile value of the standard normal distribution and $J^{-1}_{(w)}(\hat{\boldsymbol{\theta}})$ is the w-th element of the diagonal of the inverse of $J(\hat{\boldsymbol{\theta}})$, which corresponds to the variance estimate of the w-th parameter estimate.

4. Diagnostics

Diagnostic analysis in the correlated binomial regression model previously described shall be performed by considering two different types of residuals (the standardized residual and the deviance residual), two global influence measures (the generalized Cook's distance and the likelihood distance), and local influence measures based on three perturbation schemes (namely, the case-weight perturbation, the response perturbation and the covariate perturbation). Residual diagnostics are useful to check for model misspecification and for outlier observation. Global and local influence measures are tools designed to assess influential observations.

To check the underlying model assumptions or, in other words, to verify if the response variables follow a correlated binomial distribution, $\text{CB}\,(n_i, p_i, \rho_i)$, with positive correlation between the Bernoulli variables in the cluster (that is, $\rho_i > 0$), the significance of the correlation structure parameter γ is observed using confidence intervals obtained in the inferential process. If $\gamma = 0$ or $\gamma = 1$, the usual binomial regression model can be considered in the analysis.

4.1. Residuals

The standardized residual for the correlated binomial regression model is defined as

$$r_i = \frac{y_i - n_i\hat{p}_i}{\sqrt{\hat{p}_i\,(1-\hat{p}_i)\,\{n_i + \hat{\rho}_i n_i(n_i-1)\}}}, \qquad i = 1, 2, \ldots, m, \tag{4}$$

and the deviance residual for the correlated binomial regression model is defined as

$$r_i^d = \text{sign}\,(y_i - n_i\hat{p}_i)\sqrt{2\ell\,(y_i \mid \mathcal{D}_i, \hat{\gamma}) - 2\ell\left(\hat{\boldsymbol{\beta}} \mid \mathcal{D}_i, \hat{\gamma}\right)}, \qquad i = 1, 2, \ldots, m. \tag{5}$$

For both (4) and (5), $\hat{p}_i = \hat{Q}_i$, $\hat{\rho}_i = \hat{R}_i$ and $\hat{\gamma}$ and $\hat{\boldsymbol{\beta}}$ are the MLEs of the parameters γ and $\boldsymbol{\beta}$, respectively.

In expression (5), $\text{sign}(\cdot)$ is the signal function; $\mathcal{D}_i = (n_i, y_i, \boldsymbol{x}_i, \boldsymbol{r}_i)^\top$; $\ell\,(y_i \,|\, \mathcal{D}_i, \hat{\gamma})$ is the saturated log-likelihood function, with $\hat{p}_i = y_i/n_i$ and the correlation structure parameter γ replaced by the MLE $\hat{\gamma}$; and $\ell\left(\hat{\boldsymbol{\beta}} \,|\, \mathcal{D}_i, \hat{\gamma}\right)$ is the log-likelihood function evaluated at the MLE $\hat{\boldsymbol{\beta}}$.

For a discussion on residuals for models under the generalized linear models framework, such as the correlated binomial regression model addressed in this paper, we refer the interested reader to (Agresti, 2015, Section 4.4.6).

4.2. Global Influence

In order to assess the influence of observations on parameter estimates of the correlated binomial regression model, we consider the generalized Cook's distance and the likelihood distance (Cook, 1977; Cook & Weisberg, 1982; Zhu et al., 2001). Both the generalized Cook's distance and the likelihood distance can be used to quantify the impact of the i-th observation on the MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ by measuring the distance between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{(-i)}$, where $\hat{\boldsymbol{\theta}}_{(-i)}$ is the MLE of $\boldsymbol{\theta}$ based on $\mathcal{L}(\boldsymbol{\theta}\,|\,\mathcal{D})$ with the i-th observation $(n_i, y_i, \boldsymbol{x}_i, \boldsymbol{r}_i)^\top$ deleted from the data set. However, these methodologies are only effective when there is a single outlier in the data set (She & Owen, 2011).

The generalized Cook's distance (Cook, 1977; Cook & Weisberg, 1982) is defined as

$$C_i = \left(\hat{\boldsymbol{\theta}}_{(-i)} - \hat{\boldsymbol{\theta}}\right)^\top J(\hat{\boldsymbol{\theta}}) \left(\hat{\boldsymbol{\theta}}_{(-i)} - \hat{\boldsymbol{\theta}}\right), \tag{6}$$

where $J(\hat{\boldsymbol{\theta}})$ is the observed Fisher information matrix given by (3).

The likelihood distance (Zhu et al., 2001) is defined as

$$LD_i = 2\left\{\ell\left(\hat{\boldsymbol{\theta}} \mid \mathcal{D}\right) - \ell\left(\hat{\boldsymbol{\theta}}_{(-i)} \mid \mathcal{D}\right)\right\}, \tag{7}$$

where $\ell\left(\hat{\boldsymbol{\theta}}\,|\,\mathcal{D}\right)$ and $\ell\left(\hat{\boldsymbol{\theta}}_{(-i)}\,|\,\mathcal{D}\right)$ are the log-likelihood functions evaluated at the usual MLE, $\hat{\boldsymbol{\theta}}$, and at the MLE with the i-th observation $(n_i, y_i, \boldsymbol{x}_i, \boldsymbol{r}_i)^\top$ deleted, $\hat{\boldsymbol{\theta}}_{(-i)}$, respectively. Since $\ell\,(\boldsymbol{\theta}\,|\,\mathcal{D})$, for fixed $\mathcal{D}$, is maximized at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, then, for whatever any other $\boldsymbol{\theta} \neq \hat{\boldsymbol{\theta}}$, $\ell\,(\boldsymbol{\theta}\,|\,\mathcal{D})$ will be less than $\ell\left(\hat{\boldsymbol{\theta}}\,|\,\mathcal{D}\right)$; thus, expression (7) is always positive.

When the number of clusters, m, is large, Cook & Weisberg (1982) suggested the following approximation for $\hat{\boldsymbol{\theta}}_{(-i)}$ in (6) and (7):

$$\hat{\boldsymbol{\theta}}_{(-i)} = \hat{\boldsymbol{\theta}} + J^{-1}(\hat{\boldsymbol{\theta}})U\left(\hat{\boldsymbol{\theta}}_{(-i)}\right),$$

where

$$U\left(\hat{\boldsymbol{\theta}}_{(-i)}\right) = \left.\frac{\partial \ell\left(\boldsymbol{\theta} \mid \mathcal{D}_{(-i)}\right)}{\partial \boldsymbol{\theta}_{(-i)}}\right|_{\boldsymbol{\theta}_{(-i)} = \hat{\boldsymbol{\theta}}_{(-i)}}$$

is a $(k+2)$-dimensional score vector of the log-likelihood function with the i-th observation deleted.

The i-th observation is regarded as an influential case if the value of its generalized Cook's distance or likelihood distance is large. This value can be compared to the critical points of the chi-square distribution with $(k+2)$ degrees of freedom, χ^2_{k+2}.

4.3. Local Influence

Influence diagnostics relying on case-deletion can be regarded as global measures of influence since they are designed to measure global change over the sample space. However, as argued by Cook (1986), single case-deletion diagnostic metrics can experience a kind of masking. Therefore, local influence diagnostic analysis should be conducted to minimize the incorrect inference about case influences. Cook's local influence metrics are designed to investigate model sensibility to minor perturbations in the data set.

Following Cook (1986), let $\ell\left(\boldsymbol{\theta}\,|\,\mathcal{D}\right)$ be the log-likelihood function of the postulated model, and let $\boldsymbol{\omega}$ be a k-dimensional perturbation vector belonging to the perturbation space $\Omega \subset \Re^q$. Denote by $\ell\left(\boldsymbol{\theta}\,|\,\mathcal{D},\boldsymbol{\omega}\right)$ the log-likelihood function of the perturbed model, and assume $\boldsymbol{\omega}_0 \in \Omega$ to be the non-perturbation vector such that $\ell\left(\boldsymbol{\theta}\,|\,\mathcal{D},\boldsymbol{\omega}_0\right) = \ell\left(\boldsymbol{\theta}\,|\,\mathcal{D}\right)$. Thus, the influence of the perturbation $\boldsymbol{\omega}$ on model parameter estimates may be evaluated through the likelihood displacement defined by

$$LD\left(\boldsymbol{\omega}\right) = 2\left\{\ell\left(\hat{\boldsymbol{\theta}}\mid\mathcal{D}\right) - \ell\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}\mid\mathcal{D}\right)\right\}, \tag{8}$$

where $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$ under $\ell\left(\boldsymbol{\theta}\,|\,\mathcal{D}\right)$, and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ is the MLE of $\boldsymbol{\theta}$ under $\ell\left(\boldsymbol{\theta}\,|\,\mathcal{D},\boldsymbol{\omega}\right)$.

As a means to assess local influence, Cook (1986) suggested the study of $LD\left(\boldsymbol{\omega}\right)$ in (8) around $\boldsymbol{\omega}_0$ using the normal curvature of $LD\left(\boldsymbol{\omega}_0 + t\boldsymbol{d}\right)$, with $t \in \Re$ and $\boldsymbol{d}$ a unit-norm direction, which is defined as

$$C_{\boldsymbol{d}}\left(\boldsymbol{\theta}\right) = 2\left|\boldsymbol{d}^\top\Delta^\top J^{-1}(\hat{\boldsymbol{\theta}})\Delta\boldsymbol{d}\right|,$$

where $\|\boldsymbol{d}\| = 1$, $J(\hat{\boldsymbol{\theta}})$ is the observed information matrix, and $\Delta = \partial\ell\left(\boldsymbol{\theta}\,|\,\mathcal{D},\boldsymbol{\omega}\right)/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$ evaluated at $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$. The direction $\boldsymbol{d}_{max}$, which corresponds to the largest curvature, may be used to assess influential observations by means of an index plot. The largest curvature, $C_{\boldsymbol{d}_{max}}$, is given by the largest eigenvalue of $\Delta^\top J^{-1}(\hat{\boldsymbol{\theta}})\Delta$ and $\boldsymbol{d}_{max}$ corresponds to its eigenvector.

Under the correlated binomial regression model presented in Section 3, Δ is a $(k+2)\times q$ matrix of partial derivatives, which may be written as

$$\Delta = \begin{pmatrix}\Delta_{\boldsymbol{\beta}}\\ \Delta_\gamma\end{pmatrix} = \left.\begin{pmatrix}\dfrac{\partial^2\ell\left(\boldsymbol{\theta}\mid\mathcal{D},\boldsymbol{\omega}\right)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\\ \dfrac{\partial^2\ell\left(\boldsymbol{\theta}\mid\mathcal{D},\boldsymbol{\omega}\right)}{\partial\gamma\partial\boldsymbol{\omega}^\top}\end{pmatrix}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\,\boldsymbol{\omega}=\boldsymbol{\omega}_0}. \tag{9}$$

In this paper, we consider case-weight, response and covariate perturbation schemes. In case-weight perturbation, the log-likelihood function of the perturbed model is obtained from

$$\ell\left(\boldsymbol{\theta}\mid\mathcal{D},\boldsymbol{\omega}\right) = \sum_{i=1}^{m}\omega_i\ell_i\left(\boldsymbol{\theta}\mid\mathcal{D}_i\right),$$

with $\boldsymbol{\omega}_0 = \mathbf{1}_m$ the vector of no perturbation, and $\mathcal{D}_i = \left(n_i, y_i, \boldsymbol{x}_i, \boldsymbol{r}_i\right)^\top$, $i = 1, 2, \ldots, m$.

In the response perturbation scheme, $\boldsymbol{y}$ is replaced by $\boldsymbol{y}_{\boldsymbol{\omega}}$, where $y_{\omega_i} = y_i + \omega_i$, $i = 1, 2, \ldots, m$, and $\boldsymbol{\omega}_0 = \boldsymbol{0}_m$. Then, the log-likelihood function is given by

$$\ell\left(\boldsymbol{\theta} \mid \mathcal{D}, \boldsymbol{\omega}\right) = \sum_{i=1}^{m} \ell\left(\boldsymbol{\theta} \mid \mathcal{D}_{\omega_i}\right),$$

with $\mathcal{D}_{\omega_i} = (n_i, y_{\omega_i}, \boldsymbol{x}_i, \boldsymbol{r}_i)$, $i = 1, 2, \ldots, m$.

Under case-weight and response perturbation schemes, the matrix $\Delta_{\boldsymbol{\beta}}$ in (9) is a $(k+1) \times m$ matrix whose elements are

$$\Delta_{\boldsymbol{\beta}_{ij}} = \frac{\partial^2}{\partial \beta_j \partial \omega_i} \ell\left(\boldsymbol{\theta} \mid \mathcal{D}, \boldsymbol{\omega}\right) = \frac{\partial^2 \ell\left(\boldsymbol{\theta} \mid \mathcal{D}, \boldsymbol{\omega}\right)}{\partial Q_i \partial \omega_i} \frac{\partial Q_i}{\partial \beta_j},$$

for $j = 0, 1, \ldots, k$ and $i = 1, 2, \ldots, m$, and the matrix Δ_γ in (9) is a $1 \times m$ matrix with elements

$$\Delta_{\gamma_i} = \frac{\partial^2}{\partial \gamma \partial \omega_i} \ell\left(\boldsymbol{\theta} \mid \mathcal{D}, \boldsymbol{\omega}\right) = \frac{\partial^2 \ell\left(\boldsymbol{\theta} \mid \mathcal{D}, \boldsymbol{\omega}\right)}{\partial R_i \partial Q_i} \frac{\partial Q_i}{\partial \omega_i},$$

for $i = 1, 2, \ldots, m$.

Covariate perturbation is obtained by defining $\boldsymbol{x}_{\omega_i}$, where $x_{\omega_i k} = x_{ik} + \omega_i$, $i = 1, 2, \ldots, m$, and setting $\boldsymbol{\omega}_0 = \boldsymbol{0}_m$. The log-likelihood function is given by

$$\ell\left(\boldsymbol{\theta} \mid \mathcal{D}, \boldsymbol{\omega}\right) = \sum_{i=1}^{m} \ell\left(\boldsymbol{\theta} \mid \mathcal{D}_{\omega_i}\right),$$

with $\mathcal{D}_{\omega_i} = (n_i, y_i, \boldsymbol{x}_{\omega_i}, \boldsymbol{r}_i)$, $i = 1, 2, \ldots, m$.

In this case, the matrix Δ in (9) has elements

$$\Delta_{\boldsymbol{\beta}_{ij}} = \frac{\partial^2 \ell\left(\boldsymbol{\theta} \mid \mathcal{D}, \boldsymbol{\omega}\right)}{\partial Q_i \partial \omega_i} \frac{\partial Q_i}{\partial \beta_j} \frac{\partial Q_i}{\partial \omega_i} + \frac{\partial \ell\left(\boldsymbol{\theta} \mid \mathcal{D}, \boldsymbol{\omega}\right)}{\partial Q_i} \frac{\partial^2 Q_i}{\partial \beta_j \partial \omega_i}$$

and

$$\Delta_{\gamma_i} = \frac{\partial^2 \ell\left(\boldsymbol{\theta} \mid \mathcal{D}, \boldsymbol{\omega}\right)}{\partial R_i \partial Q_i} \frac{\partial Q_i}{\partial \omega_i} \frac{\partial R_i}{\partial \gamma},$$

for $j = 0, 1, \ldots, k$ and $i = 1, 2, \ldots, m$.

To perform local influence diagnostics for a partition of the vector of parameters, say $\boldsymbol{\beta}$ in $\boldsymbol{\theta} = (\boldsymbol{\beta}, \gamma)^\top$, the normal curvature is written as

$$C_{\boldsymbol{d}}\left(\boldsymbol{\beta}\right) = 2 \left| \boldsymbol{d}^\top \boldsymbol{\Delta}^\top \left(J^{-1}(\hat{\boldsymbol{\theta}}) - J_{22} \right) \boldsymbol{\Delta} \boldsymbol{d} \right|,$$

with $J_{22} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & J^{-1}(\gamma) \end{bmatrix}$ and $J\left(\hat{\gamma}\right) = -\frac{\partial^2}{\partial \gamma^2} \ell\left(\boldsymbol{\theta} \,|\, \mathcal{D}\right)\Big|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}}$. Thus, the direction $\boldsymbol{d}_{max}$, that is, the eigenvector of $\boldsymbol{\Delta}^\top \left(J^{-1}(\hat{\boldsymbol{\theta}}) - J_{22} \right) \boldsymbol{\Delta}$ evaluated at $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{w}_0$, is used to asses influential observations on $\boldsymbol{\beta}$. Similarly, we can use this same procedure to assess local influential observations on γ.

5. Simulation Results

In this section, we present results based on simulated data sets to assess the quality of MLEs of model parameters and check the performance of the considered diagnostic metrics under the correlated binomial regression model described in Section 3.

In order to verify the frequentist properties of MLEs under the correlated binomial regression model, 1000 samples/data sets were simulated assuming $m = 500$ clusters with each response variable Y_i following a $\mathrm{CB}\,(n_i, p_i, \rho_i)$ distribution, $i = 1, 2, \ldots, m$. The number of trials per cluster, n_i, was generated from a $\mathrm{B}\,(45, 0.5)$ distribution. We consider two covariates to be available, with x_{i1} and x_{i2} drawn from $\mathrm{N}\,(0, 2)$ and $\mathrm{N}\,(0, 1)$ distributions, respectively, and $v(\boldsymbol{r}_i)$ was simulated from a Uniform $(0, 1)$ distribution. We assume the link functions of Table 1 for p_i and the autoregressive (AR) correlation structure for ρ_i. The parameter values of γ, β_0, β_1 and β_2 are shown[2] in the third column of Table 2.

Table 2: Simulation results.

Fitted model	Parameter	True value	Mean	Bias	MSE	Coverage probability
CB-LOG-AR	γ	0.2	0.2012	0.0012	0.0008	0.95
	β_0	-2	-2.0023	-0.0023	0.0034	0.96
	β_1	0.5	0.5013	0.0013	0.0006	0.95
	β_2	-2	-2.0018	-0.0018	0.0043	0.94
CB-LL-AR	γ	0.1	0.1002	0.0002	0.0003	0.95
	β_0	3	3.0051	0.0051	0.0043	0.95
	β_1	0.1	0.1017	0.0017	0.0006	0.94
	β_2	-0.5	-0.5009	-0.0009	0.0025	0.95
CB-CLL-AR	γ	0.3	0.3019	0.0019	0.0010	0.95
	β_0	-1.5	-1.5023	-0.0023	0.0010	0.94
	β_1	0.5	0.5014	0.0014	0.0002	0.95
	β_2	0.5	0.5010	0.0010	0.0007	0.94
CB-PRO-AR	γ	0.05	0.0477	-0.0023	0.0003	0.89
	β_0	0.01	0.0092	-0.0008	0.0007	0.94
	β_1	0.8	0.7957	-0.0043	0.0006	0.93
	β_2	-0.7	-0.6953	0.0047	0.0009	0.94

Simulation results are presented in Table 2, where the first column indicates the regression model fitted to the data sets. In this case, CB-LOG-AR denotes the correlated binomial regression model with logit link function for p_i and AR correlation structure for ρ_i, CB-LL-AR represents the correlated binomial regression model with log-log link function for p_i and AR correlation structure for ρ_i, CB-CLL-AR denotes the correlated binomial regression model with

[2] It is worth commenting here that we have also considered $x_{i1} \sim \mathrm{Uniform}(0, 1)$ and $x_{i2} \sim \mathrm{Uniform}(0, 2)$, as well as different values of parameters γ, β_0, β_1 and β_2, however, we didn't find any significant difference in the results obtained (not presented here).

complementary log-log link function for p_i and AR correlation structure for ρ_i, and CB-PRO-AR is the correlated binomial regression model with probit link function for p_i and AR correlation structure for ρ_i. The fourth column of Table 2 provides the mean value of the obtained MLEs from each simulated data set, the fifth column shows the mean value of the bias computed for each MLE from each data set, the sixth column provides the computed mean MSE of MLEs, and the seventh column gives the estimated coverage probability based on 95% ACIs.

To summarize, the results shown in Table 2 give evidence that MLEs of the considered correlated binomial regression models have good frequentist properties. It can be noticed that both the mean bias and mean MSE are small, and the coverage probability approaches the expected nominal one of 95% for all model parameters. However, it is worth mentioning that the estimated coverage probability for γ under CB-PRO-AR model is smaller than those observed for this same parameter in the other three models.

In order to assess if the diagnostic metrics are able to correctly detect outliers and/or influential case, we simulated data sets assuming $m = 100$ clusters with $Y_i \sim \text{CB}\,(n_i, p_i, \rho_i)$, $i = 1, 2, \dots, m$, and the number of trials per cluster, n_i, was generated from a $\text{B}\,(50, 0.3)$ distribution. We consider two covariates to be available, with $x_{i1} \sim \text{N}\,(0, 2)$ and $x_{i2} \sim \text{N}\,(0, 1)$, and $v(\boldsymbol{r}_i)$ was simulated from a $\text{Uniform}\,(0, 1)$ distribution. For illustration purposes, we present the results obtained assuming the logit link function for p_i and the AR correlation structure for ρ_i, with parameter values $\beta_0 = -1.5$, $\beta_1 = 0.5$, $\beta_2 = 0.5$ and $\gamma = 0.1$. However, we note that similar results were obtained from the other link functions shown in Table 1.

Prior to generating the values of the response variable, case #7 was deliberately perturbed by transforming the observation of covariate x_2 of case #7 into an atypical one by adding 3 times the standard deviation of x_2 to its original value, namely $x_{7,2} = x_{7,2} + 3 * sd(x_2)$. After simulating the observed values of the response variable, case #40 of $\boldsymbol{y}$ was perturbed by the relation $y_{40} = y_{40} + 2 * sd(\boldsymbol{y})$; since y_{40} must be an integer, we take the rounded value of the transformation.

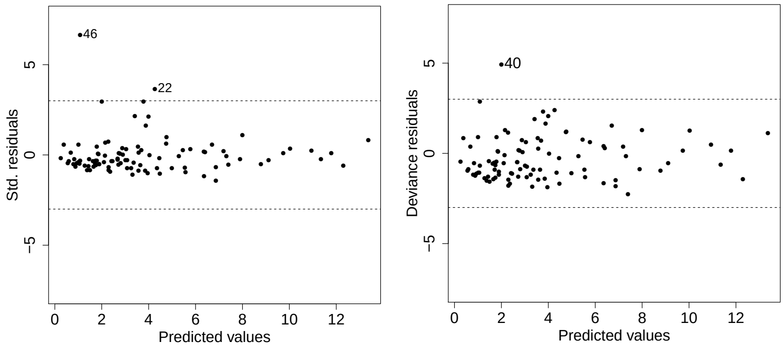
The model fit summary is shown in Table 3. Although the 95% ACIs of model parameters β_0, β_1 and β_2 contain their true values, their MLEs are not as precise as it would be expected for a large sample size. On the other hand, the MLE of γ does not appear to be affected by the perturbation in the data set.

Table 3: Model fit summary for the simulated perturbed data set: MLEs and 95% ACIs of the model parameters γ, β_0, β_1 and β_2.

Parameter	γ	β_0	β_1	β_2
True value	0.1	-1.5	0.5	0.5
MLE	0.108	-1.369	0.454	0.379
95% ACI	(0.027, 0.190)	(-1.573, -1.166)	(0.367, 0.540)	(0.238, 0.520)

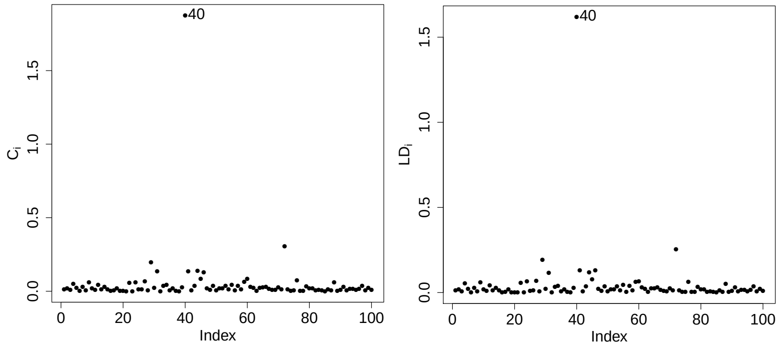
Standardized residuals and deviance residuals are presented in Figure 1(a) and Figure 1(b), respectively. The standardized residuals do not indicate observation #40 as an outlier, but they indicate observations #22 and #46 as possible outliers.

In Figure 1(b), the deviance residuals indicate only observation #40 as an outlier. We note that this behavior of the standardized residuals and deviance residuals was observed for other simulated samples as well, and, while the former ones tend to indicate non-outlier cases as possible outliers the latter ones seem to be a more robust choice for the correlated binomial regression model. From Figures 2(a) and 2(b), we notice that both the Cook's generalized distance and likelihood distance show only case #40 as being influential. Local influence metrics are depicted in Figures 3 and 4, which reveal case #40 as being locally influential only on $\hat{\boldsymbol{\beta}}$, in addition to case #7 under the covariate x_2 perturbation scheme (Figure 3(d)). We note that, if a case is locally influential then it is globally influential, however the inverse relation is not always true.



(a) (b)

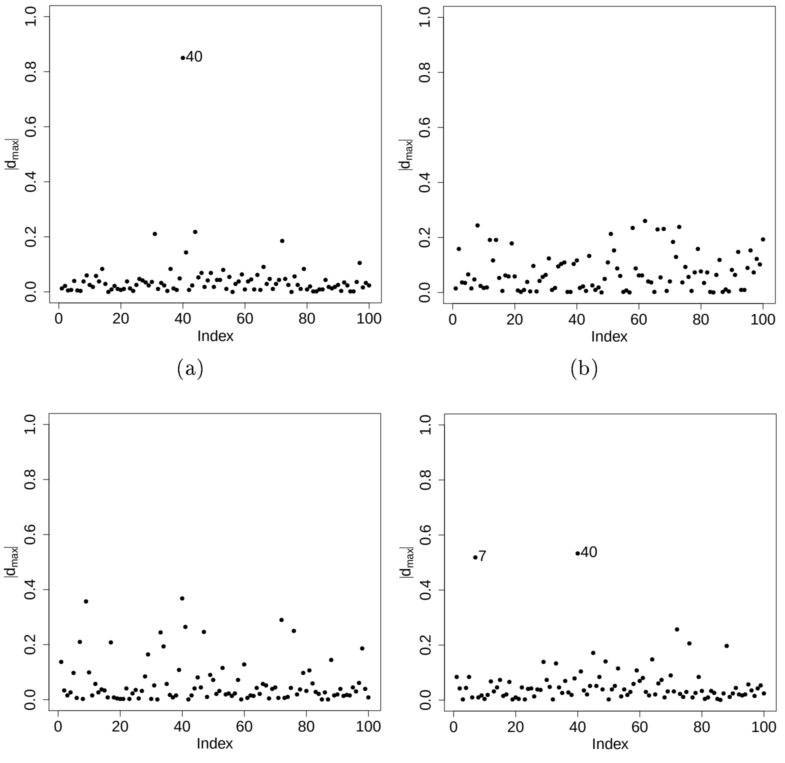
FIGURE 1: Simulated perturbed data set: (a) standardized residuals versus predicted values; (b) deviance residuals versus predicted values.



(a) (b)

FIGURE 2: Simulated perturbed data set: (a) Cook's generalized distance; (b) likelihood distance.

In order to assess the sensitivity of MLEs when influential observations are present in the data set, we also fitted the correlated binomial regression model to the data set without case #7, case #40 and cases #7 and #40, which were identified as influential for the MLE of $\boldsymbol{\beta}$. Table 4 shows the relative difference, $|(\hat{\theta}_w - \hat{\theta}_{w(-i)})/\hat{\theta}_w| \times 100\%$, between MLEs obtained using the perturbed data set and MLEs obtained using the simulated data set after removing the observations. In the first scenario, when we remove only observation #7, the MLE of γ seems to become less precise and there is not a pronounced difference in the MLE of $\boldsymbol{\beta}$. In the second option, when we remove only case #40, we notice that the MLE of β_2 is greatly affected. Removing case #40 also has a small impact on the MLE of γ. Finally, the removal of both cases #7 and #40 only seems to have a pronounced effect in the MLE of β_2.



(a)

(b)

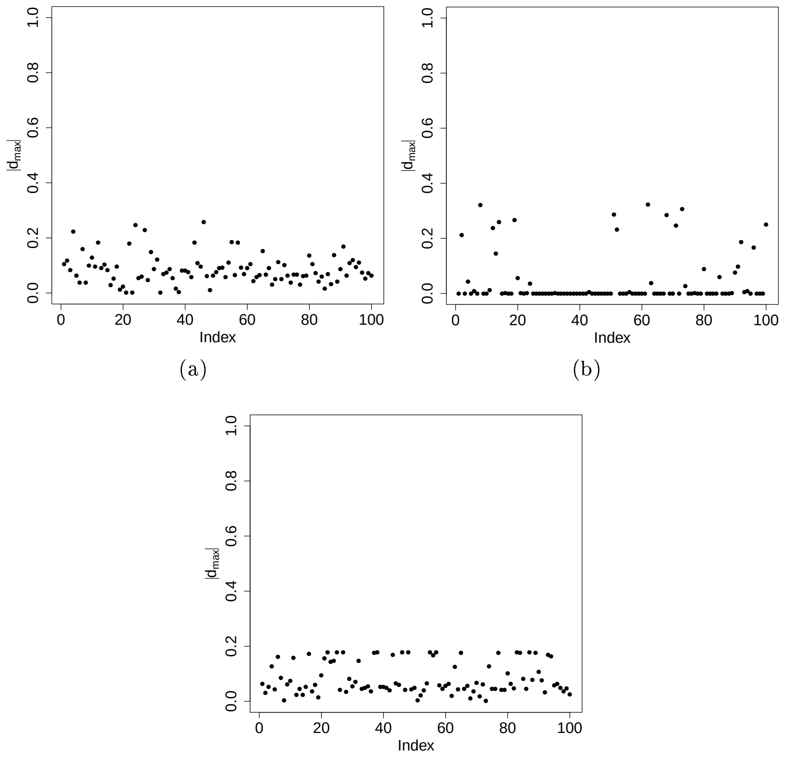
(c)

(d)

Figure 3: Simulated perturbed data set: index plot of $|\boldsymbol{d}_{\max}|$ for $\boldsymbol{\beta}$ under: (a) case-weight perturbation; (b) response perturbation; (c) covariate x_1 perturbation; (d) covariate x_2 perturbation.

Table 4: Simulated perturbed data set: MLEs of complete data set and data sets with case #7, case #40 and cases #7 and #40 deleted, and their relative changes.

Parameter	γ	β_0	β_1	β_2
True value	0.1	-1.5	0.5	0.5
MLE	0.108	-1.369	0.454	0.379
MLE (#7)	0.115	-1.364	0.452	0.377
Relative change (%)	6.48	0.40	0.36	0.48
MLE (#40)	0.098	-1.480	0.467	0.510
Relative change (%)	9.84	8.04	2.83	34.77
MLE (#7, #40)	0.105	-1.476	0.466	0.508
Relative change (%)	3.55	7.80	2.63	34.12



(a)

(b)

(c)

Figure 4: Simulated perturbed data set: index plot of $|\boldsymbol{d}_{\max}|$ for γ under: (a) case-weight perturbation; (b) response perturbation; (c) covariate d perturbation.

6. Application

Next, we analyze the motivational example presented in Section 2 to illustrate the proposed methodology. We also compare the correlated binomial regression models with the beta-binomial regression model (Prentice, 1986), the negative binomial regression model, the Poisson regression model and the usual binomial regression model.

The beta-binomial distribution (Skellam, 1948) is derived by regarding the probability of success parameter p to arise from the beta distribution, $\text{Beta}(\alpha_1, \alpha_2)$, $\alpha_1 > 0$ and $\alpha_2 > 0$. Thus, under the parameterization $p = \alpha_1/(\alpha_1 + \alpha_2)$ and $\rho = 1/(\alpha_1 + \alpha_2 + 1)$, such that $\alpha_1 = p/\zeta$ and $\alpha_2 = (1-p)/\zeta$, where $\zeta = \rho/(1-\rho)$, the beta-binomial distribution can be written as

$$P\left(Y = y \mid n, p, \zeta\right) = \binom{n}{y} \prod_{j=0}^{y-1} \left(p + \zeta j\right) \prod_{j=0}^{n-y-1} \left((1-p) + \zeta j\right) \left[\prod_{j=0}^{n-1} \left(1 + \zeta j\right)\right]^{-1},$$

where $\prod_{j=0}^{x} c_j = 0$, for any $x < 0$, $y = 0, 1, \ldots, n$, $n \in \mathbb{N} - \{0\}$, $0 < p < 1$ and $-1 \leq \rho \leq 1$. The mean and variance of this model are $E\left(Y\right) = np$ and $\text{Var}\left(Y\right) = np\left(1-p\right)\left\{1 + (n-1)\rho\right\}$ (Prentice, 1986).

Let $y_1, y_2, \ldots, y_m$ be a set of observed values of $Y_1, Y_2, \ldots, Y_m$. Then, the log-likelihood function is given by

$$\begin{aligned} \ell_{\text{BB}}\left(\boldsymbol{p}, \boldsymbol{\zeta} \mid m, \boldsymbol{n}, \boldsymbol{y}\right) = \sum_{i=1}^{m} \Bigg\{ & \log \binom{n_i}{y_i} + \sum_{j=0}^{y_i - 1} \log\left(p_i + \zeta_i j\right) \\ & + \sum_{j=0}^{n_i - y_i - 1} \log\left((1-p_i) + \zeta_i j\right) - \sum_{j=0}^{n_i - 1} \log\left(1 + \zeta_i j\right) \Bigg\}, \end{aligned} \tag{10}$$

where $p_i = Q_i$ and $\rho_i = R_i$. The MLEs for the beta-binomial regression model can be obtained by direct maximization of the log-likelihood function (10). We refer the reader to Prentice (1986) for a detailed description of the observed Fisher information matrix, as well as for the asymptotic results for parameter estimates in the beta-binomial regression model.

For the correlated binomial regression model, we considered the link functions of Table 1 for $\boldsymbol{p}$ and the AR correlation structure for $\boldsymbol{\rho}$. For the beta-binomial regression model, we also considered the link functions of Table 1 for $\boldsymbol{p}$ and the AR correlation structure for $\boldsymbol{\zeta}$. The negative binomial and binomial regression models were fitted using the link functions of Table 1 for $\boldsymbol{p}$, and the Poisson regression model was fitted with log link function for $\boldsymbol{p}$.

To model the probability of success of the Bernoulli trials, two covariates are considered: the average number of medical appointments per employee, x_1, and the average cost of a medical test, x_2. The covariate: number of days between the beginning of the plan period and the first high-cost health service occurrence per each employee, r_{ij}, is used to account for the dependence between the Bernoulli variables inside the company. In fact, we consider the variable $\min_{s,t} |r_{is} - r_{it}|$, that is, the minimum of days between the employee s and t, which assumes values between zero, when both employees use the service on the same day, and 365, when there is no use of the plan either by employee s or t. This variable is standardized in the interval $[0, 1]$ using the transformation $\min_{s,t} |r_{is} - r_{it}| = \min_{s,t} |r_{is} - r_{it}|/365$. It is intuitive to assume that the greater the difference between the times of using the plan, the lower the relation between the use of the service. For this reason, the continuous AR correlation structure given by $R_i = \gamma^{\frac{\min_{s,t}|r_{is}-r_{it}|}{365}}$, $i = 1, 2, \ldots, m$ and $s, t = 1, 2, \ldots, n_i$, is considered in the analysis.

The obtained values of AIC and BIC for each model fitted to the health care provider data set are shown in Table 5. From the AIC and BIC values of all models fitted to the data set, it can be seen that the correlated binomial regression model with complementary log-log link function is identified as the best modeling choice. Therefore, in Table 6 we present the model fit summary for the correlated binomial regression model with complementary log-log link function for $\boldsymbol{p}$ and AR correlation structure for $\boldsymbol{\rho}$. From the 95% ACI of the correlation structure parameter, γ, presented in this table, it can be observed that the zero value is not contained in such an interval, thus providing evidence that the data is overdispersed and that the correlated binomial regression model can be used to analyze the health care provider data set.

Table 5: Health care provider data set: AIC and BIC values for all regression models fitted with different link functions †.

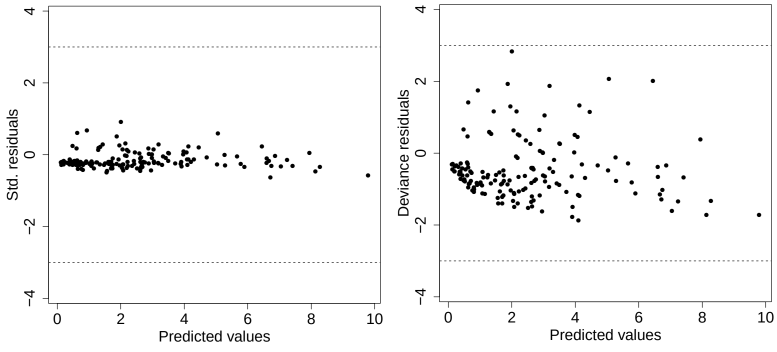
Model	Link function for $\boldsymbol{p}$	AIC	BIC
Correlated Binomial	Logit	397.861	410.161
	Log-log	398.390	410.691
	Complementary log-log	397.811	410.112
	Probit	398.143	410.444
Beta-Binomial	Logit	440.544	452.844
	Log-log	440.518	452.819
	Complementary log-log	440.545	452.846
	Probit	440.530	452.830
Negative Binomial	Logit	466.482	478.783
	Log-log	466.661	478.961
	Complementary log-log	466.431	478.731
	Probit	466.561	478.862
Binomial	Logit	538.471	547.696
	Log-log	554.946	564.172
	Complementary log-log	538.495	547.720
	Probit	538.242	547.468
Poisson	Log	668.343	677.569

† Correlated binomial regression models were fitted with AR correlation structure for $\boldsymbol{\rho}$ and beta-binomial regression models were fitted with AR correlation structure for $\boldsymbol{\zeta}$.

Table 6: Model fit summary for the health care provider data set: MLEs and 95% ACIs of model parameters γ, β_0, β_1 and β_2.

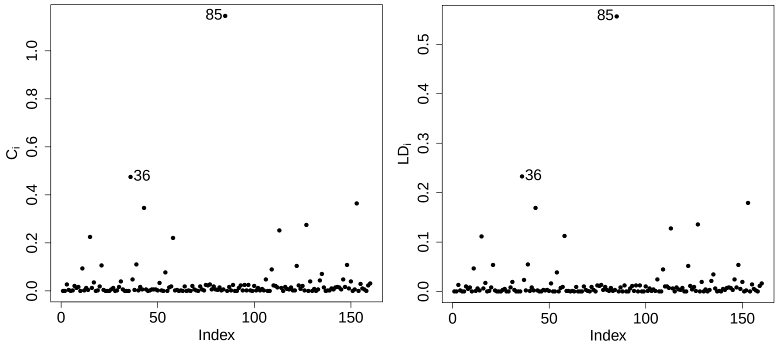
Parameter	γ	β_0	β_1	β_2
MLE	0.223	-3.833	0.206	0.322
95% ACI	(0.121, 0.325)	(-4.411, -3.254)	(0.032, 0.381)	(0.161, 0.484)

The assumption of independence and the presence of outliers may be observed by examining the residuals plotted in time order, if the order was available. Moreover, residuals can also be used to check for model misspecification and for outliers. This can be done by examining the residuals plotted against the predicted values. The standardized residuals based on the predicted values of Y_i, and the deviance residuals based on the log-likelihood function, are presented in Figure 5. Both the standardized residuals, in Figure 5(a), and the deviance residuals, in Figure 5(b), indicate a good specification of the model.



(a) (b)

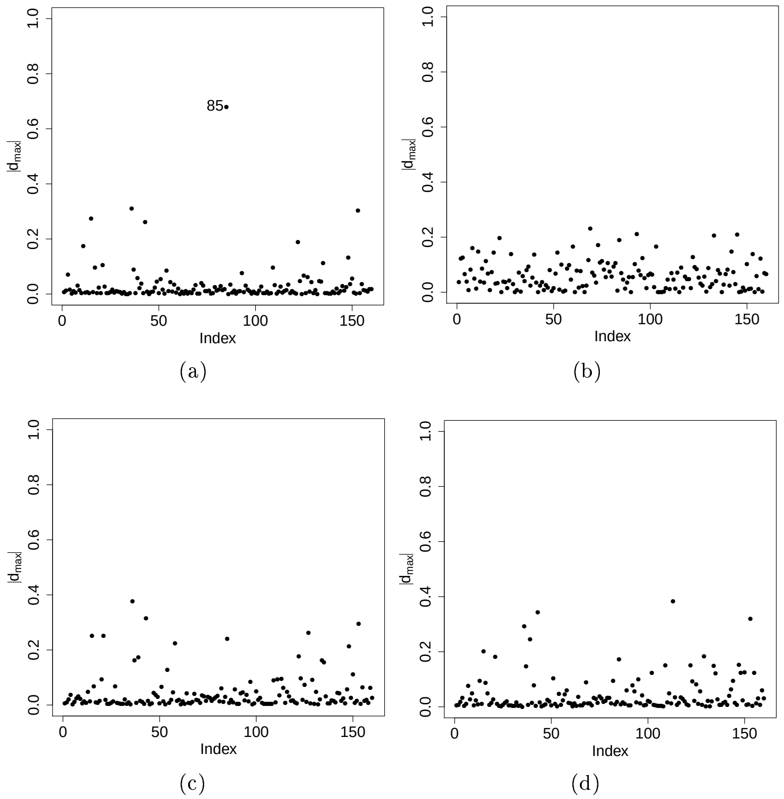
FIGURE 5: Health care provider data set: (a) standardized residuals versus predicted values; (b) deviance residuals versus predicted values.



(a) (b)

FIGURE 6: Health care provider data set: (a) Cook's generalized distance; (b) likelihood distance.

The Cook's generalized distance and the likelihood distance are shown in Figure 6(a) and Figure 6(b), respectively. Both metrics indicate case #36 and case #85 as influential. Local influence metrics are depicted in Figure 7. Based on case-weight perturbation (Figure 7(a)), observation #85 appears as influential. Under the response perturbation (Figure 7(b)), covariate x_1 perturbation (Figure 7(c)) and covariate x_2 perturbation (Figure 7(d)) schemes, no observation is highlighted as influential on $\hat{\boldsymbol{\beta}}$. Notice also, from Figure 8, that there is no influential observation on $\hat{\gamma}$ under any perturbation scheme.



(a)

(b)

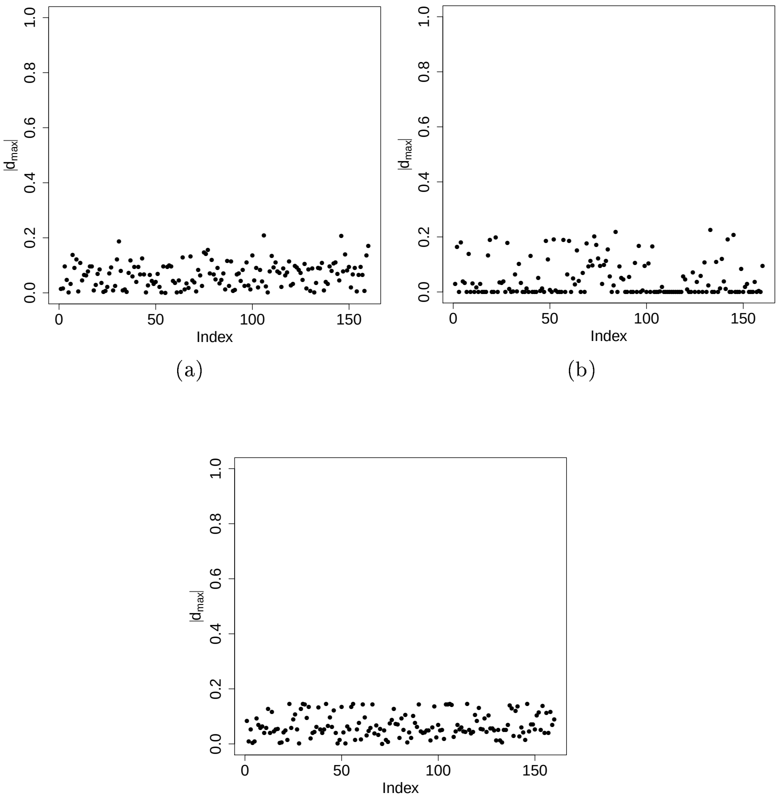
(c)

(d)

FIGURE 7: Health care provider data set: index plot of $|\boldsymbol{d}_{max}|$ for $\boldsymbol{\beta}$ under: (a) case-weight perturbation; (b) response perturbation; (c) covariate x_1 perturbation; (d) covariate x_2 perturbation.

In Table 7, we present the relative difference between MLEs obtained using the complete data set and MLEs obtained using the data set without case #36, case #85 and cases #36 and #85, which were identified as influential for the MLE of $\boldsymbol{\beta}$. We notice that, when we remove only the globally influential observation #36, there is a pronounced difference in the MLE of β_1. When the locally influential case #85 is removed, the MLE of β_1 is greatly affected and the difference in the MLE of β_2 is also noticeable. If both cases #36 and #85 are removed, there is a moderate effect in the MLEs of β_1 and β_2.

In Table 8, we show the model fit summary for the correlated binomial regression models with complementary log-log link function for $\boldsymbol{p}$ and AR correlation structure for $\boldsymbol{\rho}$ fitted to the data set without case #36, case #85 and cases #36 and #85. It can be seen that removing case #36 does not affect which model parameters are statistically significant at a 95% confidence level. On the other hand, removing case #85 or cases #36 and #85 makes the zero value to be contained in the 95% ACI of parameter β_1.



(a)

(b)

(c)

Figure 8: Health care provider data set: index plot of $|\boldsymbol{d}_{max}|$ for γ under: (a) case-weight perturbation; (b) response perturbation; (c) covariate d perturbation.

Table 7: Health care provider data set: MLEs of complete data set and data sets with case #36, case #85 and cases #36 and #85 deleted, and their relative changes.

Parameter	γ	β_0	β_1	β_2
MLE	0.223	-3.833	0.206	0.322
MLE (#36)	0.222	-4.031	0.260	0.317
Relative change (%)	0.44	5.19	25.88	1.58
MLE (#85)	0.225	-3.572	0.117	0.245
Relative change (%)	0.96	6.79	43.18	24.00
MLE (#36, #85)	0.225	-3.774	0.173	0.240
Relative change (%)	0.67	1.53	16.40	25.41

Table 8: Model fit summary for the health care provider data set: MLEs and 95% ACIs of model parameters γ, β_0, β_1 and β_2 with influential observations removed.

Removed cases	Parameter	γ	β_0	β_1	β_2
(#36)	MLE	0.222	-4.031	0.260	0.317
	95% ACI	(0.120, 0.324)	(-4.610, -3.453)	(0.086, 0.434)	(0.156, 0.478)
(#85)	MLE	0.225	-3.572	0.117	0.245
	95% ACI	(0.123, 0.328)	(-4.151, -2.994)	(-0.057, 0.291)	(0.083, 0.406)
(#36, #85)	MLE	0.225	-3.774	0.173	0.240
	95% ACI	(0.123, 0.327)	(-4.352, -3.196)	(-0.002, 0.347)	(0.079, 0.402)

Since we are dealing with a real data set, the removal of influential cases (observations) should not be decided lightly. We notice, for instance, that the collection of more data could reveal the pronounced cases as not being influential. On the other hand, the researcher needs to be made aware that his/her data set contains observations that have been indicated to have a more predominant influence in model parameter estimates. Therefore, based on the analysis conducted in this work and considering the complete data set, the decision regarding the renewal of contracts establishes that the probability of high-cost health service in the i-th company, for this real data set, is given by

$$\hat{p_i} = 1 - \exp\{-\exp\{-3.833 + 0.206x_{i1} + 0.322x_{i2}\}\},$$

with x_{i1}: average number of medical appointments per employee, and x_{i2}: average cost of medical tests. The correlation between any two individuals within the i-th company, for this real data set, is given by $\hat{\rho_i} = 0.223^{v(\boldsymbol{r}_i)}$, with $v(\boldsymbol{r}_i)$: the minimum of days between the employees / 365. We note that the choice of variables to model the probability of success $\boldsymbol{p}$ and the correlation $\boldsymbol{\rho}$ was made based on the covariates provided in the data set that were statistically significant at a 95% confidence level.

7. Conclusions

In this paper, we presented a frequentist approach for the correlated binomial regression model, which is a useful tool to model binomial data subject to overdispersion. Model parameters were estimated by direct maximization of the log-likelihood function. The class of correlated binomial regression models can simultaneously model the probability of success parameter $\boldsymbol{p}$ and the correlation parameter $\boldsymbol{\rho}$. Therefore, we have considered different link functions for the regression structure of $\boldsymbol{p}$ and the AR correlation structure was assumed for $\boldsymbol{\rho}$.

The results based on simulated data sets indicated good asymptotic properties of model parameter estimates. Sensitivity of MLEs under different perturbation schemes was assessed using simulated perturbed data sets and the considered regression diagnostic metrics were indicated to have a good performance under the correlated binomial regression framework.

A real data set on a health care provider in Brazil was analyzed to illustrate the proposed methodology. MLEs of model parameters were computed and diagnostic metrics were used to detect outlier observations as well as globally and locally influential observations with a more pronounced effect in parameter estimation. For this application, the correlated binomial regression model was shown to be a good modeling choice. Moreover, assuming the response variable to follow a correlated binomial distribution provides a more realistic approach to the data, since employees within the same company can be regarded as not necessarily independent.

[Received: March 2020 — Accepted: March 2021]

References

Agresti, A. (2015), *Foundations of Linear and Generalized Linear Models*, Wiley Series in Probability and Statistics, first edn, Wiley, New Jersey.

Akaike, H. (1974), ‘A new look at the statistical model identification’, *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Altham, P. M. E. (1978), ‘Two generalizations of the binomial distribution’, *Journal of the Royal Statistical Society. Series C* **27**(2), 162–167.

Cook, R. D. (1977), ‘Detection of influential observations in linear regression’, *Technometrics* **19**(1), 15–18.

Cook, R. D. (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*, Monographs on statistics and applied probability, Chapman and Hall, London.

Diniz, C. A. R., Tutia, M. H. & Leite, J. G. (2010), ‘Bayesian analysis of a correlated binomial model’, *Brazilian Journal of Probability and Statistics* **24**(1), 68–77.

Efron, B. (1986), ‘Double exponential families and their use in generalized linear regression’, *Journal of the American Statistical Association* **81**(395), 709–721.

Fu, J. & Sproule, R. (1995), ‘A generalization of the binomial distribution’, *Communications in Statistics - Theory and Methods* **24**(10), 2645–2658.

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

Lehmann, E. L. & Casella, G. (1998), *Theory of point estimation*, second edn, Springer, New York.

Luceño, A. (1995), ‘A family of partially correlated poisson models for overdispersion’, *Computational Statistics and Data Analysis* **20**(5), 511–520.

McCullagh, P. & Nelder, J. A. (1989), *Generalized Linear Models*, second edn, Chapman and Hall, London.

Nocedal, J. & Wright, S. J. (2006), *Numerial Optimization*, second edn, Springer-Verlag, New York.

Pires, R. M. & Diniz, C. A. R. (2012), 'Correlated binomial regression models', *Computational Statistics and Data Analysis* **56**(8), 2513–2525.

Prentice, R. L. (1986), 'Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors', *Journal of the American Statistical Association* **81**(394), 321–327.

R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org

Schwarz, G. (1978), 'Estimating the dimension of a model', *Annals of Statistics* **6**(2), 461–464.

She, Y. & Owen, A. B. (2011), 'Outlier detection using nonconvex penalized regression', *Journal of the American Statistical Association* **106**(494), 626–639.

Sherman, M. (2011), *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*, Wiley Series in Probability and Statistics, John Wiley and Sons.

Skellam, J. G. (1948), 'A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials', *Journal of the Royal Statistical Society, Series B* **10**(2), 257–261.

Zhu, H., Lee, S.-Y., Wei, B.-C. & Zhou, J. (2001), 'Case-deletion measures for models with incomplete data', *Biometrika* **88**(3), 727–737.

Appendix. First- and Second-Order Partial Derivatives of the Log-Likelihood Function

In this appendix, we first show the score functions of the log-likelihood function $\ell(\boldsymbol{\theta}\,|\,\mathcal{D}) = \log \mathcal{L}(\boldsymbol{\theta}\,|\,\mathcal{D})$, where $\mathcal{L}(\boldsymbol{\theta}\,|\,\mathcal{D})$ is given by (2). These quantities are obtained as follows:

$$U(\beta_r) = \frac{\partial \ell(\boldsymbol{\theta} \mid \mathcal{D})}{\partial \beta_r} = \sum_{i=1}^{m} \Bigg\{ \frac{\partial g^{-1}(\eta_i)}{\partial \beta_r} \Bigg[C \Big[y_i \left(g^{-1}(\eta_i)\right)^{-1} - (n_i - y_i)\left(1 - g^{-1}(\eta_i)\right)^{-1} \Big] \\ + D \Big[\frac{y_i}{n_i} \left(g^{-1}(\eta_i)\right)^{-1} - \frac{(n_i - y_i)}{n_i} \left(1 - g^{-1}(\eta_i)\right)^{-1} \Big] \Bigg] (C + D)^{-1} \Bigg\},$$

for $r = 0, 1, \ldots, k$, and

$$U(\gamma) = \frac{\partial \ell(\boldsymbol{\theta} \mid \mathcal{D})}{\partial \gamma} = \sum_{i=1}^{m} \left\{ \frac{\partial h\left(v\left(\boldsymbol{r}_i\right), \gamma\right)}{\partial \gamma} \cdot \frac{(-A+B)}{(C+D)} \right\}.$$

In addition, the second-order partial derivatives (Hessian) of the log-likelihood function are as follows:

$$\frac{\partial^2 \ell(\boldsymbol{\theta} \mid \mathcal{D})}{\partial \gamma^2} = \sum_{i=1}^{m} \left\{ \frac{\partial^2 h\left(v\left(\boldsymbol{r}_i\right), \gamma\right)}{\partial \gamma^2} \cdot \frac{(-A+B)}{(C+D)} - \left(\frac{\partial h\left(v\left(\boldsymbol{r}_i\right), \gamma\right)}{\partial \gamma} \right)^2 \cdot \frac{(-A+B)^2}{(C+D)^2} \right\},$$

$$\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta} \mid \mathcal{D})}{\partial \beta_r \partial \gamma} = & \sum_{i=1}^{m} \Bigg\{ \Bigg[\frac{\partial h\left(v\left(\boldsymbol{r}_i\right), \gamma\right)}{\partial \gamma} \Bigg[A \Big[-y_i \left(g^{-1}(\eta_i)\right)^{-1} + (n_i - y_i) \left(1 - g^{-1}(\eta_i)\right)^{-1} \Big] \\
& + B \left[\frac{y_i}{n_i} \left(g^{-1}(\eta_i)\right)^{-1} - \frac{(n_i - y_i)}{n_i} \left(1 - g^{-1}(\eta_i)\right)^{-1} \right] \Bigg] \frac{\partial g^{-1}(\eta_i)}{\partial \beta_r} (C+D)^{-1} \\
& - (C+D)^{-2} \Bigg[\frac{\partial g^{-1}(\eta_i)}{\partial \beta_r} \Bigg[C \Big[y_i \left(g^{-1}(\eta_i)\right)^{-1} - (n_i - y_i) \left(1 - g^{-1}(\eta_i)\right)^{-1} \Big] \\
& + D \left[\frac{y_i}{n_i} \left(g^{-1}(\eta_i)\right)^{-1} - \frac{(n_i - y_i)}{n_i} \left(1 - g^{-1}(\eta_i)\right)^{-1} \right] \Bigg] \left[\frac{\partial h\left(v\left(\boldsymbol{r}_i\right), \gamma\right)}{\partial \gamma} (-A+B) \right] \Bigg] \Bigg\}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta} \mid \mathcal{D})}{\partial \beta_r \partial \beta_s} = & \sum_{i=1}^{m} \Bigg\{ \Bigg[\frac{\partial g^{-1}(\eta_i)}{\partial \beta_r} \cdot \frac{\partial g^{-1}(\eta_i)}{\partial \beta_s} \Bigg[C \Big[y_i^2 \Big(g^{-1}(\eta_i)\Big)^{-2} - y_i \Big(g^{-1}(\eta_i)\Big)^{-2} \\
& - 2y_i(n_i - y_i) \Big(g^{-1}(\eta_i)\Big)^{-1} \Big(1 - g^{-1}(\eta_i)\Big)^{-1} + (n_i - y_i)^2 \Big(1 - g^{-1}(\eta_i)\Big)^{-2} \\
& - (n_i - y_i) \Big(1 - g^{-1}(\eta_i)\Big)^{-2} \Big] + D \Bigg[\frac{y_i^2}{n_i^2} \Big(g^{-1}(\eta_i)\Big)^{-2} - \frac{y_i}{n_i} \Big(g^{-1}(\eta_i)\Big)^{-2} \\
& - 2 \frac{y_i}{n_i^2} (n_i - y_i) \Big(g^{-1}(\eta_i)\Big)^{-1} \Big(1 - g^{-1}(\eta_i)\Big)^{-1} + \frac{(n_i - y_i)^2}{n_i^2} \Big(1 - g^{-1}(\eta_i)\Big)^{-2} \\
& - \frac{(n_i - y_i)}{n_i} \Big(1 - g^{-1}(\eta_i)\Big)^{-2} \Bigg] \Bigg] + \frac{\partial^2 g^{-1}(\eta_i)}{\partial \beta_r \partial \beta_s} \Bigg[C \Big[y_i \Big(g^{-1}(\eta_i)\Big)^{-1} - (n_i - y_i) \\
& \times \Big(1 - g^{-1}(\eta_i)\Big)^{-1} \Big] + D \Bigg[\frac{y_i}{n_i} \Big(g^{-1}(\eta_i)\Big)^{-1} - \frac{(n_i - y_i)}{n_i} \Big(1 - g^{-1}(\eta_i)\Big)^{-1} \Bigg] \Bigg] \Bigg] \\
& \times (C+D)^{-1} - \Bigg[\frac{\partial g^{-1}(\eta_i)}{\partial \beta_r} \Bigg[C \Big[y_i \Big(g^{-1}(\eta_i)\Big)^{-1} - (n_i - y_i) \Big(1 - g^{-1}(\eta_i)\Big)^{-1} \Big] \\
& + D \left[\frac{y_i}{n_i} \Big(g^{-1}(\eta_i)\Big)^{-1} - \frac{(n_i - y_i)}{n_i} \Big(1 - g^{-1}(\eta_i)\Big)^{-1} \right] \Bigg]^2 \frac{\partial g^{-1}(\eta_i)}{\partial \beta_s} \Bigg] (C+D)^{-2} \Bigg\},
\end{aligned}$$

for $r, s = 0, 1, \ldots, k$, where

$$\begin{aligned}
A &= \binom{n_i}{y_i} \left(g^{-1}(\eta_i)\right)^{y_i} \left(1 - g^{-1}(\eta_i)\right)^{n_i - y_i},\\
B &= \left(g^{-1}(\eta_i)\right)^{\frac{y_i}{n_i}} \left(1 - g^{-1}(\eta_i)\right)^{\frac{n_i - y_i}{n_i}} I_{A_{2i}}(y_i),\\
C &= \binom{n_i}{y_i} \left(g^{-1}(\eta_i)\right)^{y_i} \left(1 - g^{-1}(\eta_i)\right)^{n_i - y_i} \left(1 - h\left(v\left(\boldsymbol{r}_i\right), \gamma\right)\right) \quad \text{and}\\
D &= \left(g^{-1}(\eta_i)\right)^{\frac{y_i}{n_i}} \left(1 - g^{-1}(\eta_i)\right)^{\frac{n_i - y_i}{n_i}} h\left(v\left(\boldsymbol{r}_i\right), \gamma\right) I_{A_{2i}}(y_i).
\end{aligned}$$

The explicit formulae of the related derivatives are given in Tables A1 and A2 for some link functions and correlation structures, respectively.

Table A1: First- and second-order derivatives of each link function with respect to $\boldsymbol{\beta}$.

Link function	$\frac{\partial g^{-1}(\eta_i)}{\partial \beta_r}$	$\frac{\partial^2 g^{-1}(\eta_i)}{\partial \beta_r \partial \beta_s}$
Logit	$x_{ir} \exp\{\eta_i\} \left[1 + \exp\{\eta_i\}\right]^{-2}$	$-x_{ir} x_{is} \exp\{\eta_i\} \left[\exp\{\eta_i\} - 1\right] \left[1 + \exp\{\eta_i\}\right]^{-3}$
Log-log	$x_{ir} \exp\{-\eta_i - \exp\{-\eta_i\}\}$	$-x_{ir} x_{is} [\exp\{-\eta_i - \exp\{-\eta_i\}\} - \exp\{-2\eta_i - \exp\{-\eta_i\}\}]$
Complementary log-log	$x_{ir} \exp\{\eta_i - \exp\{\eta_i\}\}$	$x_{ir} x_{is} [\exp\{\eta_i - \exp\{\eta_i\}\} - \exp\{2\eta_i - \exp\{\eta_i\}\}]$
Probit	$x_{ir} \phi(\eta_i)$ *	$-(2\pi)^{-\frac{1}{2}} x_{ir} x_{is} \eta_i \exp\{-\eta_i^2/2\}$

* $\phi(\cdot)$ is the standard normal probability density function.

Table A2: First- and second-order derivatives of some correlation structures with respect to γ.

Correlation structure	$\frac{\partial h\left(v\left(\boldsymbol{r}_i\right), \gamma\right)}{\partial \gamma}$	$\frac{\partial^2 h\left(v\left(\boldsymbol{r}_i\right), \gamma\right)}{\partial \gamma^2}$
Exponential	$-v\left(\boldsymbol{r}_i\right) \exp\{-\gamma v\left(\boldsymbol{r}_i\right)\}$	$\left[v\left(\boldsymbol{r}_i\right)\right]^2 \exp\{-\gamma v\left(\boldsymbol{r}_i\right)\}$
Gaussian	$-2\gamma \left[v\left(\boldsymbol{r}_i\right)\right]^2 \exp\{-\left[\gamma v\left(\boldsymbol{r}_i\right)\right]^2\}$	$\frac{2\left[v\left(\boldsymbol{r}_i\right)\right]^2 \exp\{-\left[\gamma v\left(\boldsymbol{r}_i\right)\right]^2\}}{\left(2\left[\gamma v\left(\boldsymbol{r}_i\right)\right]^2 - 1\right)^{-1}}$
AR	$v\left(\boldsymbol{r}_i\right) \lambda^{v\left(\boldsymbol{r}_i\right)} \lambda^{-1}$	$\gamma^{v\left(\boldsymbol{r}_i\right) - 2} v\left(\boldsymbol{r}_i\right) \left[v\left(\boldsymbol{r}_i\right) - 1\right]$