Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 542–560
DOI: 10.1214/16-BJPS323

Bias correction in power series generalized nonlinear models

Priscila G. Silva, Audrey H. M. A. Cysneiros and Gauss M. Cordeiro

Universidade Federal de Pernambuco

Abstract. Power series generalized nonlinear models [*Comput. Statist. Data Anal.* **53** (2009) 1155–1166] can be used when the Poisson assumption of equidispersion is not valid. In these models, we consider a more general family of discrete distributions for the response variable and a nonlinear structure for the regression parameters, although the dispersion parameter and other shape parameters are assumed known. We derive a general matrix formula for the second-order bias of the maximum likelihood estimate of the regression parameter vector in these models. We use the results by [*J. Roy. Statist. Soc. B* **30** (1968) 248–275] and bootstrap technique [*Ann. Statist.* **7** (1979) 1–26] to obtain the bias-corrected maximum likelihood estimate. Simulation studies are performed using different estimates. We also present an empirical application.

1 Introduction

Count data occur in several different areas. In recent years, the number of published papers dealing with statistical analysis for univariate count data within the framework of regression models has been increased steadily. Poisson and negative binomial distributions are the most useful models in the regression analysis of count data [see, the book by Cameron and Trivedi (1998)]. The Poisson distribution is the cornerstone model for count data. For many observed count data, however, it is common to have the sample variance to be greater or smaller than the sample mean which are referred to as over-dispersion and under-dispersion, respectively. These types of data may arise due to one or more possible causes such as heterogeneity and aggregation for over-dispersion and repulsion for under-dispersion. Consequently, there have been both studies of the effect of over-dispersion on inferences made under a Poisson model and other models have been suggested for accommodating over-dispersion in statistical analysis. Several methods have been proposed for dealing with extra-Poisson variation when doing regression analysis of count data.

Power series generalized nonlinear models (PSGNLMs), pioneered by Cordeiro, Andrade and De Castro (2009), are defined by a modified power series family of distributions for the response (parameterized in terms of the mean) and a possible nonlinear link function for the mean response. This class of models

Key words and phrases. Bias correction, discrete distribution, maximum likelihood, scoring method, nonlinear model.

Received August 2015; accepted June 2016.

unifies several important discrete models in the same framework thus extending the classical log-nonlinear models, binomial nonlinear models and negative binomial nonlinear models to cope with several other discrete distributions.

The random component of the PSGNLM is defined by a subclass of the modified power series family of distributions originally defined and studied by Gupta (1974). However, we express the family parameterized in terms of the mean parameter as developed and studied by Consul (1990). This wider discrete family of distributions combined with the systematic component of the exponential family nonlinear model (EFNLM), first defined by Cordeiro and Paula (1989), constitutes a flexible tool for statistical modeling of discrete data and a rich illustration on the use of univariate discrete regression models for practical applications.

A central object in asymptotic likelihood theory is the calculation of the second-order biases of the maximum likelihood estimates (MLEs). To improve the accuracy of these estimates, substantial effort has gone into computing the cumulants of log-likelihood derivatives which are, however, notoriously cumbersome. The MLEs typically have biases of order $O(n^{-1})$ for large sample size n, which are commonly ignored in practice, the justification being that they are small when compared to the standard errors of the parameter estimates that are of order $O(n^{-1/2})$. For small samples sizes, however, these biases can be appreciable and of the same magnitude of the corresponding standard errors. In these cases, the biases cannot be neglected, and for turning feasible estimation of their size in practical applications, corresponding formulae for their calculation need to be established for a wide range of regression models.

The paper is organized as follows. In Section 2, we define the PSGNLMs. In Section 3, we obtain the bias-corrected MLEs in these models. Simulation results are presented and discussed in Section 4. Concluding remarks are given in Section 5.

2 Power series generalized nonlinear models

We consider discrete random variables $Y_1, \ldots, Y_n$ in Y which are independent and each Y_i follows a family of distributions with mean parameter $\mu_i > 0$ and dispersion parameter $\phi > 0$ defined by the probability mass function with respect to Lebesgue measure

$$\pi(y; \mu_i, \phi) = \frac{a(y, \phi) g(\mu_i, \phi)^y}{f(\mu_i, \phi)}, \qquad y \in A_s, \tag{2.1}$$

where the support of Y_i is a subset A_s of integers $\{s, s+1, \ldots\}$ defined here not depending upon unknown parameters, $s \geq 0$, $a(y; \phi)$ is positive, and the analytic functions $f = f(\mu; \phi)$ and $g = g(\mu; \phi)$ (of the mean parameter μ and the common dispersion parameter ϕ) are positive, finite and twice-differentiable. The dispersion parameter ϕ is assumed known. We have $E(Y) = \mu = \frac{f'g}{fg'}$ and

$\mathrm{Var}(Y) = V(\mu, \phi) = \frac{g}{g'}$. From now on, the primes denote differentiation with respect to μ. We introduce a nonlinear regression structure for the mean vector $\mu = E(Y)$ of the class of distributions (2.1) given by the systematic component

$$h(\mu_i) = \eta_i = \eta(x_i; \boldsymbol{\beta}), \qquad i = 1, \ldots, n, \tag{2.2}$$

where $h(\cdot)$ is a known one-to-one differentiable link function, $\eta(\cdot; \cdot)$ is a specified nonlinear function of unknown parameters, x_i is a $q \times 1$ vector and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ (for $p < n$) is a set of unknown parameters to be estimated. Further, we assume that $\boldsymbol{\beta}$ is defined in a subset $\Omega_{\boldsymbol{\beta}}$ of $I\!R^p$ $(p < n)$ and $\eta(x_i; \boldsymbol{\beta})$ is an injective and continuously differentiable function with respect to $\boldsymbol{\beta}$ such that the $n \times p$ derivative matrix of the nonlinear predictor $\widetilde{X} = \widetilde{X}(\boldsymbol{\beta}) = \partial\eta/\partial\boldsymbol{\beta}^\top$ has rank p for all $\boldsymbol{\beta}$. The local model matrix $\widetilde{X}$ in general depends on the unknown parameter vector $\boldsymbol{\beta}$.

Let $\ell = \ell(\boldsymbol{\beta})$ be the total log-likelihood function for the PSGNLM defined above. We have

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log\{a(y_i, \phi)\} + \sum_{i=1}^{n} [y_i \log\{g(\mu_i, \phi)\} - \log\{f(\mu_i, \phi)\}]. \tag{2.3}$$

The log-likelihood is assumed to satisfy the usual regularity conditions of large sample likelihood theory; see, for instance, Cox and Hinkley (1974).

The expected information matrix for $\boldsymbol{\beta}$ conditioning on ϕ is given by $K_{\boldsymbol{\beta}} = \widetilde{X}^\top W \widetilde{X}$, where $W = \mathrm{diag}\{V_i^{-1} h_i'^{-2}\}$ and $V_i = V(\mu; \phi)$. The information matrix depends only on the model matrix, the variance function and the first derivative of the link function. A nonlinear optimization method such as the Fisher scoring algorithm is required to obtain the MLE $\hat{\boldsymbol{\beta}}$; see Cordeiro, Andrade and De Castro (2009).

3 Bias of the estimate of $\boldsymbol{\beta}$

Bias correction has been extensively studied in the statistical literature and there has been considerable interest in finding simple matrix expressions for second-order biases of MLEs in some classes of regression models that do not involve cumulants of log-likelihood derivatives. The methodology has been applied to several regression models in recent years. We cite the following models: normal nonlinear models [Cook, Tsai and Wei (1986)], generalized linear models [Cordeiro and McCullagh (1991)], multivariate nonlinear regression models [Cordeiro and Vasconcellos (1997)], symmetric nonlinear regression models [Cordeiro et al. (2000)], Student t regression models with unknown degrees of freedom [Vasconcellos and Silva (2005)] and beta regression models [Ospina, Cribari-Neto and Vasconcellos (2006)].

The purpose of this section is to use Cox and Snell (1968) formula (20) for the n^{-1} bias of the MLE in order to obtain the second-order bias of $\hat{\boldsymbol{\beta}}$. We derive a simple matrix formula for the bias of $\hat{\boldsymbol{\beta}}$. We shall use the following notation for the derivatives of the log-likelihood function: $U_r = \partial\ell/\partial\beta_r$, $U_{rs} = \partial^2\ell/\partial\beta_r\,\partial\beta_s$, $U_{rst} = \partial^3\ell/\partial\beta_r\,\partial\beta_s\,\partial\beta_t$, and so on. The notation used for the moments of such derivatives is that of Lawley (1956): $\kappa_{rs} = \mathrm{E}(U_{rs})$, $\kappa_{r,s} = \mathrm{E}(U_r U_s)$, $\kappa_{rst} = \mathrm{E}(U_{rst})$, etc., where all $\kappa' s$ refer to a total over the sample and are, in general, typically of order $O(n)$. We also define the derivatives of the cumulants by $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, etc. Further, we use the notation proposed by Cordeiro and Paula (1989): $\tilde{x}_{ir} = \partial\eta_i/\partial\beta_r$, $\tilde{x}_{irs} = \partial^2\eta_i/\partial\beta_r\,\partial\beta_s$ and $\tilde{x}_{irst} = \partial^3\eta_i/\partial\beta_r\,\partial\beta_s\,\partial\beta_t$.

The first, second- and third-order derivatives of the log-likelihood function (2.3) are

$$U_r = \sum_{i=1}^{n} d_{0i}\tilde{x}_{ir}, \qquad U_{rs} = \sum_{i=1}^{n}(d_{1i}\tilde{x}_{is}\tilde{x}_{ir} + d_{0i}\tilde{x}_{irs})$$

and

$$U_{rst} = \sum_{i=1}^{n}\bigl(d_{3i}\tilde{x}_{it}\tilde{x}_{is}\tilde{x}_{ir} + d_{1i}(\tilde{x}_{ist}\tilde{x}_{ir} + \tilde{x}_{is}\tilde{x}_{irt} + \tilde{x}_{it}\tilde{x}_{irs}) + d_{0i}\tilde{x}_{irst}\bigr),$$

where $d_{0i} = y_i t_i - q_i$, $d_{ji} = y_i t_i^{(j)} - q_i^{(j)}\{(h_i')^j\}^{-1}$, $d_{3i} = d_{2i} - d_{1i}h_i^{(2)}\{(h_i')^2\}^{-1}$, $t_i = g_i'\{g_i h_i'\}^{-1}$ and $q_i = f_i'\{f_i h_i'\}^{-1}$. Here, the superscript (j) indicates the jth differentiation with respect to the mean μ for $j = 1, 2$ and $i = 1, \dots, n$. Taking expected values of such derivatives, we obtain the joint cumulants

$$\kappa_{rs} = \sum_{i=1}^{n} w_{1i}\tilde{x}_{is}\tilde{x}_{ir}, \qquad \kappa_{rs}^{(t)} = \sum_{i=1}^{n}\{\tilde{w}_{1i}\tilde{x}_{it}\tilde{x}_{is}\tilde{x}_{ir} + \tilde{x}_{ist}\tilde{x}_{ir} + w_{1i}\tilde{x}_{is}\tilde{x}_{irt}\},$$

and

$$\kappa_{rst} = \sum_{i=1}^{n}\bigl\{w_{3i}\tilde{x}_{it}\tilde{x}_{is}\tilde{x}_{ir} + w_{1i}(\tilde{x}_{ist}\tilde{x}_{ir} + \tilde{x}_{is}\tilde{x}_{irt} + \tilde{x}_{it}\tilde{x}_{irs})\bigr\},$$

where

$$w_{ji} = \left(\frac{f_i' g_i}{f_i g_i'} t_i^{(j)} - q_i^{(j)}\right)\frac{1}{h_i'}, \qquad w_{3i} = w_{2i} - \frac{w_{1i}h_i''}{(h_i')^2},$$

$$\tilde{w}_{ji} = \varphi_{ji} - \frac{(j-1)q_i V_i t_i^{(j)} h_i'' - q_i^{(j+1)}}{(h_i')^{j+1}} + j\,\frac{q_i^{(j)} h_i''}{(h_i')^{j+2}},$$

$$\varphi_{ji} = \frac{q_i' V_i t_i^{(j)} + q_i V_i' t_i^{(j)} + q_i V_i t_i^{(j+1)}}{(h_i')^j}.$$

These quantities involve derivatives that depend upon well-known functions f, g, h and V of the PSGNLMs.

Cox and Snell (1968) obtained a general formula for the second-order bias (i.e. of order $O(n^{-1})$) of the MLE $\hat{\boldsymbol{\beta}}$ of the parameter vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$. Let $B(\hat{\beta}_b)$ be the second-order bias of the estimate $\hat{\beta}_b$. We can write (for $b = 1, \dots, p$)

$$B(\hat{\beta}_b) = \sum_{r,s,t} \kappa^{br}\kappa^{st}\left(\kappa_{rs}^{(t)} - \frac{1}{2}\kappa_{rst}\right), \tag{3.1}$$

where the indices r, s and t refer to the components of $\boldsymbol{\beta}$. Here, $-\kappa^{rs} = \kappa^{r,s}$ denotes the (r, s)th element of the inverse expected information matrix. The quantity $\kappa_{rs}^{(t)} - \frac{1}{2}\kappa_{rst}$ in (3.1) can be written as

$$\kappa_{rs}^{(t)} - \frac{1}{2}\kappa_{rst} = \sum_{i=1}^{n} c_i \tilde{x}_{it}\tilde{x}_{is}\tilde{x}_{ir} + \frac{1}{2}\sum_{i=1}^{n} w_{1i}(\tilde{x}_{ist}\tilde{x}_{ir} + \tilde{x}_{is}\tilde{x}_{irt} - \tilde{x}_{it}\tilde{x}_{irs}),$$

where $c_i = \tilde{w}_{1i} - \frac{1}{2}\{w_{2i} - w_{1i}h_i''(h_i')^{-2}\}$, for $i = 1, \dots, n$. Thus,

$$B(\hat{\beta}_b) = \sum_{r,s,t} \kappa^{br}\kappa^{st}\sum_i c_i \tilde{x}_{it}\tilde{x}_{is}\tilde{x}_{ir} + \frac{1}{2}\sum_{r,s,t}\kappa^{br}\kappa^{st}\sum_i w_{1i}\tilde{x}_{ist}\tilde{x}_{ir}.$$

In matrix notation, we can write the $O(n^{-1})$ bias of $\hat{\boldsymbol{\beta}}$ as

$$B(\hat{\boldsymbol{\beta}}) = (\widetilde{X}^\top W \widetilde{X})^{-1}\widetilde{X}^\top W\delta. \tag{3.2}$$

Here, $\delta = (Z_d c + \frac{1}{2}D\delta_1)$, Z_d and D are diagonal matrices of order n, $Z = \widetilde{X}(\widetilde{X}^\top W\widetilde{X})^{-1}\widetilde{X}^\top$, $d_i = \mathrm{tr}\{(\widetilde{X}^\top W\widetilde{X})^{-1}\widetilde{\widetilde{X}}_i\}$, $\widetilde{\widetilde{X}}_i$ is a square matrix of order p defined by the elements $\tilde{x}_{irs}$, c and δ_1 are vectors of order $n \times 1$ whose elements are c_i and w_{1i}, respectively.

Thus, the bias vector $B(\hat{\boldsymbol{\beta}})$ is simply the set of coefficients from the weighted linear regression of δ on the columns of $\widetilde{X}$ with weighted matrix W. In the regression calculations, all quantities have to be evaluated at $\hat{\boldsymbol{\beta}}$. For generalized linear models (GLMs), (3.2) coincides with the result (4.2) due to Cordeiro and McCullagh (1991). For the linear model, $\widetilde{\widetilde{X}}_i = 0$ and, consequently, $\delta = Z_d c$. Equation (3.2) is the main result of this paper and can be used to produce a *bias-reduced* estimate by subtracting the bias approximation from the MLE. Alternatively, an examination of the form of the bias may suggest a reparametrization of the model to yield less biased estimates.

A number of remarks are worth making with respect to (3.2). First, $B(\hat{\boldsymbol{\beta}})$ is a function of the local model matrix $\widetilde{X}$, the first two derivatives of the scalars t, q and link function and the first derivatives of the scalars f, g and variance function. Second, to evaluate the n^{-1} bias we need only to compute the asymptotic covariance matrix Z of the estimate $\hat{\eta}$ and the diagonal matrices Z_d and D and the square matrices $\widetilde{\widetilde{X}}_i$ for $i = 1, \dots, n$. It is obvious that (3.2) does depend on the fitted model through the quantities above. Third, it is possible to obtain a closed-form expression for $B(\hat{\boldsymbol{\beta}})$ in models with closed-form information matrix. Fourth, the right-hand side of (3.2) can be evaluated at $\hat{\boldsymbol{\beta}}$ to define the bias–corrected estimate $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \hat{B}(\hat{\boldsymbol{\beta}})$, where $\hat{B}(\cdot)$ is the value of $B(\cdot)$ at $\hat{\boldsymbol{\beta}}$. The bias-corrected estimate $\tilde{\boldsymbol{\beta}}$ is expected to have better sampling properties than the classical estimate $\hat{\boldsymbol{\beta}}$. In fact, several simulation results presented in the literature by Botter and Cordeiro (1998), Cordeiro et al. (2000), Vasconcellos and Silva (2005) and Ospina, Cribari-Neto and Vasconcellos (2006) have shown that the bias-corrected estimates $\tilde{\boldsymbol{\beta}}$ have smaller biases than their corresponding uncorrected estimates, thus suggesting that the bias corrections have the effect of shrinking the corrected estimates toward to the true parameter values. However, we can not conclude that the bias-corrected estimates offer always some improvement over the MLEs, since they can have larger mean squared errors than the uncorrected estimates.

An alternative approach to obtain bias-corrected MLEs is through Efron (1979) bootstrap resampling. Consider a random sample $y = (y_1, \dots, y_n)^\top$ of a variable Y with distribution function $F = F_\theta(y)$, where θ is the parameter that indexes the distribution, that is, it is viewed as a functional of F, say $\theta = t(F)$. Let $\hat{\theta}$ be an estimator for θ based on y, say $\hat{\theta} = s(y)$. The application of the bootstrap technique consists in obtaining, from the original sample y, a large number of pseudo-samples $y^* = (y_1^*, \dots, y_n^*)^\top$ and then extracting information from these copies to improve inference. In the parametric version, the bootstrap samples are generated from $F(\hat{\theta})$, which is denoted by $F_{\hat{\theta}}$. Hence, the bias can be expressed as $B_F(\hat{\theta}, \theta) = \mathrm{E}(\hat{\theta} - \theta) = \mathrm{E}_F[s(y)] - t(F)$, where the subscript F indicates that expectation is taken with respect to F. The bootstrap bias estimate can be obtained by replacing the true distribution F, which generated the original sample, by $F_{\hat{\theta}}$ in the above expression. Then, the parametric estimate of the bias is given by $B_{\hat{F}_{\hat{\theta}}}(\hat{\theta}, \theta) = \mathrm{E}_{\hat{F}_{\hat{\theta}}}[s(y)] - t(\hat{F}_{\hat{\theta}})$.

If N bootstrap samples $y^{*1}, \dots, y^{*N}$ are generated independently from the original sample y, and the corresponding bootstrap replications $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*N}$ are calculated, where $\hat{\theta}^{*i} = s(y^{*i})$ for $i = 1, \dots, N$, then it is possible to approximate the expectation $\mathrm{E}_{F_{\hat{\theta}}}[s(y)]$ by the sample mean $\hat{\theta}^{*(\cdot)} = \frac{1}{N}\sum_{i=1}^N \hat{\theta}^{*i}$. So, the bootstrap bias estimate based on N replications of $\hat{\theta}$ is given by $\hat{B}_{\hat{F}_{\hat{\theta}}}(\hat{\theta}, \theta) = \hat{\theta}^{*(\cdot)} - s(y)$. Finally, we define the bootstrap bias-corrected estimate by $\breve{\theta}_2 = s(y) - \hat{B}_{\hat{F}_{\hat{\theta}}}(\hat{\theta}, \theta) = 2s(y) - \hat{\theta}^{*(\cdot)}$.

Asymptotic confidence intervals for any regression parameter can be constructed based on the multivariate normal approximations $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, K(\boldsymbol{\beta})^{-1})$ and $\tilde{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, K(\boldsymbol{\beta})^{-1})$. Hence, for large n, the asymptotic confidence interval (ACI) and the corrected asymptotic confidence interval (CACI) for $\boldsymbol{\beta}$ with confidence level $100(1-\gamma)\%$ are given by $(\hat{\boldsymbol{\beta}} - z_{1-\frac{\gamma}{2}}(K(\hat{\boldsymbol{\beta}})^{-1})^{1/2}, \hat{\boldsymbol{\beta}} + z_{1-\frac{\gamma}{2}}(K(\hat{\boldsymbol{\beta}})^{-1})^{1/2})$ and $(\tilde{\boldsymbol{\beta}} - z_{1-\frac{\gamma}{2}}(K(\tilde{\boldsymbol{\beta}})^{-1})^{1/2}, \tilde{\boldsymbol{\beta}} + z_{1-\frac{\gamma}{2}}(K(\tilde{\boldsymbol{\beta}})^{-1})^{1/2})$, respectively.

The Bootstrap-t confidence interval (BtCI) is calculated from the distribution of the T statistic defined by $T = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\widehat{\mathrm{se}}(\hat{\boldsymbol{\beta}})$, which can be computed from a random sample $y = (y_1, \dots, y_n)^\top$ under the assumption of normality, where $\widehat{\mathrm{se}}(\hat{\boldsymbol{\beta}})$ is the estimated standard error of $\hat{\boldsymbol{\beta}}$. We generate N bootstrap samples $(y^{*1}, \dots, y^{*N})$ from the original sample y and, for each bootstrap sample, we determine $T^{*b} = (\hat{\boldsymbol{\beta}}^{*b} - \boldsymbol{\beta})/\widehat{\mathrm{se}}^{*b}$, for $b = 1, \dots, N$, where $\hat{\boldsymbol{\beta}} = s(y)$ is the estimated value of $\boldsymbol{\beta}$ calculated from the original sample y, $\hat{\boldsymbol{\beta}}^{*b} = s(y^{*b})$ is the estimated value of $\boldsymbol{\beta}$ and $\widehat{\mathrm{se}}^{*b}$ is the estimated standard error of $\hat{\boldsymbol{\beta}}^{*b}$ for the bootstrap sample y^{*b}. The percentiles $\gamma/2$ and $1 - \gamma/2$ of T^{*b} are estimated by $\hat{t}^{(\gamma/2)}$ and $\hat{t}^{(1-\gamma/2)}$, respectively, such that

$$\frac{\#(T^{*b} \le \hat{t}^{(\gamma/2)})}{N} = \frac{\gamma}{2} \quad \text{and} \quad \frac{\#(T^{*b} \le \hat{t}^{(1-\gamma/2)})}{N} = 1 - \frac{\gamma}{2}.$$

In this way, the Bootstrap-t confidence interval is given by $(\hat{\boldsymbol{\beta}} - \hat{t}^{(1-\gamma/2)}\widehat{\mathrm{se}}(\hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}} - \hat{t}^{(\gamma/2)}\widehat{\mathrm{se}}(\hat{\boldsymbol{\beta}}))$. The quantities $\hat{t}^{(\gamma/2)}$ and $\hat{t}^{(1-\gamma/2)}$ are calculated as follows: we sort all N bootstrap values of T^{*b} and the numbers of replications $N \times (\gamma/2)$ and $N \times (1 - \gamma/2)$ are the quantities $\hat{t}^{(\gamma/2)}$ and $\hat{t}^{(1-\gamma/2)}$, respectively, assuming that $N \times (\gamma/2)$ and $N \times (1 - \gamma/2)$ are integers. If $N \times (\gamma/2)$ and $N \times (1 - \gamma/2)$ are not integers, we can use the following approach: assuming that $0 < \gamma < 1/2$, let k be the greatest integer $\le (N + 1) \times (\gamma/2)$. Then, the quantities $\hat{t}^{(\gamma/2)}$ and $\hat{t}^{(1-\gamma/2)}$ correspond to the kth and $(N + 1 - k)$th ordered elements of T^{*b}, respectively.

4 Numerical evidence

We present some Monte Carlo simulation results on the finite sample behavior of the MLE of $\boldsymbol{\beta}$ and its bias-adjusted counterpart. First, we use the nonlinear regression model $\eta_i = \beta_0 + \beta_1 x_{1i} + \exp(\beta_2 x_{2i})$, $i = 1, \dots, n$. The response was generated from the Consul, generalized negative binomial (GNB) and generalized Poisson (GP) distributions, respectively. For the GNB and GP models, we generate the response variate by fixing the parameters at $\beta_0 = \beta_1 = \beta_2 = 0, 25$, whereas for the Consul model, we take $\beta_0 = \beta_1 = 0, 25$ and $\beta_2 = 1$. The independent variable x_1 and x_2 are chosen as independent random draws from the uniform $U(0, 1)$ distribution. The sample size was taken as $n = 25, 35, 45$ and 100 and the simulations are based on 10,000 replications. All simulations are performed using Ox [Doornik (2009)]. We obtain the MLE $\hat{\boldsymbol{\beta}}$ and compute the adjusted estimate $\tilde{\boldsymbol{\beta}}$. Then, we perform $B = 600$ replications of the parametric bootstrap in order to compute the bootstrap bias-corrected estimate, say $\check{\boldsymbol{\beta}}$.

The figures in Tables 1, 2 and 3 refer to the estimates of $\boldsymbol{\beta}$ for the GNB, Consul and GP models, respectively. Here, we take different sample sizes. In these tables, we present the relative biases for different estimates. It is noteworthy that the corrected estimates $\tilde{\boldsymbol{\beta}}$ have relative biases in magnitude smaller than those of the corresponding MLEs for almost all sample sizes. The bias-corrected estimates obtained by the analytical approach are usually larger than the true parameter values,

Table 1 *Point estimation of $\boldsymbol{\beta}$—GNB nonlinear model*

		β_0			β_1			β_2		
n	Estimate	bias	RB (%)	MSE	bias	RB (%)	MSE	bias	RB (%)	MSE
25	$\hat{\boldsymbol{\beta}}$	−0.01346	−5.3830	0.00018	0.01991	7.9626	0.00040	−0.36879	−147.5200	0.13601
	$\tilde{\boldsymbol{\beta}}$	0.01081	4.3219	0.00012	0.01274	5.0974	0.00016	0.29076	116.3000	0.08454
	$\breve{\boldsymbol{\beta}}$	0.02302	9.2065	0.00053	0.00259	1.0360	0.00001	−0.12964	−51.8560	0.01681
35	$\hat{\boldsymbol{\beta}}$	−0.01428	−5.7134	0.00020	0.00891	3.5657	0.00008	−0.18050	−72.2020	0.03258
	$\tilde{\boldsymbol{\beta}}$	0.00413	1.6527	0.00002	0.00483	1.9303	0.00002	0.05699	22.7950	0.00325
	$\breve{\boldsymbol{\beta}}$	0.01267	5.0671	0.00016	−0.00714	−2.8548	0.00005	0.01945	7.7794	0.00038
45	$\hat{\boldsymbol{\beta}}$	−0.01095	−4.3806	0.00012	0.00733	2.9328	0.00005	−0.12069	−48.2760	0.01457
	$\tilde{\boldsymbol{\beta}}$	0.00157	0.6277	0.00000	0.00377	1.5079	0.00001	0.01979	7.9149	0.00039
	$\breve{\boldsymbol{\beta}}$	0.00418	1.6701	0.00002	−0.00211	−0.8436	0.00000	0.02625	10.5000	0.00069
100	$\hat{\boldsymbol{\beta}}$	−0.00638	−2.5503	0.00004	0.00329	1.3179	0.00001	−0.03543	−14.1730	0.00126
	$\tilde{\boldsymbol{\beta}}$	−0.00027	−0.1093	0.00000	0.00088	0.3531	0.00000	0.00258	1.0336	0.00001
	$\breve{\boldsymbol{\beta}}$	0.00004	0.0179	0.00000	−0.00032	−0.1279	0.00000	0.02822	11.2860	0.00080

Table 2 *Point estimation of $\boldsymbol{\beta}$—Consul nonlinear model*

		β_0			β_1			β_2		
n	Estimate	bias	RB (%)	MSE	bias	RB (%)	MSE	bias	RB (%)	MSE
25	$\hat{\boldsymbol{\beta}}$	−0.04912	−19.6480	0.00241	0.04250	17.0000	0.00181	−0.09605	−9.6052	0.00923
	$\tilde{\boldsymbol{\beta}}$	0.00162	0.6481	0.00000	0.00678	2.7120	0.00005	0.03357	3.3573	0.00113
	$\breve{\boldsymbol{\beta}}$	0.00240	0.9583	0.00001	−0.00810	−3.2395	0.00007	0.03982	3.9815	0.00159
35	$\hat{\boldsymbol{\beta}}$	−0.03749	−14.9980	0.00141	0.01488	5.9517	0.00022	−0.03500	−3.5004	0.00123
	$\tilde{\boldsymbol{\beta}}$	0.00026	0.1024	0.00000	0.00496	1.9853	0.00002	0.00401	0.4006	0.00002
	$\breve{\boldsymbol{\beta}}$	0.00708	2.8332	0.00005	−0.01029	−4.1159	0.00011	0.01781	1.7808	0.00032
45	$\hat{\boldsymbol{\beta}}$	−0.03018	−12.0710	0.00091	0.01474	5.8950	0.00022	−0.02971	−2.9714	0.00088
	$\tilde{\boldsymbol{\beta}}$	−0.00099	−0.3971	0.00000	0.00521	2.0840	0.00003	0.00228	0.2285	0.00001
	$\breve{\boldsymbol{\beta}}$	0.00254	1.0157	0.00001	−0.00446	−1.7848	0.00002	0.00980	0.9801	0.00010
	$\hat{\boldsymbol{\beta}}$	−0.01278	−5.1129	0.00016	0.00388	1.5520	0.00002	−0.01204	−1.2035	0.00014
100	$\tilde{\boldsymbol{\beta}}$	−0.00011	−0.0456	0.00000	0.00088	0.3512	0.00000	0.00082	0.0822	0.00000
	$\breve{\boldsymbol{\beta}}$	0.00187	0.7470	0.00000	−0.00058	−0.2337	0.00000	0.00088	0.0876	0.00000

Table 3 *Point estimation of $\boldsymbol{\beta}$—GP nonlinear model*

		β_0			β_1			β_2		
n	Estimate	bias	RB (%)	MSE	bias	RB (%)	MSE	bias	RB (%)	MSE
25	$\hat{\boldsymbol{\beta}}$	−0.01477	−5.9072	0.00022	0.01921	7.6838	0.00037	−0.38494	−153.9800	0.14818
	$\tilde{\boldsymbol{\beta}}$	0.01101	4.4019	0.00012	0.01354	5.4166	0.00018	0.34852	139.4100	0.12146
	$\breve{\boldsymbol{\beta}}$	0.02582	10.3280	0.00067	0.00305	1.2188	0.00001	−0.15162	−60.6460	0.02299
35	$\hat{\boldsymbol{\beta}}$	−0.01491	−5.9645	0.00022	0.00894	3.5778	0.00008	−0.18724	−74.8940	0.03506
	$\tilde{\boldsymbol{\beta}}$	0.00439	1.7553	0.00002	0.00513	2.0533	0.00003	0.06964	27.8570	0.00485
	$\breve{\boldsymbol{\beta}}$	0.01488	5.9534	0.00022	−0.00756	−3.0221	0.00006	0.00205	0.8208	0.00000
45	$\hat{\boldsymbol{\beta}}$	−0.01020	−4.0782	0.00010	0.00731	2.9228	0.00005	−0.25702	−102.8100	0.06606
	$\tilde{\boldsymbol{\beta}}$	0.00291	1.1633	0.00001	0.00389	1.5575	0.00002	−0.08515	−34.0600	0.00725
	$\breve{\boldsymbol{\beta}}$	0.00531	2.1250	0.00003	−0.00220	−0.8800	0.00000	0.02368	9.4729	0.00056
	$\hat{\boldsymbol{\beta}}$	−0.00664	−2.6571	0.00004	0.00337	1.3489	0.00001	−0.03711	−14.8450	0.00138
100	$\tilde{\boldsymbol{\beta}}$	−0.00032	−0.1270	0.00000	0.00093	0.3708	0.00000	0.00377	1.5099	0.00001
	$\breve{\boldsymbol{\beta}}$	0.00007	0.0287	0.00000	−0.00001	−0.0034	0.00000	0.02493	9.9719	0.00062

although the bootstrap correction in several cases underestimate the parameters. Among the three estimates, $\hat{\boldsymbol{\beta}}$ is the one with poorest performance. All estimates become more efficient if the sample size increases as it is expected.

Our simulation results indicate that the estimated biases of $\tilde{\boldsymbol{\beta}}$ and $\check{\boldsymbol{\beta}}$ are in absolute value smaller than the estimated biases of $\hat{\boldsymbol{\beta}}$ independent of n. The only exception corresponds to the estimated biases of $\check{\beta}_0$ with $n = 25$ which are equal to 0.02302 and 0.02582 for the GNB and GP models, respectively. These values are larger than the estimated biases 0.01346 and 0.01477 of $\hat{\beta}_0$ for these models. We can verify that the estimates $\tilde{\boldsymbol{\beta}}$ and $\check{\boldsymbol{\beta}}$ are more precise than $\hat{\boldsymbol{\beta}}$ based on their mean squared errors. Between the estimates $\tilde{\boldsymbol{\beta}}$ and $\check{\boldsymbol{\beta}}$, there is no indication that one estimate is superior to the other.

The performance of the estimates was evaluated by assuming that $\beta_0 = \beta_1 = 0.25$ for fixed $n = 35$ and varying β_2. Tables 4, 5 and 6 report results for the BNG, Consul and GP models, respectively. For all models, the estimates $\tilde{\boldsymbol{\beta}}$ are closer to the true parameters than the other estimates. For example, for the GP model and $\beta_2 = 0.5$, the estimated biases of $\tilde{\boldsymbol{\beta}}$ are equal in magnitude to $0.00029, 0.00487$ and 0.02325, whereas for $\hat{\boldsymbol{\beta}}$ they are $0.02128, 0.00967$ and 0.08210 and for $\check{\boldsymbol{\beta}}$ they are $0.00581, 0.00745$ and 0.05261. The estimate $\hat{\boldsymbol{\beta}}$ tends to underestimate β_0 and β_2 and to overestimate β_1, whereas the bias-corrected estimates tend to yield the opposite effects. In the majority of cases, $\tilde{\boldsymbol{\beta}}$ yields positive biases for all parameters. However, in almost all cases, $\check{\boldsymbol{\beta}}$ yields positive biases for β_0 and β_2 and negative biases for β_1.

We present some simulated intervals ACI, CACI and BtCI for the parameters β_0, β_1 and β_2 in the Consul, GNB and GP regression models. We evaluate 81 confidence intervals with nominal coverage 0.95 and sample sizes $n = 35, 45$ and 100. These intervals were constructed in such a way that they contained the true value of the parameter with probability 0.95, with probability 0.025 of the lower limit to be greater than the true value of the parameter, and with probability 0.025 of the upper limit to be smaller than the true value. For each confidence interval, the observed coverage probability was computed by the frequency that the 10,000 confidence intervals contain the true value of the parameter. Tables 7–9 give the simulated confidence intervals for $\beta_0 = 0.25$, $\beta_1 = 0.25$ and $\beta_2 = 1.0$ when $n = 35, 45$ and 100, respectively, under the above regression models.

We can observe that the average lengths of the ACI and CACI are close, although the average length of the CACI is slightly inferior for all sample sizes. Further, the BtCI has the longest average length than the corresponding two other intervals for all sample sizes. When the sample size increases, the average lengths of all intervals decrease as expected, that is, the interval estimates become more precise. Tables 7–9 indicate that the BtCI yields greatest coverage probabilities and then following by the ACI and CACI, respectively. When the sample size increases, the coverage probabilities tend to the nominal coverage.

Table 4 *Point estimation of β_2—GNB nonlinear model—$n = 35$*

		β_0			β_1			β_2		
β_2	Estimate	bias	RB (%)	MSE	bias	RB (%)	MSE	bias	RB (%)	MSE
0.25	$\hat{\boldsymbol{\beta}}$	−0.01428	−5.7134	0.00020	0.00891	3.5657	0.00008	−0.18050	−72.2020	0.03258
	$\tilde{\boldsymbol{\beta}}$	0.00413	1.6527	0.00002	0.00483	1.9303	0.00002	0.05699	22.7950	0.00325
	$\breve{\boldsymbol{\beta}}$	0.01267	5.0671	0.00016	−0.00714	−2.8548	0.00005	0.01945	7.7794	0.00038
0.5	$\hat{\boldsymbol{\beta}}$	−0.02102	−8.4071	0.00044	0.01004	4.0162	0.00010	−0.07023	−14.0450	0.00493
	$\tilde{\boldsymbol{\beta}}$	−0.00033	−0.1310	0.00000	0.00506	2.0246	0.00003	0.01622	3.2440	0.00026
	$\breve{\boldsymbol{\beta}}$	0.00474	1.8960	0.00002	−0.00772	−3.0890	0.00006	0.05825	11.6500	0.00339
0.75	$\hat{\boldsymbol{\beta}}$	−0.02324	−9.2954	0.00054	0.00970	3.8808	0.00009	−0.03610	−4.8128	0.00130
	$\tilde{\boldsymbol{\beta}}$	−0.00097	−0.3862	0.00000	0.00459	1.8369	0.00002	0.00700	0.9340	0.00005
	$\breve{\boldsymbol{\beta}}$	0.00481	1.9242	0.00002	−0.00807	−3.2281	0.00007	0.02531	3.3747	0.00064
1	$\hat{\boldsymbol{\beta}}$	−0.02421	−9.6828	0.00059	0.00945	3.7819	0.00009	−0.02691	−2.6913	0.00072
	$\tilde{\boldsymbol{\beta}}$	−0.00156	−0.6236	0.00000	0.00440	1.7601	0.00002	0.00495	0.4948	0.00002
	$\breve{\boldsymbol{\beta}}$	0.00522	2.0878	0.00003	−0.00772	−3.0881	0.00006	0.00556	0.5556	0.00003

Table 5 *Point estimation of β_2—Consul nonlinear model—$n = 35$*

		β_0			β_1			β_2		
β_2	Estimate	bias	RB (%)	MSE	bias	RB (%)	MSE	bias	RB (%)	MSE
0.25	$\hat{\boldsymbol{\beta}}$	−0.02424	−9.6954	0.00059	0.01209	4.8341	0.00015	−0.49742	−198.9700	0.24743
	$\tilde{\boldsymbol{\beta}}$	0.00847	3.3889	0.00007	0.00651	2.6025	0.00004	1.58820	635.2700	2.52230
	$\breve{\boldsymbol{\beta}}$	0.03081	12.3250	0.00095	−0.00953	−3.8137	0.00009	0.69055	276.2200	0.47686
0.5	$\hat{\boldsymbol{\beta}}$	−0.03138	−12.5510	0.00098	0.01367	5.4674	0.00019	−0.28810	−57.6190	0.08300
	$\tilde{\boldsymbol{\beta}}$	0.00251	1.0046	0.00001	0.00526	2.1042	0.00003	−0.02344	−4.6882	0.00055
	$\breve{\boldsymbol{\beta}}$	0.01308	5.2311	0.00017	−0.00957	−3.8289	0.00009	0.49339	98.6780	0.24343
0.75	$\hat{\boldsymbol{\beta}}$	−0.03604	−14.4140	0.00130	0.01455	5.8185	0.00021	−0.06748	−8.9966	0.00455
	$\tilde{\boldsymbol{\beta}}$	0.00030	0.1192	0.00000	0.00494	1.9777	0.00002	0.02725	3.6327	0.00074
	$\breve{\boldsymbol{\beta}}$	0.00719	2.8743	0.00005	−0.00961	−3.8437	0.00009	0.10189	13.5860	0.01038
1	$\hat{\boldsymbol{\beta}}$	−0.03749	−14.9980	0.00141	0.01488	5.9517	0.00022	−0.03500	−3.5004	0.00123
	$\tilde{\boldsymbol{\beta}}$	0.00026	0.1024	0.00000	0.00496	1.9853	0.00002	0.00401	0.4006	0.00002
	$\breve{\boldsymbol{\beta}}$	0.00708	2.8332	0.00005	−0.01029	−4.1159	0.00011	0.01781	1.7808	0.00032

Table 6 *Point estimation of β_2—GP nonlinear model—$n = 35$*

		β_0			β_1			β_2		
β_2	Estimate	bias	RB (%)	MSE	bias	RB (%)	MSE	bias	RB (%)	MSE
0.25	$\hat{\boldsymbol{\beta}}$	−0.01491	−5.9645	0.00022	0.00894	3.5778	0.00008	−0.18724	−74.8940	0.03506
	$\tilde{\boldsymbol{\beta}}$	0.00439	1.7553	0.00002	0.00513	2.0533	0.00003	0.06964	27.8570	0.00485
	$\breve{\boldsymbol{\beta}}$	0.01488	5.9534	0.00022	−0.00756	−3.0221	0.00006	0.00205	0.8208	0.00000
0.5	$\hat{\boldsymbol{\beta}}$	−0.02128	−8.5119	0.00045	0.00967	3.8681	0.00009	−0.08210	−16.4190	0.00674
	$\tilde{\boldsymbol{\beta}}$	0.00029	0.1153	0.00000	0.00487	1.9487	0.00002	0.02325	4.6502	0.00054
	$\breve{\boldsymbol{\beta}}$	0.00581	2.3258	0.00003	−0.00745	−2.9800	0.00006	0.05261	10.5220	0.00277
	$\hat{\boldsymbol{\beta}}$	−0.02421	−9.6819	0.00059	0.00916	3.6634	0.00008	−0.04057	−5.4091	0.00165
0.75	$\tilde{\boldsymbol{\beta}}$	−0.00082	−0.3265	0.00000	0.00424	1.6979	0.00002	0.00863	1.1511	0.00007
	$\breve{\boldsymbol{\beta}}$	0.00479	1.9146	0.00002	−0.00799	−3.1969	0.00006	0.03449	4.5987	0.00119
1	$\hat{\boldsymbol{\beta}}$	−0.02564	−10.2550	0.00066	0.00877	3.5066	0.00008	−0.03433	−3.4326	0.00118
	$\tilde{\boldsymbol{\beta}}$	−0.00176	−0.7053	0.00000	0.00396	1.5857	0.00002	0.01903	1.9033	0.00036
	$\breve{\boldsymbol{\beta}}$	0.00555	2.2211	0.00003	−0.00818	−3.2702	0.00007	0.02874	2.8744	0.00083

Table 7 *95% asymptotics CI, bootstrap-t CI and coverage probabilities for the indicated parameters in the Consul model*

		Lower limit			Upper limit			Coverage		
Estimate	Intervals	$n = 35$	$n = 45$	$n = 100$	$n = 35$	$n = 45$	$n = 100$	$n = 35$	$n = 45$	$n = 100$
β_0	ACI	−0.460	−0.378	−0.181	0.888	0.824	0.652	89.720	90.100	90.910
	CACI	−0.417	−0.346	−0.167	0.921	0.849	0.664	89.720	89.910	89.920
	BtCI	−0.519	−0.438	−0.233	1.094	1.001	0.761	94.350	94.760	94.540
β_1	ACI	−0.683	−0.591	−0.326	1.198	1.110	0.836	88.810	89.460	90.180
	CACI	−0.689	1.185	−0.598	1.098	−0.328	0.833	88.770	89.420	90.130
	BtCI	−0.886	−0.773	−0.439	1.409	1.293	1.281	94.610	94.830	94.950
β_2	ACI	0.565	0.621	0.759	1.360	1.317	1.218	89.380	89.950	90.150
	CACI	0.624	0.665	0.776	1.379	1.335	1.227	90.550	90.610	90.780
	BtCI	0.402	0.491	0.689	1.292	1.281	1.225	87.030	88.790	92.800

Table 8 *95% asymptotics CI, bootstrap-t CI and coverage probabilities for the indicated parameters in the GNB model*

		Lower limit			Upper limit			Coverage		
Estimate	Intervals	$n = 35$	$n = 45$	$n = 100$	$n = 35$	$n = 45$	$n = 100$	$n = 35$	$n = 45$	$n = 100$
β_0	ACI	−0.323	−0.253	−0.095	0.769	0.724	0.583	89.680	90.240	90.580
	CACI	−0.299	−0.235	−0.088	0.792	0.740	0.590	89.780	90.210	90.610
	BtCI	−0.373	−0.303	−0.139	0.924	0.858	0.667	94.070	94.240	94.260
β_1	ACI	−0.528	−0.460	−0.229	1.041	0.957	0.732	89.790	89.740	89.690
	CACI	−0.537	−0.468	−0.232	1.039	0.954	0.731	89.800	89.750	89.740
	BtCI	−0.679	−0.600	−0.319	1.220	1.110	0.833	94.630	94.630	93.320
β_2	ACI	0.625	0.671	0.784	1.316	1.283	1.195	90.210	90.180	90.600
	CACI	0.663	0.702	0.797	1.341	1.304	1.204	90.680	90.210	90.760
	BtCI	0.522	0.589	0.737	1.306	1.289	1.216	89.280	91.010	93.260

Table 9 *95% asymptotics CI, bootstrap-t CI and coverage probabilities for the indicated parameters in the GP model*

		Lower limit			Upper limit			Coverage		
Estimate	Intervals	$n = 35$	$n = 45$	$n = 100$	$n = 35$	$n = 45$	$n = 100$	$n = 35$	$n = 45$	$n = 100$
	ACI	−0.359	−0.292	−0.125	0.820	0.760	0.603	90.150	90.670	90.230
β_0	CACI	−0.332	−0.272	−0.116	0.836	0.772	0.610	89.940	90.580	90.200
	BtCI	−0.409	−0.348	−0.164	0.991	0.906	0.696	94.580	95.000	94.570
	ACI	−0.541	−0.465	−0.235	1.047	0.971	0.741	89.910	89.730	90.250
β_1	CACI	−0.548	−0.469	−0.237	1.044	0.967	0.739	89.860	89.700	89.740
	BtCI	−0.688	−0.599	−0.324	1.219	1.122	0.842	94.440	94.660	94.900
	ACI	−0.242	−0.099	0.150	1.068	0.977	0.813	89.710	89.450	90.360
β_2	CACI	−0.032	0.022	0.182	1.071	0.995	0.828	89.600	89.940	90.920
	BtCI	−1.561	−0.741	0.032	0.796	0.803	0.781	79.050	81.250	89.990

Table 10 *Calls to a technical support help line in the weeks immediately following a product release*

weeks	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8
calls	0	2	2	1	1	3	5	8	5	9	17	9	24	16	23	27

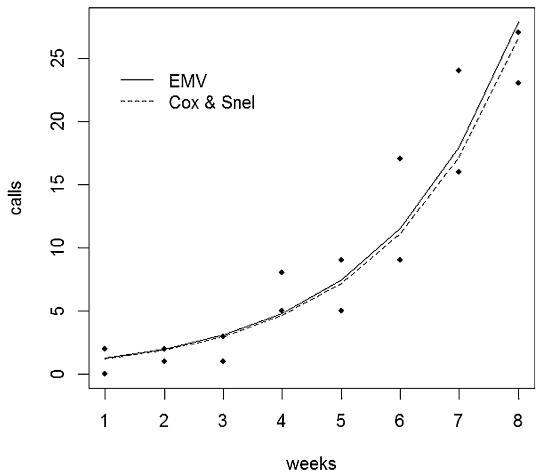

Figure 1 *Fitted number of calls under the Poisson model whose estimates are in Table* 11.

5 Application

In this section, we present an illustration the example 54.2 provided by SAS/ STAT(R) 9.2 User's Guide. This example shows how to analyze count data for calls to a technical support help line in the weeks immediately following a product release. This information could be used to decide upon the allocation of technical support resources for new products. You can model the number of daily calls as a Poisson random variable, where the average number of calls modeled by a nonlinear function of the number of weeks that have elapsed since the product's release. The data are given in Table 10. During the first several weeks after a new product is released, the number of questions that technical support receives concerning the product increases in a sigmoidal fashion. The expression for the mean value in the classic Poisson regression involves the identity link. The mean function is modeled as follows:

$$\mu_i = \beta_0 \exp(\beta_1 \text{weeks}_i), \qquad i = 1, \dots, 16. \tag{5.1}$$

The likelihood for every observation calls_i is $\text{calls}_\text{i} \sim \text{P}(\mu_\text{i})$.

In Figure 1, we plot the estimated number of calls against number of weeks using the uncorrected and corrected estimates.

Table 11 gives the uncorrected and corrected estimates and their asymptotic standard errors between parentheses. The corrected estimates in $\tilde{\boldsymbol{\beta}}$ are smaller than

Table 11 *Estimates of the model parameters (standard errors between parentheses)—Poisson model*

	$\hat{\boldsymbol{\beta}}$		$\tilde{\boldsymbol{\beta}}$	
Parameter	Estimate	Standard error	Estimate	Standard error
β_0	0.82610	0.25494	0.80406	0.25252
β_1	0.43983	0.046276	0.43760	0.047134

the uncorrected estimates in $\boldsymbol{\beta}$. The bias correction suggests that the uncorrected estimates underestimate $\boldsymbol{\beta}$.

6 Concluding remarks

Recently, there has been considerable interest to obtain explicit expressions for second-order biases of the maximum likelihood estimates (MLEs) in some regression models when they do not involve cumulants of log-likelihood derivatives. We discuss a general class of power series generalized nonlinear models (PSGNLMs) [Cordeiro, Andrade and De Castro (2009)], which is quite useful to analyze count data. We derive a simple matrix formula for the biases of the MLEs of the model parameters in these models. In addition, some Monte Carlo simulations have been investigated to compare the performance of the MLE $\hat{\boldsymbol{\beta}}$ and the two bias-corrected counterparts $\tilde{\boldsymbol{\beta}}$ and $\breve{\boldsymbol{\beta}}$ based on an analytical bias correction [Cox and Snell (1968)] and a bootstrap parametric technique, respectively.

The simulations show that the estimated biases of $\hat{\boldsymbol{\beta}}$ are much larger than those of the corresponding bias-corrected estimates. They also indicate that the bias-corrected estimates $\tilde{\boldsymbol{\beta}}$ and $\breve{\boldsymbol{\beta}}$ are closer to the true parameter values than the unadjusted estimates $\hat{\boldsymbol{\beta}}$, thus correctly signalizing the direction of the bias corrections. Overall, the estimate $\tilde{\boldsymbol{\beta}}$ is the best one in terms of bias size, since it usually yields estimated bias smaller in magnitude than those of $\hat{\boldsymbol{\beta}}$ and $\breve{\boldsymbol{\beta}}$. The bias correction has less impact as n increases and the corrected estimates $\tilde{\boldsymbol{\beta}}$ and $\breve{\boldsymbol{\beta}}$ tend to have slightly smaller standard errors than the uncorrected estimates $\hat{\boldsymbol{\beta}}$ at least for samples of moderate to large sizes. In these cases, the bias correction can lead to substantial improvement in terms of bias and mean square error. Overall, the simulations indicate that bias correction in PSGNLMs can then be used to obtain improved estimates with more reliable finite sample behavior.

Acknowledgments

Research supported by CAPES, CNPq and FACEPE, Brazil.

References

Botter, D. A. and Cordeiro, G. M. (1998). Improved estimators for generalized linear models with dispersion covariates. *Journal of Statistical Computation and Simulation* **62**, 91–104. MR1682556

Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*, 434 p. New York: Cambridge University Press. MR1648274

Consul, P. C. (1990). New class of location-parameter discrete probability distributions and their characterizations. *Communications in Statistics, Theory and Methods* **19**, 4653–4666. MR1114865

Cook, D. R., Tsai, C. L. and Wei, B. C. (1986). Bias in nonlinear regression. *Biometrika* **73**, 615–623. MR0897853

Cordeiro, G. M., Andrade, M. G. and De Castro, M. (2009). Power series generalized nonlinear models. *Computational Statistics and Data Analysis* **53**, 1155–1166. MR2657079

Cordeiro, G. M., Ferrari, S. L. P., Uribe-Opazo, M. A. and Vasconcellos, K. L. P. (2000). Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. *Statistics and Probability Letters* **46**, 317–328. MR1743990

Cordeiro, G. M. and McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B* **53**, 629–643. MR1125720

Cordeiro, G. M. and Paula, G. A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika* **76**, 93–100. MR0991426

Cordeiro, G. M. and Vasconcellos, K. L. P. (1997). Bias correction for a class of multivariate nonlinear regression models. *Statistics and Probability Letters* **35**, 155–164. MR1483269

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall. MR0370837

Cox, D. R. and Snell, E. (1968). A general definition of residuals. *Journal of the Royal Statistical Society B* **30**, 248–275. MR0237052

Doornik, J. A. (2009). *An Object-Oriented Matrix Programming Language Ox 6*. London: Timberlake Consultants.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* **7**, 1–26. MR0515681

Gupta, R. C. (1974). Modified power series distribution and some of its applications. *Sankhyâ B* **36**, 288–298. MR0391334

Lawley, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303. MR0082237

Ospina, R., Cribari–Neto, F. and Vasconcellos, K. L. P. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics and Data Analysis* **51**, 960–981. MR2297500

Vasconcellos, K. L. P. and Silva, S. G. (2005). Corrected estimates for student t regression models with unknown degrees of freedom. *Journal of Statistical Computation and Simulation* **75**, 409–423. MR2154788

P. G. Silva
A. H. M. A. Cysneiros
G. M. Cordeiro
Departamento de Estatística
Universidade Federal de Pernambuco
50740–540 Recife, PE
Brazil
E-mail: pgds1@de.ufpe.br
audrey@de.ufpe.br
gauss@de.ufpe.br
URL: https://www.ufpe.br/de/