On Bias Correction in a Class of Inflated Beta Regression Models

Raydonal Ospina[1] & Silvia L. P. Ferrari[2]

[1] Departamento de Estatística, Universidade Federal de Pernambuco, Brazil

[2] Departamento de Estatística, Universidade de São Paulo, Brazil

Correspondence: Silvia L. P. Ferrari, Departamento de Estatística, Universidade de São Paulo, Rua do Matão, 1010
São Paulo/SP 05508-090, Brazil. E-mail: silviaferrari.usp@gmail.com

Abstract

Inflated beta regression models bear practical applicability in modeling rates and proportions measured continuously in the presence of zeros and/or ones. In this article, the second-order bias of maximum likelihood estimators for zero-or-one inflated beta regression model parameters is derived. This enables one to obtain corrected estimators that are approximately unbiased. Numerical results exhibit that corrected estimators show better performance in terms of mean-square error and bias when compared to maximum likelihood estimators.

Keywords: inflated beta distribution, inflated beta regression model, bias, bias correction, maximum likelihood estimation

1. Introduction

Conventional estimation methods in statistical models may not be viable or appropriate in small samples. An important research area is the study of maximum likelihood estimator (MLE) behavior in small-sized samples, particularly bias analysis. In practice, the bias is usually ignored, the argument being that it is small when compared to the standard error of the parameter estimator. In fact, it can be shown that, in general, the bias is of order $O(n^{-1})$, where n is the sample size, while the asymptotic standard error is of order $O(n^{-1/2})$. However, for some models, the bias may be considerably greater than the corresponding standard error in small- or moderate-sized samples. Therefore, it is useful to obtain the second-order biases of MLEs, which allow us to evaluate the quality of the estimates and to obtain bias corrected estimates, particularly for small and moderate sample sizes.

Cox and Snell (1968) presented a general expression for the bias of order $O(n^{-1})$ of the MLE in both the one-parameter and multiparameter cases. Based on their formula, Cook et al. (1986) obtained the bias of the MLEs in nonlinear regression models with normal error. Cordeiro and McCullagh (1991), Botter and Cordeiro (1998), Cordeiro and Cribari–Neto (1998), and Cordeiro and Toyama (2008) obtained formulas for the second-order biases of the MLEs in generalized linear models and extensions. A detailed review on the bias correction literature can be found in Ospina et al. (2006).

The focus of this work is the analytic bias correction in inflated beta regression models. Beta regression models are suited for modeling data in the form of rates, fractions, and proportions; see, for instance, Paolino (2001), Kieschnick and McCullough (2003), Ferrari and Cribari–Neto (2004), Vasconcellos and Cribari–Neto (2005), Smithson and Verkuilen (2006), Cribari–Neto and Zeileis (2010), and Grün et al. (2012), among others. Among the different specifications of the beta regression models, the proposed structure of Ferrari and Cribari–Neto (2004) is the one presented here, as the parameterization they use allows for direct modeling of the distribution mean using a linear predictor and a general link function, in a similar way to what is done in generalized linear models.

In many practical situations, proportions and fractions present zeros and/or ones. In these cases, a statistical model that allows the addition of a point mass in the extremes of the interval $(0, 1)$ seems to be more appropriate than the beta law. Ospina and Ferrari (2010) developed inflated beta distributions as natural alternatives to the beta distributions for modeling data observed in $[0, 1)$, $(0, 1]$, or $[0, 1]$. In addition, Ospina and Ferrari (2012) studied beta regression models that allow for a point mass at one of the extremes of the unit interval and in which it is possible to model the parameters of the inflated beta distribution using adequate nonlinear predictors and link functions. In their work, inferential results are discussed focusing on likelihood theory. Also, diagnostic tools to identify violations of model assumptions are proposed.

Works focusing on bias correction of the MLEs for the parameters that index the beta distribution were conducted by Cordeiro et al. (1997) and Cribari–Neto and Vasconcellos (2002). Vasconcellos and Cribari–Neto (2005) obtained corrected bias MLEs for the parameters of the beta distribution, where such parameters are modeled through regression structures. Ospina et al. (2006) obtained a formula to calculate the biases of the MLEs in the beta regression model proposed by Ferrari and Cribari–Neto (2004). The authors noticed that, for small or moderate sample sizes, the MLE of the precision parameter can be markedly biased. Our objective is to extend the results obtained by Ospina et al. (2006) to the zero-or-one inflated beta regression models.

This article is organized as follows. In Section 2, the zero-or-one inflated beta regression models is defined and likelihood inference is presented. In Section 3, analytical expressions for the term of order $O(n^{-1})$ of the biases of the MLEs for the zero-or-one inflated beta regression model parameters are obtained. This term will be used to define corrected estimators that have bias of order $O(n^{-2})$. In Section 4, the corrected estimators are numerically evaluated using Monte Carlo simulations. Finally, our conclusions are presented in Section 5.

2. Model and Likelihood Inference

2.1 Model Definition and Likelihood Function

For simplicity, the linear version with constant precision of the zero-or-one inflated beta regression model proposed by Ospina and Ferrari (2012) is adopted. Consider independent random variables $y_1, \ldots, y_n$, each with an inflated beta density function at point c ($c = 0$ or $c = 1$)(Note 1) given by

$$\mathrm{bi}_c(y; \alpha, \mu, \phi) = \begin{cases} \alpha, & \text{if } y = c, \\ (1 - \alpha)f(y; \mu, \phi), & \text{if } y \in (0, 1), \end{cases} \tag{1}$$

with $0 < \alpha < 1$, $0 < \mu < 1$, $\phi > 0$, and $f(y; \mu, \phi)$ being the beta density function

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi-1}(1 - y)^{(1-\mu)\phi-1}, \quad y \in (0, 1). \tag{2}$$

Here,

$$\mathrm{E}(y^r) = \alpha c + (1 - \alpha)\mu_r,$$

$$\mathrm{Var}(y) = (1 - \alpha)\mathrm{Var}(y \mid y \in (0, 1)) + \alpha(1 - \alpha)(c - \mu)^2,$$

where $\mu_r = \mathrm{E}(y^r \mid y \in (0, 1)) = (\mu\phi)_{(r)}/(\phi)_{(r)}$, with $a_{(r)} = a(a + 1) \cdots (a + r - 1)$, is the rth moment of the beta distribution (2), and $\mathrm{Var}(y \mid y \in (0, 1)) = \mu(1 - \mu)/(1 + \phi)$. Note that, ϕ can be interpreted as a precision parameter and $\alpha = P(y = c)$ represents the probability of observing zero ($c = 0$) or one ($c = 1$). For more details about this distribution, refer to Ospina and Ferrari (2010).

The inflated beta linear regression model at point c is defined as follows. The independent random variables $y_1, \ldots, y_n$ are such y_t, for $t = 1, \ldots, n$, has density (1) with parameters $\alpha = \alpha_t$, $\mu = \mu_t$, and ϕ. It is assumed that α_t and μ_t are defined as

$$h(\alpha_t) = \sum_{i=1}^{M} z_{ti}\gamma_i = \zeta_t, \qquad g(\mu_t) = \sum_{i=1}^{m} x_{ti}\beta_i = \eta_t, \tag{3}$$

where $\gamma = (\gamma_1, \ldots, \gamma_M)^\top$ and $\beta = (\beta_1, \ldots, \beta_m)^\top$ are unknown regression parameter vectors, such that $\gamma \in \mathbf{R}^M$ and $\beta \in \mathbf{R}^m$, and $z_{t1}, \ldots, z_{tM}$ and $x_{t1}, \ldots, x_{tm}$ are observations of known exogenous variables, with $m + M < n$. Note that the values of z's and x's may fully or partially coincide. It is assumed that the link functions $h : (0, 1) \to \mathbf{R}$ and $g : (0, 1) \to \mathbf{R}$ are strictly monotone and twice differentiable. Here, ϕ is a precision parameter that is constant for all observations.

The likelihood function for the parameter vector $\theta = (\gamma^\top, \beta^\top, \phi)^\top$, of the zero-or-one inflated beta linear regression model is given by

$$L(\theta) = \prod_{t=1}^{n} \mathrm{bi}_c(y_t; \alpha_t, \mu_t, \phi) = L_1(\gamma)L_2(\beta, \phi), \tag{4}$$

with

$$L_1(\gamma) = \prod_{t=1}^{n} \alpha_t^{\mathbb{1}_{\{c\}}(y_t)}(1 - \alpha_t)^{1-\mathbb{1}_{\{c\}}(y_t)} \quad \text{and} \quad L_2(\beta, \phi) = \prod_{t:y_t \in (0,1)} f(y_t; \mu_t, \phi),$$

where $1\!\!1_A(y) = 1$ if $y \in A$, and $1\!\!1_A(y) = 0$ if $y \notin A$, and the parameters μ_t and α_t satisfy (3), that is, $\alpha_t = h^{-1}(\zeta_t)$ and $\mu_t = g^{-1}(\eta_t)$. Note that the likelihood function (4) can be factored into two terms, one that depends solely on the parameter vector γ and another that depends solely on β and ϕ. Therefore, the parameter vectors γ and $(\beta^\top, \phi)^\top$ are separable (Pace & Salvan, 1997, p. 128) and maximum likelihood inference on $(\beta^\top, \phi)^\top$ can be conducted separately from that for γ, as if the value of γ were known, and vice-versa. Further, note that the discrete component $L_1(\gamma)$ involves only the parameters used to model the probability of occurrence of zero or one. In contrast, the continuous component $L_2(\beta, \phi)$ only involves the parameters used to model the conditional distribution of the response variable, given that it belongs to the interval $(0, 1)$.

The log-likelihood function for θ is given by

$$\ell(\theta) = \ell_1(\gamma) + \ell_2(\beta, \phi) = \sum_{t=1}^{n} \ell_t(\alpha_t) + \sum_{t:y_t \in (0,1)} \ell_t(\mu_t, \phi), \tag{5}$$

where

$$\ell_t(\alpha_t) = 1\!\!1_{\{c\}}(y_t) \log \alpha_t + (1 - 1\!\!1_{\{c\}}(y_t)) \log(1 - \alpha_t),$$
$$\ell_t(\mu_t, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t)\phi) + (\mu_t \phi - 1) \log y_t + \{(1 - \mu_t)\phi - 1\} \log(1 - y_t).$$

By the separability of the parameter vectors γ and $(\beta^\top, \phi)^\top$, one can independently obtain the score for γ and the score for $(\beta^\top, \phi)^\top$.

2.2 Score Function

The components of the score vector, obtained by differentiation of the log-likelihood function with respect to the parameters, are given, for $R = 1, \ldots, M$ and $r = 1, \ldots, m$, as

$$U_R = \frac{\partial \ell_1(\gamma)}{\partial \gamma_R} = \sum_{t=1}^{n} \frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \frac{\partial \zeta_t}{\partial \gamma_R} = \sum_{t=1}^{n} \frac{1\!\!1_{\{c\}}(y_t) - \alpha_t}{\alpha_t(1 - \alpha_t)} \frac{d\alpha_t}{d\zeta_t} z_{tR}, \tag{6}$$

$$U_r = \frac{\partial \ell_2(\beta, \phi)}{\partial \beta_r} = \sum_{t:y_t \in (0,1)} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_r} = \phi \sum_{t=1}^{n} (1 - 1\!\!1_{\{c\}}(y_t))(y_t^* - \mu_t^*) \frac{d\mu_t}{d\eta_t} x_{tr},$$

where $d\alpha_t/d\zeta_t = dh^{-1}(\zeta_t)/d\zeta_t = 1/h'(\alpha_t)$, $d\mu_t/d\eta_t = dg^{-1}(\eta_t)/d\eta_t = 1/g'(\mu_t)$, and the conditional moment $\mu_t^* = E(y_t^* \mid y_t \in (0, 1)) = \psi(\mu_t \phi) - \psi((1 - \mu_t)\phi)$, with $\psi(\cdot)$ denoting the digamma function (Note 2) and $y_t^* = \log\{y_t/(1 - y_t)\}$ if $y_t \in (0, 1)$, and $y_t^* = 0$ otherwise. On the other hand,

$$U_\phi(\beta, \phi) = \frac{\partial \ell_2(\beta, \phi)}{\partial \phi} = \sum_{t:y_t \in (0,1)} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \phi} = \sum_{t=1}^{n} (1 - 1\!\!1_{\{c\}}(y_t))\{\mu_t(y_t^* - \mu_t^*) + s(y_t) + \psi(\phi) - \psi((1 - \mu_t)\phi)\},$$

where $s(y_t) = \log(1 - y_t)$ if $y_t \in (0, 1)$, and $s(y_t) = 0$ otherwise.

Defining the vectors $y^* = (y_1^*, \ldots, y_n^*)^\top$, $y^c = (1\!\!1_{\{c\}}(y_1), \ldots, 1\!\!1_{\{c\}}(y_n))^\top$, $\mu^* = (\mu_1^*, \ldots, \mu_n^*)^\top$, $\alpha^* = (\alpha_1, \ldots, \alpha_n)^\top$, and the diagonal matrices $H = \text{diag}\{1 - 1\!\!1_{\{c\}}(y_1), \ldots, 1 - 1\!\!1_{\{c\}}(y_n)\}$, $G = \text{diag}\{d\alpha_1/d\zeta_1, \ldots, d\alpha_n/d\zeta_n\}$, $P = \text{diag}\{1/[\alpha_1(1 - \alpha_1)], \ldots, 1/[\alpha_n(1 - \alpha_n)]\}$, and $T = \text{diag}\{d\mu_1/d\eta_1, \ldots, d\mu_n/d\eta_n\}$, the score vectors for γ and β can be written as

$$U_\gamma(\gamma) = Z^\top P G(y^c - \alpha^*),$$
$$U_\beta(\beta, \phi) = \phi X^\top T H(y^* - \mu^*),$$

respectively. Here, Z is an $n \times M$ matrix with the t-th row given by $z_t^\top = (z_{t1}, \ldots, z_{tM})$, and X is an $n \times m$ matrix with the t-th row given by $x_t^\top = (x_{t1}, \ldots, x_{tm})$. Finally, let $D^* = \text{diag}\{d_1^*, \ldots, d_n^*\}$ with $d_t^* = \mu_t(y_t^* - \mu_t^*) + s(y_t) + \psi(\phi) - \psi((1 - \mu_t)\phi)$. The score function for the precision parameter is given by $U_\phi = \text{tr}(H D^*)$, where $\text{tr}(\cdot)$ represents the trace of a square matrix.

From the separability of the parameters γ and $(\beta^\top, \phi)^\top$, the MLE of γ is obtained independently from that of $(\beta^\top, \phi)^\top$ as the solution to the nonlinear system $U_\gamma(\gamma) = 0$. Likewise, the MLE of $(\beta^\top, \phi)^\top$ is obtained as the solution of the nonlinear system $(U_\beta(\beta, \phi)^\top, U_\phi(\beta, \phi))^\top = 0$. Note that, in both cases, such estimators do not have closed form. They can be obtained numerically by maximizing the log-likelihood function using a nonlinear

optimization algorithm, such as the Newton algorithm (Newton-Raphson, Fisher score, BHHH, etc.) or a quasi-Newton algorithm (BFGS); refer to Press et al. (1992, Chapters 9 and 10).

2.3 Fisher's Information Matrix

Let $Q = GPG = \text{diag}\{q_1, \ldots, q_n\}$, $\Delta = \text{diag}\{\delta_1, \ldots, \delta_n\}$, $W = \text{diag}\{w_1, \ldots, w_n\}$, $D = \text{diag}\{d_1, \ldots, d_n\}$ and $\mathbf{c} = (\mathbf{c}_1, \ldots, \mathbf{c}_n)^\top$, where $q_t = p_t(\mathrm{d}\alpha_t/\mathrm{d}\zeta_t)^2$, $\delta_t = 1 - \alpha_t$, $w_t = \psi'(\mu_t\phi) + \psi'(1 - \mu_t)\phi)$, $d_t = (1 - \mu_t)^2\psi''((1 - \mu_t)\phi) + \mu_t^2\psi'(\mu_t\phi) - \psi'(\phi)$ and $\mathbf{c}_t = \phi[\mu_t\psi'(\mu_t\phi) - (1 - \mu_t)\psi'((1 - \mu_t)\phi)]$, for $t = 1, \ldots, n$. The Fisher information matrix $K(\theta)$ for the zero-or-one inflated beta regression model is given by

$$K(\theta) = \begin{pmatrix} K_{\gamma\gamma} & 0 \\ 0 & K_\vartheta(\vartheta) \end{pmatrix} = \begin{pmatrix} K_{\gamma\gamma} & 0 & 0 \\ 0 & K_{\beta\beta} & K_{\beta\phi} \\ 0 & K_{\phi\beta} & K_{\phi\phi} \end{pmatrix}, \tag{7}$$

where $K_{\gamma\gamma} = Z^\top QZ$, $K_{\beta\beta} = \phi^2 X^\top \Delta TWTX$, $K_{\beta\phi} = K_{\beta\phi}^\top = X^\top \Delta T\mathbf{c}$, $K_{\phi\phi} = \text{tr}(\Delta D)$. Note that the matrix in (7) does not depend on the inflation point c. Also, note that $K_{\gamma\gamma}$ does not depend on (β, ϕ) and that $K_{\beta\beta}$, $K_{\beta\phi}$, $K_{\beta\phi}$ and $K_{\phi\phi}$ do not depend on γ. Additionally, γ is orthogonal to $\vartheta = (\beta^\top, \phi)^\top$, which implies that the respective elements of the score vector are asymptotically uncorrelated. Consequently, the MLE of γ is asymptotically independent of the MLEs of β and ϕ.

Let $W_{\beta\beta} = \phi^2 \Delta TWT$, $W_{\beta\phi} = \Delta T\mathbf{c}$, $W_{\phi\beta} = W_{\beta\phi}^\top$ and $W_{\phi\phi} = \text{tr}(\Delta D)$. Also, let

$$\widetilde{X} = \begin{pmatrix} X & 0 \\ 0 & 1 \end{pmatrix} \tag{8}$$

and

$$\widetilde{W} = \begin{pmatrix} W_{\beta\beta} & W_{\beta\phi} \\ W_{\phi\beta} & W_{\phi\phi} \end{pmatrix}. \tag{9}$$

The Fisher information matrix for $\vartheta = (\beta^\top, \phi)^\top$ can be written as $K_\vartheta(\vartheta) = \widetilde{X}^\top \widetilde{W}\widetilde{X}$.

From the standard formula for the inverse of partitioned matrix (see, for instance, Rao, 1973, p. 33), it can deduced that the Fisher information matrix inverse is given by

$$K(\theta)^{-1} = \begin{pmatrix} K_{\gamma\gamma}^{-1} & 0 \\ 0 & K_\vartheta(\vartheta)^{-1} \end{pmatrix} = \begin{pmatrix} K^{\gamma\gamma} & 0 & 0 \\ 0 & K^{\beta\beta} & K^{\beta\phi} \\ 0 & K^{\phi\beta} & K^{\phi\phi} \end{pmatrix}, \tag{10}$$

where

$$K^{\gamma\gamma} = (Z^\top QZ)^{-1},$$

$$K^{\beta\beta} = (X^\top W_{\beta\beta}X)^{-1}\left\{I_m + \frac{X^\top T\mathbf{c}\mathbf{c}^\top T^\top X(X^\top W_{\beta\beta}X)^{-1}}{\text{tr}(D) - \mathbf{c}^\top T^\top X(X^\top W_{\beta\beta}X)^{-1}X^\top T\mathbf{c}}\right\},$$

$$K^{\beta\phi} = (K^{\phi\beta})^\top = -[\text{tr}(D) - \mathbf{c}^\top T^\top X(X^\top W_{\beta\beta}X)^{-1}X^\top T\mathbf{c}]^{-1}(X^\top W_{\beta\beta}X)^{-1}X^\top T\mathbf{c},$$

$$K^{\phi\phi} = [\text{tr}(D) - \mathbf{c}^\top T^\top X(X^\top W_{\beta\beta}X)^{-1}X^\top T\mathbf{c}]^{-1},$$

with I_m being the $m \times m$ identity matrix. The inverse of Fisher's information matrix is useful for computing asymptotic standard errors of MLEs. For more details see Ospina and Ferrari (2012).

3. Bias Correction of MLEs

3.1 Cox and Snell's Formula

Let $\mathcal{L}(\omega)$ be the log-likelihood function of a parameter vector $\omega = (\omega_1, \ldots, \omega_k)$ for a sample of n observations, and let $\widehat{\omega} = (\widehat{\omega}_1, \ldots, \widehat{\omega}_k)$ be the MLE of ω obtained as the solution of the system of equations $U(\omega) = \partial\mathcal{L}(\omega)/\partial\omega = 0$. Cox AND Snell (1968) obtained the $O(n^{-1})$ term of the bias of $\widehat{\omega}_r$ as

$$B(\widehat{\omega}_r) = \sum_{s,t,u} \kappa^{rs}\kappa^{tu}\left(\kappa_{st,u} + \frac{1}{2}\kappa_{stu}\right), \qquad r = 1, \ldots, k, \tag{11}$$

where $-\kappa^{rs} = \kappa^{r,s}$ is the (r, s) element of the Fisher information matrix inverse, $\kappa_{stu} = \text{E}(\partial^3\mathcal{L}(\omega)/\partial\omega_s\partial\omega_t\partial\omega_u)$, and $\kappa_{st,u} = \text{E}(\partial^2\mathcal{L}(\omega)/\partial\omega_s\partial\omega_t \times \partial\mathcal{L}(\omega)/\partial\omega_u)$. In many situations, it is convenient to use the Bartlett identity

$\kappa_{st,u} + \frac{1}{2}\kappa_{stu} = \kappa_{st}^{(u)} - \frac{1}{2}\kappa_{stu}$, , where $\kappa_{st}^{(u)} = \partial\kappa_{st}/\partial\omega_u$, to facilitate expressing the bias in matrix notation. The great usefulness of (11) is in defining a corrected MLE, which is unbiased up to order $O(n^{-1})$, given by

$$\widetilde{\omega}_r = \widehat{\omega}_r - \widehat{B}(\widehat{\omega}_r),$$

where $\widehat{B}(\widehat{\omega}_r)$ is given by (11) evaluated at $\widehat{\omega}$. This new estimator, $\widetilde{\omega}_r$, has bias of order $O(n^{-2})$, as $E(\widetilde{\omega}_r) = \omega_r + O(n^{-2})$, and may be preferred over $\widehat{\omega}_r$ that has bias of order $O(n^{-1})$.

Our aim is to obtain an expression to calculate the second-order bias of the vector of the MLEs for zero-or-one inflated beta regression models using formula (11).

3.2 Bias of $\widehat{\gamma}$ and $(\widehat{\beta}^\top, \widehat{\phi})^\top$

Now, we obtain an expression for the second order biases of the MLEs in a class of inflated beta regression models using the Cox and Snell's (1968) general formula. This expression will, in turn, allow us to obtain bias corrected estimates of the unknown parameters. The notation to be used is introduced below. The derivatives of the log-likelihood function (5) with respect to the unknown parameters are indicated by indices, where the letters $R, S, \ldots$ correspond to the derivatives with respect to the elements of γ, the letters $r, s, \ldots$ correspond to the derivatives with respect to the elements β, and ϕ corresponds to the derivatives with respect to ϕ. For instance, $U_R = \partial\ell/\partial\gamma_R$, $U_r = \partial\ell/\partial\beta_r$, $U_\phi = \partial\ell/\partial\phi$, $U_{\phi s} = \partial^2\ell/\partial\phi\partial\beta_s$, $U_{rs\phi} = \partial^3\ell/\partial\beta_r\partial\beta_s\partial\phi$, and so forth. The notation for the cumulants of log-likelihood derivatives is borrowed from Lawley (1956): $\kappa_{RS} = E(U_{RS})$, $\kappa_{rs} = E(U_{rs})$, $\kappa_{\phi\phi} = E(U_\phi U_\phi)$, $\kappa_{rs,\phi} = E(U_{rs}U_\phi)$, $\kappa_{RSU} = E(U_{RSU})$, $\kappa_{rst} = E(U_{rst})$, etc., where all κs are of order $O(n)$. The derivatives of cumulants are denoted by $\kappa_{RS}^{(U)} = \partial\kappa_{RS}/\partial\gamma_U$, $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, $\kappa_{r\phi}^{(\phi)} = \partial\kappa_{r\phi}/\partial\phi$, and so on.

In the zero-or-one inflated beta regression model, the separability between γ and $\vartheta = (\beta^\top, \phi)^\top$ enables us to independently obtain the second-order bias of $\widehat{\gamma}$ and $(\widehat{\beta}^\top, \widehat{\phi})^\top$. From (11) and the aforementioned Bartlett identity, the second-order bias of the b-th element of $\widehat{\gamma} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_M)$ reduces to

$$B(\widehat{\gamma}_b) = \sum_{R,S,U} \kappa^{bR}\kappa^{SU}\left\{\kappa_{RS}^{(U)} - \frac{1}{2}\kappa_{RSU}\right\}. \tag{12}$$

The terms in (12) are obtained from the cumulants of the log-likelihood function given in the Appendix.

Let $\widetilde{\mathcal{Z}} = Q^{-1}W_0\delta_{\gamma\gamma}$, where $W_0 = \text{diag}\{w_{01}, \ldots, w_{0n}\}$ is an $n \times n$ diagonal matrix with t-th diagonal element given by

$$w_{0t} = \frac{1}{2}p_t\left(\frac{\partial}{\partial\zeta_t}\frac{d\alpha_t}{d\zeta_t}\right)\left(\frac{d\alpha_t}{d\zeta_t}\right)^2,$$

and $\delta_{\gamma\gamma}$ represents the $n \times 1$ vector obtained from the main diagonal of $ZK^{\gamma\gamma}Z^\top = Z(Z^\top QZ)^{-1}Z^\top$. We obtain

$$B(\widehat{\gamma}) = (Z^\top QZ)^{-1}Z^\top Q\widetilde{\mathcal{Z}}, \tag{13}$$

where Z is the covariate matrix used to model the discrete component of the model.

Now, from (11), the second-order bias of the a-th element of $\widehat{\vartheta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_m, \widehat{\phi})$ can be written as

$$
\begin{aligned}
B(\widehat{\vartheta}_a) = &\sum_{r,s,u}\kappa^{ar}\kappa^{su}\left\{\kappa_{rs}^{(u)} - \frac{1}{2}\kappa_{rsu}\right\} + \kappa^{a\phi}\sum_{s,u}\kappa^{su}\left\{\kappa_{\phi s}^{(u)} - \frac{1}{2}\kappa_{\phi su}\right\}\\
&+ \sum_{r,u}\kappa^{ar}\kappa^{\phi u}\left\{\kappa_{r\phi}^{(u)} - \frac{1}{2}\kappa_{r\phi u}\right\} + \sum_{r,s}\kappa^{ar}\kappa^{s\phi}\left\{\kappa_{rs}^{(\phi)} - \frac{1}{2}\kappa_{rs\phi}\right\}\\
&+ \kappa^{a\phi}\sum_{u}\kappa^{\phi u}\left\{\kappa_{\phi\phi}^{(u)} - \frac{1}{2}\kappa_{\phi\phi u}\right\} + \kappa^{a\phi}\sum_{s}\kappa^{s\phi}\left\{\kappa_{\phi s}^{(\phi)} - \frac{1}{2}\kappa_{\phi s\phi}\right\}\\
&+ \kappa^{\phi\phi}\sum_{r}\kappa^{ar}\left\{\kappa_{r\phi}^{(\phi)} - \frac{1}{2}\kappa_{r\phi\phi}\right\} + \kappa^{a\phi}\kappa^{\phi\phi}\left\{\kappa_{\phi\phi}^{(\phi)} - \frac{1}{2}\kappa_{\phi\phi\phi}\right\},
\end{aligned}
\tag{14}
$$

for $a = 1, \ldots, m+1$. Note that β and ϕ are not orthogonal parameters, as the block $K_{\beta\phi}$ in the information matrix is not a null matrix. It forces us to calculate all the terms in $B(\widehat{\vartheta}_a)$. The cumulants of $\ell_2(\beta, \phi)$ that appear in (14) are given in the Appendix. After some calculations we arrive at

$$
\begin{aligned}
B(\widehat{\beta}) = &K^{\beta\beta}X^\top[\Delta W_1\delta_{\beta\beta} + \Delta(W_3 + W_2)XK^{\beta\phi} + (\text{diagonal}(\Delta W_4))^\top K^{\phi\phi}]\\
&+ K^{\beta\phi}[\text{tr}(\Delta W_3 XK^{\beta\beta}X^\top) + K^{\phi\phi}\text{tr}(\Delta S) + \text{diagonal}(\Delta(W_5 + W_4))XK^{\beta\phi}],
\end{aligned}
\tag{15}
$$

where diagonal($\cdot$) represents the row vector formed by the main diagonal of a square matrix, $K^{\beta\beta}$, $K^{\phi\phi}$ and $K^{\beta\phi}$ are blocks of the Fisher information matrix inverse, and W_i, for $i = 1, \ldots, 5$ are given in the Appendix.

Now, let

$$\widetilde{\delta} = \begin{pmatrix} \Delta W_1 \delta_{\beta\beta} + \Delta(W_3 + W_2)XK^{\beta\phi} + (\text{diagonal}(\Delta W_4))^\top K^{\phi\phi} \\ \text{tr}(\Delta W_3 XK^{\beta\beta}X^\top) + K^{\phi\phi}\text{tr}(\Delta S) + \text{diagonal}(\Delta(W_5 + W_4))XK^{\beta\phi} \end{pmatrix},$$

and $K^{\beta*} = (K^{\beta\beta} \ K^{\beta\phi})$.

The second-order bias of $\widehat{\beta}$ can be written as

$$B(\widehat{\beta}) = K^{\beta*}\widetilde{X}^\top \widetilde{\delta},$$

where $\widetilde{X}$ is given in (8).

As earlier, after some algebraic manipulations, we have that

$$
\begin{aligned}
B(\widehat{\phi}) &= K^{\phi\beta}X^\top(\Delta W_1 \delta_{\beta\beta} + \Delta(W_3 + W_2)XK^{\beta\phi} + (\text{diagonal}(\Delta W_4))^\top K^{\phi\phi}) \\
&\quad + K^{\phi\phi}(\text{tr}(\Delta W_3 XK^{\beta\beta}X^\top) + K^{\phi\phi}\text{tr}(\Delta S) + \text{diagonal}(\Delta(W_5 + W_4))XK^{\beta\phi}).
\end{aligned}
\tag{16}
$$

Let $K^{\phi*} = (K^{\phi\beta} \ K^{\phi\phi})$ be a $1 \times (m + 1)$ matrix. The second-order bias of $\widehat{\phi}$ can be written as

$$B(\widehat{\phi}) = K^{\phi*}\widetilde{X}^\top \widetilde{\delta}.$$

Let $\vartheta = (\beta^\top, \phi)^\top$. The second-order bias of the MLE of ϑ can be written as

$$B(\widehat{\vartheta}) = (\widetilde{X}^\top \widetilde{W} \widetilde{X})^{-1}\widetilde{X}^\top \widetilde{W} \widetilde{\xi},$$

where $\widetilde{\xi} = \widetilde{W}^{-1}\widetilde{\delta}$ with $\widetilde{W}$ given in (9). Consequently, $B(\widehat{\vartheta})$ can be estimated from a generalized least squares regression in the auxiliary variable $\widetilde{\xi}$. Note that the expression for the correction of the second-order bias of $\widehat{\vartheta}$ involves the parameters of the discrete and continuous components of the model.

3.3 Bias of $\widehat{\mu}$ and $\widehat{\alpha}$

Let $\zeta = (\zeta_1, \ldots, \zeta_n)^\top = Z\gamma$ and $\eta = (\eta_1, \ldots, \eta_n)^\top = X\beta$ be the linear predictors. The MLEs of ζ and η are $\widehat{\zeta} = Z\widehat{\gamma}$ and $\widehat{\eta} = X\widehat{\beta}$, respectively. Hence, $\widehat{\zeta} - \zeta = Z(\widehat{\gamma} - \gamma)$ and $\widehat{\eta} - \eta = X(\widehat{\beta} - \beta)$. Consequently, the second-order bias of $\widehat{\zeta}$ is

$$B(\widehat{\zeta}) = ZB(\widehat{\gamma}) = Z(Z^\top QZ)^{-1}Z^\top Q\widetilde{Z}, \tag{17}$$

where $B(\widehat{\gamma})$ is the second-order bias of $\widehat{\gamma}$ given in (13). Similarly, the second-order bias of $\widehat{\eta}$ is

$$B(\widehat{\eta}) = XB(\widehat{\beta}) = XK^{\beta*}\widetilde{X}^\top \widetilde{\delta}. \tag{18}$$

Now, let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top$, where $\alpha_t = h^{-1}(\zeta_t)$ and $\mu_t = g^{-1}(\eta_t)$, for $t = 1, \ldots, n$, with g and h being the corresponding link functions. To obtain the second-order bias of $\widehat{\alpha}_t$, $h^{-1}(\widehat{\zeta}_t)$ is expanded in Taylor series up to the second order in a neighbourhood of ζ_t as

$$h^{-1}(\widehat{\zeta}_t) = h^{-1}(\zeta_t) + \frac{dh^{-1}(\zeta_t)}{d\zeta_t}(\widehat{\zeta}_t - \zeta_t) + \frac{1}{2}\frac{d^2 h^{-1}(\zeta_t)}{d\zeta_t^2}(\widehat{\zeta}_t - \zeta_t)^2 + O_p(n^{-3/2}).$$

By arranging the terms and taking the expected value, the second-order bias of $\widehat{\alpha}_t$ is given by

$$B(\widehat{\alpha}_t) = B(h^{-1}(\widehat{\zeta}_t)) = B(\widehat{\zeta}_t)\frac{d\alpha_t}{d\zeta_t} + \frac{1}{2}\text{V}(\widehat{\zeta}_t)\frac{d^2\alpha_t}{d\zeta_t^2},$$

where $\text{V}(\widehat{\zeta}_t)$ is the asymptotic variance of $\widehat{\zeta}_t$, obtained from the (t, t)-th element of the matrix $Z(Z^\top QZ)^{-1}Z^\top$ and $B(\widehat{\zeta}_t)$ is the $O(n^{-1})$ bias of $\widehat{\zeta}_t$. Consequently, we can write

$$B(\widehat{\alpha}) = G_1 Z(Z^\top QZ)^{-1}Z^\top Q\widetilde{Z} + \frac{1}{2}G_2 \text{diagonal}(Z(Z^\top QZ)^{-1}Z^\top),$$

where $G_1 = \text{diag}(d\alpha_1/d\zeta_1, \ldots, d\alpha_n/d\zeta_n)$, $G_2 = \text{diag}(d^2\alpha_1/d\zeta_1^2, \ldots, d^2\alpha_n/d\zeta_n^2)$, and $B(\widehat{\zeta})$ is the n^{-1} bias of $\widehat{\zeta}$ given in (17).

Similarly, the second-order bias of $\widehat{\mu}$ can be obtained. By expanding $g^{-1}(\widehat{\eta}_t)$ in Taylor series up to the second order in a neighbourhood of η_t, we have

$$g^{-1}(\widehat{\eta}_t) = g^{-1}(\eta_t) + \frac{\mathrm{d}g^{-1}(\eta_t)}{\mathrm{d}\eta_t}(\widehat{\eta}_t - \eta_t) + \frac{1}{2}\frac{\mathrm{d}^2 g^{-1}(\eta_t)}{\mathrm{d}^2 \eta_t}(\widehat{\eta}_t - \eta_t)^2 + O_p(n^{-3/2}).$$

By rearranging the terms and taking the expected value, the second-order bias of $\widehat{\mu}_t$ is given by

$$B(\widehat{\mu}_t) = B(g^{-1}(\widehat{\eta}_t)) = B(\widehat{\eta}_t)\frac{\mathrm{d}\mu_t}{\mathrm{d}\eta_t} + \frac{1}{2}\mathrm{V}(\widehat{\eta}_t)\frac{\mathrm{d}^2 \mu_t}{\mathrm{d}\eta_t^2},$$

where $\mathrm{V}(\widehat{\eta}_t)$ is the asymptotic variance of $\widehat{\eta}_t$ obtained from the (t,t)-th element of $XK^{\beta\beta}X^\top$, and $B(\widehat{\eta}_t)$ being the $O(n^{-1})$ bias of $\widehat{\eta}_t$ given in (18). We can now write

$$B(\widehat{\boldsymbol{\mu}}) = TXK^{\beta*}\widetilde{X}^\top\widetilde{\delta} + \frac{1}{2}T_1\mathrm{diagonal}(XK^{\beta\beta}X^\top),$$

where $T = \mathrm{diag}(\mathrm{d}\mu_1/\mathrm{d}\eta_1, \ldots, \mathrm{d}\mu_n/\mathrm{d}\eta_n)$, $T_1 = \mathrm{diag}(\mathrm{d}^2\mu_1/\mathrm{d}\eta_1^2, \ldots, \mathrm{d}^2\mu_n/\mathrm{d}\eta_n^2)$.

3.4 Bias of Smooth Functions of $\widehat{\phi}$

For the zero-or-one inflated beta regression model, a reparameterization of the precision parameter, ϕ, can be considered. Let $\sigma = \mathcal{T}(\phi)$, where $\mathcal{T}(\cdot)$ is a continuous strictly monotone twice differentiable function. From the invariance property of the MLEs, we have $\widehat{\sigma} = \mathcal{T}(\widehat{\phi})$. From a Taylor series expansion of $\mathcal{T}(\phi)$ up to the third term in a neighbourhood of ϕ, we get

$$\mathcal{T}(\widehat{\phi}) = \mathcal{T}(\phi) + \mathcal{T}'(\phi)(\widehat{\phi} - \phi) + \frac{1}{2}\mathcal{T}''(\phi)(\widehat{\phi} - \phi)^2 + O_p(n^{-3/2}).$$

By taking the expected value, the second-order bias of $\widehat{\sigma}$ can be written as

$$B(\widehat{\sigma}) = \mathcal{T}'(\phi)B(\widehat{\phi}) + \frac{1}{2}\mathcal{T}''(\phi)\mathrm{V}(\widehat{\phi}),$$

where $B(\widehat{\phi})$ is the second-order bias of $\widehat{\phi}$ given in (16), and $\mathrm{V}(\widehat{\phi}) = K^{\phi\phi}$ is the asymptotic variance of $\widehat{\phi}$. In Table 1, different specifications for σ are presented along with the second-order bias of the respective MLE.

Table 1. Bias of $\widehat{\sigma} = \mathcal{T}(\widehat{\phi})$

$\sigma = \mathcal{T}(\phi)$	$B(\widehat{\sigma})$
$1/(\phi+1)$	$-\frac{B(\widehat{\phi})}{(\phi+1)^2} + \frac{\mathrm{V}(\widehat{\phi})}{(\phi+1)^3}$
$1/(\phi+1)^2$	$-\frac{2B(\widehat{\phi})}{(\phi+1)^3} + \frac{3\mathrm{V}(\widehat{\phi})}{(\phi+1)^4}$
$\log\phi$	$\frac{B(\widehat{\phi})}{\phi} - \frac{\mathrm{V}(\widehat{\phi})}{2\phi^2}$

Through simulated data, the behavior of the relative bias of $B(\widehat{\sigma})/\sigma$ as a function of ϕ is illustrated. We consider a zero-inflated beta regression model with

$$g(\mu_t) = \beta_0 + \beta_1 x_t, \qquad h(\alpha_t) = \gamma_0 + \gamma_1 z_t,$$

$t = 1, \ldots, n$, where g and h are logit link functions. The values of the covariates x_t and z_t are independent realizations of a uniform random variable, $\mathcal{U}(0,1)$. The sample size is $n = 100$ and the true parameter values are $\gamma_0 = -0.5$, $\gamma_1 = 0.5$, $\beta_0 = -0.5$, and $\beta_1 = 1.5$. Here, ϕ takes the values $2, 5, 10, 20, 40, 60, 100, 150, 500$, and 1000.

In Figure 1 the behavior of relative bias of $B(\widehat{\sigma})/\sigma$ as a function of ϕ is presented. Note that the relative bias of the MLEs corresponding to the parameterizations $\sigma = \phi$ and $\sigma = 1/(\phi+1)^2$ goes to a non-zero constant value as ϕ increases. Also, note that for the parameterization $\sigma = 1/(\phi+1)$, the relative bias behaves approximately as a constant function at zero and this is valid even for moderate sample sizes (for instance, $n = 80$). Finally, for the parameterization $\sigma = \log\phi$, it is observed that the relative bias decreases to zero as ϕ grows. In light of these facts, the use of the parameterization $\sigma = 1/(\phi+1)$ can be suggested because even with moderate values of n and small values of ϕ, the relative bias remains stable and close to zero.
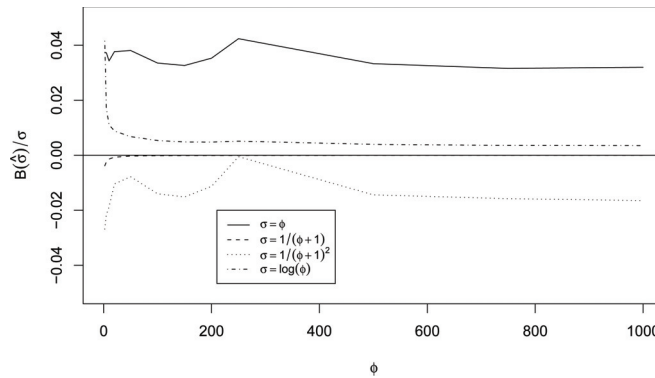
Figure 1. Second-order relative bias of $\widehat{\sigma}$

4. Numerical Evaluation

Through Monte Carlo simulations, we now investigate the performances of MLEs for the inflated beta regression model parameters and their corrected versions in finite-sized samples. First, we consider a zero-inflated beta regression model with

$$h(\alpha_t) = \gamma_0 + \gamma_1 z_t, \qquad g(\mu_t) = \beta_0 + \beta_1 x_t, \tag{19}$$

$t = 1, \ldots, n$. Here, h and g are logit link functions. The true parameter values are $\gamma_0 = -0.5$, $\gamma_1 = 1.5$, $\beta_0 = 0.5$, $\beta_1 = 1.8$, and $\phi = 120$; the covariate values, z_t and x_t, are independently selected from the $\mathcal{U}(0, 1)$ distribution. For this experiment, the sample sizes are $n = 30, 60$, and 90, and 5000 Monte Carlo replications. For each simulated sample, we fitted the regression model (19); that is, by maximizing the log-likelihood function, we obtained the estimates $\widehat{\gamma} = (\widehat{\gamma_0}, \widehat{\gamma_1})$, $\widehat{\vartheta} = (\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\phi})$, and $\widehat{\sigma} = 1/(\widehat{\phi} + 1)$. In addition, from the results in Section 3, we calculated the corrected MLEs, $\widetilde{\gamma} = (\widetilde{\gamma_0}, \widetilde{\gamma_1})$, $\widetilde{\vartheta} = (\widetilde{\beta_0}, \widetilde{\beta_1}, \widetilde{\phi})$, and $\widetilde{\sigma}$.

Table 2. Simulation results for estimation of γ_0 and γ_1

n	Estimator	Bias	Rel. bias	$\sqrt{\text{MSE}}$
30	$\widehat{\gamma_0}$	−0.0617	0.1235	0.7496
	$\widetilde{\gamma_0}$	−0.0514	0.1028	0.7376
60	$\widehat{\gamma_0}$	−0.0340	0.0680	0.5607
	$\widetilde{\gamma_0}$	−0.0285	0.0571	0.5556
90	$\widehat{\gamma_0}$	−0.0113	0.0227	0.4003
	$\widetilde{\gamma_0}$	−0.0079	0.0158	0.3980
30	$\widehat{\gamma_1}$	0.1786	0.1191	1.6488
	$\widetilde{\gamma_1}$	0.1472	0.0981	1.6193
60	$\widehat{\gamma_1}$	0.0878	0.0585	0.9763
	$\widetilde{\gamma_1}$	0.0732	0.0488	0.9671
90	$\widehat{\gamma_1}$	0.0394	0.0263	0.7442
	$\widetilde{\gamma_1}$	0.0295	0.0197	0.7397

In the Monte Carlo experiment, we used the "multiply-with-carry" algorithm (GM) as a pseudorandom number generator, with period 2^{60}. The log-likelihood function is maximized through the BFGS method with analytical derivatives, which, in general, is the method that presents the best performance (Mittelhammer, Judge, & Miller, 2000, p.199). All the simulations were programmed using the Ox matrix programming language `Ox` (Cribari–Neto & Zarkos, 2003). For each sample size, the bias, relative bias, and root-mean-square error of the 5000 estimates were calculated. Tables 2 and 3 present the simulation results.

Notice that the bias, relative bias, and mean square error decrease as the sample size increases, as expected. In Table 2, we observe that the bias of the MLEs of the regression parameters that model the discrete component are non-negligible for small sample sizes. We note that the analytical correction improves the performance of MLEs when the sample size is small in terms of the bias and relative bias. For example, when $n = 30$, the relative biases are 0.1235 for $\widehat{\gamma}_0$, and 0.1028 for $\widetilde{\gamma}_0$.

The MLEs and their corrected versions show similar performance in terms of root-mean-square error. We note that in general, the corrected estimators present a negligible precision gain in comparison to the MLEs of the parameters that model the discrete component, even in small sample sizes.

Table 3. Simulation results for estimation of β_0, β_1 and ϕ

n	Estimator	Bias	Rel. Bias	$\sqrt{\text{MSE}}$
30	$\widehat{\beta}_0$	0.0004	0.0009	0.1308
	$\widetilde{\beta}_0$	−0.0001	−0.0001	0.1541
60	$\widehat{\beta}_0$	−0.0016	−0.0030	0.0821
	$\widetilde{\beta}_0$	−0.0024	−0.0046	0.0818
90	$\widehat{\beta}_0$	−0.0004	−0.0007	0.0678
	$\widetilde{\beta}_0$	−0.0008	−0.0015	0.0677
30	$\widehat{\beta}_1$	0.0043	0.0024	0.2531
	$\widetilde{\beta}_1$	−0.0213	−0.0118	1.0772
60	$\widehat{\beta}_1$	0.0060	0.0033	0.1757
	$\widetilde{\beta}_1$	−0.0005	−0.0002	0.1748
90	$\widehat{\beta}_1$	0.0024	0.0013	0.1433
	$\widetilde{\beta}_1$	−0.0016	−0.0008	0.1429
30	$\widehat{\phi}$	42.3195	0.3526	86.5712
	$\widetilde{\phi}$	−8.6485	−0.0720	68.2596
60	$\widehat{\phi}$	16.7394	0.1395	41.2450
	$\widetilde{\phi}$	0.8568	0.0071	33.6214
90	$\widehat{\phi}$	8.9504	0.0745	27.4915
	$\widetilde{\phi}$	2.5856	0.0215	24.8168
30	$\widehat{\sigma}$	−0.0010	−0.1409	0.0029
	$\widetilde{\sigma}$	0.0002	0.0285	0.0074
60	$\widehat{\sigma}$	−0.0005	−0.0683	0.0020
	$\widetilde{\sigma}$	−0.0001	−0.0199	0.0021
90	$\widehat{\sigma}$	−0.0003	−0.0441	0.0015
	$\widetilde{\sigma}$	−0.0001	−0.0199	0.0015

Table 3 shows the results regarding the MLEs of the regression parameters of the continuous component of the model and their corrected versions. We note that the performance of these estimators is similar, regardless of the sample size. For example, for $n = 60$, the mean of $\widehat{\beta}_1$ differs from the mean of $\widetilde{\beta}_1$ in the third decimal place. Likewise, the root-mean-square errors are close. For example, for $n = 60$, $\sqrt{\text{MSE}} = 0.1757$ for $\widehat{\beta}_1$ and $\sqrt{\text{MSE}} = 0.1748$ for $\widetilde{\beta}_1$. Further, note that the biases and relative biases are close to zero. This indicates that

the MLEs for the regression parameters of the continuous component present good sample properties; that is, the estimated values of the regression parameters are close to the true values of the regression parameters, even when the sample size is relatively small. These results are consistent with those obtained by Ospina et al. (2006) for beta regression models.

The figures in Table 3 indicate that the MLEs of ϕ are markedly biased, and the estimated bias is positive for the different sample sizes considered. The bias correction proposed in this article significantly reduces the bias. For example, for $n = 60$, the bias of $\widehat{\phi}$ is 16.7394 while the bias of $\widetilde{\phi}$ is 0.8568. Consequently, the bias correction becomes very important, because if ϕ is overestimated, the variance of the response variable may be underestimated. In hypothesis testing, this may lead to the investigator erroneously rejecting the hypothesis that the regression parameters are zero. We also notice that the corrected estimator $\widetilde{\phi}$ presents smaller root-mean-square errors than $\widehat{\phi}$ for all the sample sizes considered. In other words, the corrected estimator is less biased and more accurate than the uncorrected MLE of ϕ. Therefore, we recommend the use of the bias corrected estimator for ϕ in practical applications.

Finally, we note that the MLE of $\sigma = 1/(\phi + 1)$ has a relative bias smaller than that of ϕ. For instance, for $n = 60$, $\widehat{\phi}$ has a relative bias equal to 0.1395 while the relative bias of $\widehat{\sigma}$ is -0.0683. In order to compare the accuracy of the MLE of σ with that of ϕ it is more convenient to use the scaled root-mean-square error ($\sqrt{\text{MSE}}$/parameter). In all of the cases, the MLE of σ outperforms the MLE of ϕ. For example, for $n = 60$, the scaled root-mean-square errors for $\widehat{\phi}$ is $\sqrt{\text{MSE}}/\phi = 0.34371$ and for $\widehat{\sigma}$ is $\sqrt{\text{MSE}}/\sigma = 0.2420$. It is therefore more convenient to parameterize the zero-or-one inflated beta regression model in terms of σ instead of the precision parameter ϕ, since the MLE of σ is less biased and more accurate than the MLE of ϕ.

5. Concluding Remarks

In this article, we derived expressions for the second-order bias of MLEs for inflated beta regression model parameters. We showed that the second-order biases obtained using the Cox and Snell (1968) formula may be written in terms of generalized least square regressions, which facilitates the calculation. The expressions found allow one to construct analytically modified MLEs with reduced bias.

The simulation results show that the bias corrections of the MLEs for the regression parameters that model the discrete and continuous components of inflated beta regression models are effective in reducing bias. However, bias corrections are not imperative because the uncorrected MLEs are not markedly biased. On the other hand, the MLE of the precision parameter is highly biased, and therefore, we recommend the use of the bias correction obtained in this article. Also, almost unbiased estimators are obtained if the model is parameterized in terms of $\sigma = 1/(\phi + 1)$ and the bias correction proposed in this paper is employed.

Appendix

Cumulants and their derivatives with respect to γ

For $R = 1, \ldots, M$, we have

$$U_{RSU} = \sum_{t=1}^{n} \left\{ \left(\frac{-2(1 - 1\!\!1_{\{c\}}(y_t))}{(1 - \alpha_t)^3} + \frac{2\,1\!\!1_{\{c\}}(y_t)}{\alpha_t^3} \right) \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 + 3 \left(\frac{-(1 - 1\!\!1_{\{c\}}(y_t))}{(1 - \alpha_t)^2} - \frac{1\!\!1_{\{c\}}(y_t)}{\alpha_t^2} \right) \left(\frac{\partial}{\partial \zeta_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right.$$
$$\left. + \left(\frac{1\!\!1_{\{c\}}(y_t)}{\alpha_t} - \frac{(1 - 1\!\!1_{\{c\}}(y_t))}{(1 - \alpha_t)} \right) \left[\left(\frac{\partial^2}{\partial \alpha_t^2} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} + \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right)^2 \right] \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU}.$$

By taking expected value, we have

$$\kappa_{RSU} = \sum_{t=1}^{n} \left\{ 2p_t^* \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 - 3p_t \left(\frac{\partial}{\partial \zeta_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU}, \tag{20}$$

where $p_t^* = \{-1/(1 - \alpha_t)^2\} + \{1/\alpha_t^2\}$. Also,

$$\kappa_{RS}^{(U)} = \frac{\partial \kappa_{RS}}{\partial \gamma_U} = \sum_{t=1}^{n} \left\{ p_t^* \left(\frac{d\alpha_t}{d\zeta_t} \right)^2 - 2p_t \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t} \right) \frac{d\alpha_t}{d\zeta_t} \right\} \frac{d\alpha_t}{d\zeta_t} z_{tS} z_{tR} z_{tU}, \tag{21}$$

for $U = 1, \ldots, M$. From the global orthogonality between γ and $(\beta^\top, \phi)^\top$ we obtain $\kappa_{RS}^{(\phi)} = \kappa_{RS}^{(u)} = 0$.

Now, from (20) and (21) we have

$$\kappa_{RS}^{(U)} - \frac{1}{2}\kappa_{RSU} = -\frac{1}{2}\sum_{t=1}^{n}\left\{p_t\left(\frac{\partial}{\partial\zeta_t}\frac{d\alpha_t}{d\zeta_t}\right)\frac{d\alpha_t}{d\zeta_t}\right\}\frac{d\alpha_t}{d\zeta_t}z_{tS}\,z_{tR}z_{tU},$$

where $p_t = 1/[\alpha_t(1-\alpha_t)]$. Let $W_0 = \text{diag}\{w_{01},\ldots,w_{0n}\}$ be the $n \times n$ diagonal matrix with t-th diagonal element given by

$$w_{0t} = \frac{1}{2}p_t\left(\frac{\partial}{\partial\zeta_t}\frac{d\alpha_t}{d\zeta_t}\right)\left(\frac{d\alpha_t}{d\zeta_t}\right)^2. \tag{22}$$

Hence,

$$\kappa_{RS}^{(U)} - \frac{1}{2}\kappa_{RSU} = \sum_{t=1}^{n}w_{0t}z_{tR}z_{tS}\,z_{tU}.$$

Consequently, the Cox and Snell (1968) formula to calculate the second-order bias of the MLE of the b-th element of γ is

$$B(\widehat{\gamma_b}) = \sum_{R,S,U}\kappa^{bR}\kappa^{SU}\left\{\kappa_{RS}^{(U)} - \frac{1}{2}\kappa_{RSU}\right\} = \sum_{t=1}^{n}w_{0t}\sum_{R}\kappa^{bR}z_{tR}\sum_{S,U}z_{tS}\kappa^{SU}z_{tU}.$$

Let $\underset{\sim}{e}_b$ be the b-th column of the $M \times M$ identity matrix. We have

$$\sum_{t=1}^{n}w_{0t}\sum_{R}\kappa^{bR}z_{tR}\sum_{S,U}z_{tS}\kappa^{SU}z_{tU} = = \underset{\sim}{e}_b^{\top}K^{\gamma\gamma}\sum_{t=1}^{n}w_{0t}z_t(z_t^{\top}K^{\gamma\gamma}z_t),$$

where z_t is the column vector obtained from the t-th row of Z. Now, let $\delta_{\gamma\gamma}$ represent the $n \times 1$ vector obtained from the main diagonal of $ZK^{\gamma\gamma}Z^{\top}$. We have

$$B(\widehat{\gamma_b}) = \sum_{R,S,U}\kappa^{bR}\kappa^{SU}\left\{\kappa_{RS}^{(U)} - \frac{1}{2}\kappa_{RSU}\right\} = \underset{\sim}{e}_b^{\top}K^{\gamma\gamma}Z^{\top}W_0\delta_{\gamma\gamma}. \tag{23}$$

Cumulants and their derivatives with respect to β and ϕ

We define the quantities $m_t = [\psi''(\mu_t\phi) - \psi''((1-\mu_t)\phi)]$, $u_t = -\phi\{2w_t + \phi[\mu_t m_t + \psi''((1-\mu_t)\phi)]\}$, $s_t = (1-\mu_t)^3\psi''((1-\mu_t)\phi) + \mu_t^3\psi''(\mu_t\phi) - \psi''(\phi)$, $\ddagger_t = [\mu_t w_t - \psi'((1-\mu_t)\phi)] + \phi[\mu_t^2\psi''(\mu_t\phi)]$, $c_t = \phi[\mu_t\psi'(\mu_t\phi) - (1-\mu_t)\psi'((1-\mu_t)\phi)]$, $\nabla_t = 2[\mu_t w_t - \psi'((1-\mu_t)\phi)] + \phi[\mu_t^2\psi''(\mu_t\phi) - (1-\mu_t)^2\psi''((1-\mu_t)\phi)]\frac{d\mu_t}{d\eta_t}$,

$$a_t = 3\left(\frac{\partial}{\partial\mu_t}\frac{d\mu_t}{d\eta_t}\right)\left(\frac{d\mu_t}{d\eta_t}\right)^2, \qquad b_t = \frac{d\mu_t}{d\eta_t}\left[\left(\frac{\partial^2}{\partial\mu_t^2}\frac{d\mu_t}{d\eta_t}\right)\frac{d\mu_t}{d\eta_t} + \left(\frac{\partial}{\partial\mu_t}\frac{d\mu_t}{d\eta_t}\right)^2\right].$$

We have

$$U_{rsu} = -\phi\sum_{t:y_t\in(0,1)}\left\{\phi^2 m_t\left(\frac{d\mu_t}{d\eta_t}\right)^3 + \phi w_t a_t - [y_t^* - \mu_t^*]b_t\right\}x_{ts}x_{tr}x_{tu},$$

$$U_{rs\phi} = \sum_{t:y_t\in(0,1)}\left\{\left(u_t\left(\frac{d\mu_t}{d\eta_t}\right) + [y_t^* - \mu_t^*]\left(\frac{\partial}{\partial\mu_t}\frac{d\mu_t}{d\eta_t}\right) - c_t\left(\frac{\partial}{\partial\mu_t}\frac{d\mu_t}{d\eta_t}\right)\right\}\frac{d\mu_t}{d\eta_t}x_{tr}x_{ts},$$

$$U_{r\phi\phi} = -\sum_{t:y_t\in(0,1)}\nabla_t x_{tr}, \qquad U_{\phi\phi\phi} = -\sum_{t:y_t\in(0,1)}s_t,$$

for $r, s, u = 1, \ldots, m$. Also,

$$\kappa_{rsu} = -\phi^2 \sum_{t=1}^{n}(1-\alpha_t)\left\{\phi m_t\left(\frac{d\mu_t}{d\eta_t}\right)^3 + w_t a_t\right\}x_{ts}x_{tr}x_{tu}, \qquad \kappa_{\phi\phi\phi} = -\sum_{t=1}^{n}(1-\alpha_t)s_t.$$

$$\kappa_{rs\phi} = \sum_{t=1}^{n}(1-\alpha_t)\left\{u_t\left(\frac{d\mu_t}{d\eta_t}\right) - \mathbf{c}_t\left(\frac{\partial}{\partial\mu_t}\frac{d\mu_t}{d\eta_t}\right)\right\}\frac{d\mu_t}{d\eta_t}x_{tr}x_{ts}, \qquad \kappa_{r\phi\phi} = -\sum_{t=1}^{n}(1-\alpha_t)\nabla_t x_{tr},$$

$$\kappa_{rs}^{(u)} = -\phi^2 \sum_{t=1}^{n}(1-\alpha_t)\left(\phi m_t\left(\frac{d\mu_t}{d\eta_t}\right)^3 + \frac{2}{3}w_t a_t\right)x_{tr}x_{ts}x_{tu}, \qquad \kappa_{rs}^{(\phi)} = \sum_{t=1}^{n}(1-\alpha_t)\left\{u_t\left(\frac{d\mu_t}{d\eta_t}\right)^2\right\}x_{ts}x_{tr},$$

$$\kappa_{r\phi}^{(u)} = -\sum_{t=1}^{n}(1-\alpha_t)\left\{[\phi(w_t + \mu_t\phi m_t + \phi\psi''((1-\mu_t)\phi))]\frac{d\mu_t}{d\eta_t} + \mathbf{c}_t\left(\frac{\partial}{\partial\mu_t}\frac{d\mu_t}{d\eta_t}\right)\right\}\frac{d\mu_t}{d\eta_t}x_{tr}x_{tu},$$

$$\kappa_{r\phi}^{(\phi)} = -\sum_{t=1}^{n}(1-\alpha_t)\ddagger_t\frac{d\mu_t}{d\eta_t}x_{tr}, \qquad \kappa_{\phi\phi}^{(u)} = \sum_{t=1}^{n}-(1-\alpha_t)\nabla_t\frac{d\mu_t}{d\eta_t}x_{tu}, \qquad \kappa_{\phi\phi}^{(\phi)} = -\sum_{t=1}^{n}(1-\alpha_t)s_t.$$

Let $W_1 = \text{diag}\{w_{11},\ldots,w_{1n}\}$, $W_2 = \text{diag}\{w_{21},\ldots,w_{2n}\}$, $W_3 = \text{diag}\{w_{31},\ldots,w_{3n}\}$, $W_4 = \text{diag}\{w_{41},\ldots,w_{4n}\}$ and $W_5 = \text{diag}\{w_{51},\ldots,w_{5n}\}$ with

$$w_{1t} = -\frac{\phi^2}{2}[\phi m_t\left(\frac{d\mu_t}{d\eta_t}\right)^3 + \frac{1}{3}w_t a_t], \qquad w_{2t} = \frac{1}{2}\left\{u_t\left(\frac{d\mu_t}{d\eta_t}\right) + \mathbf{c}_t\left(\frac{\partial}{\partial\mu_t}\frac{d\mu_t}{d\eta_t}\right)\right\}\frac{d\mu_t}{d\eta_t},$$

$$w_{3t} = -\frac{1}{2}\left\{\phi^2[\mu_t m_t + \psi''((1-\mu)\phi)]\frac{d\mu_t}{d\eta_t} + \mathbf{c}_t\left(\frac{\partial}{\partial\mu_t}\frac{d\mu_t}{d\eta_t}\right)\right\}\frac{d\mu_t}{d\eta_t},$$

$$w_{4t} = \frac{-1}{2}\phi[\mu_t^2\psi''(\mu_t\phi) - (1-\mu_t)^2\psi''((1-\mu_t)\phi)]\frac{d\mu_t}{d\eta_t}, \qquad w_{5t} = -\frac{1}{2}\nabla_t\frac{d\mu_t}{d\eta_t}.$$

After some algebra we obtain

$$\kappa_{rs}^{(u)} - \frac{1}{2}\kappa_{rsu} = \sum_{t=1}^{n}\delta_t w_{1t}x_{tr}x_{ts}x_{tu}, \qquad \kappa_{rs}^{(\phi)} - \frac{1}{2}\kappa_{rs\phi} = \sum_{t=1}^{n}\delta_t w_{2t}x_{tr}x_{ts},$$

$$\kappa_{r\phi}^{(u)} - \frac{1}{2}\kappa_{r\phi u} = \sum_{t=1}^{n}\delta_t w_{3t}x_{tr}x_{tu}, \qquad \kappa_{r\phi}^{(\phi)} - \frac{1}{2}\kappa_{r\phi\phi} = \sum_{t=1}^{n}\delta_t w_{4t}x_{tr},$$

$$\kappa_{\phi\phi}^{(u)} - \frac{1}{2}\kappa_{\phi\phi u} = \sum_{t=1}^{n}\delta_t w_{5t}x_{tu}, \qquad \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2}\kappa_{\phi\phi\phi} = \sum_{t=1}^{n}-\frac{1}{2}\delta_t s_t.$$

We can now obtain the second-order bias of the MLE a-th element of $\widehat{\beta}$ from (14), with $a = 1,\ldots,m$. For $a = 1,\ldots,m$ we have

$$\sum_{r,s,u}\kappa^{ar}\kappa^{su}\left\{\kappa_{rs}^{(u)} - \frac{1}{2}\kappa_{rsu}\right\} = \underset{\sim}{i}_a^{\top}K^{\beta\beta}X^{\top}\Delta W_1\delta_{\beta\beta},$$

where $\underset{\sim}{i}_a$ is the a-th column vector of the $m \times m$ identity matrix and $\delta_{\beta\beta}$ is the $n \times 1$ vector obtained from the main diagonal of $XK^{\beta\beta}X^{\top}$. We have

$$\sum_{r,s,\phi}\kappa^{ar}\kappa^{s\phi}\left\{\kappa_{rs}^{(\phi)} - \frac{1}{2}\kappa_{rs\phi}\right\} = \underset{\sim}{i}_a^{\top}K^{\beta\beta}(X^{\top}\Delta W_2 X)K^{\beta\phi},$$

$$\sum_{\phi,s,u}\kappa^{a\phi}\kappa^{su}\left\{\kappa_{\phi s}^{(u)} - \frac{1}{2}\kappa_{\phi su}\right\} = \underset{\sim}{i}_a^{\top}K^{\beta\phi}\,\text{tr}(\Delta W_3 XK^{\beta\beta}X^{\top}),$$

$$\sum_{r,\phi,u}\kappa^{ar}\kappa^{\phi u}\left\{\kappa_{r\phi}^{(u)} - \frac{1}{2}\kappa_{r\phi u}\right\} = \underset{\sim}{i}_a^{\top}K^{\beta\beta}(X^{\top}\Delta W_3 X)K^{\beta\phi}.$$

Similarly, if ρ_t is the t-th column vector $n \times n$ identity matrix we obtain

$$\sum_{\phi,s,\phi} \kappa^{a\phi} \kappa^{s\phi} \left\{ \kappa_{\phi s}^{(\phi)} - \frac{1}{2}\kappa_{\phi s\phi} \right\} = \underline{i}_a^\top \, K^{\beta\phi} \mathrm{diagonal}(W_4) X K^{\beta\phi},$$

$$\sum_{\phi,\phi,u} \kappa^{a\phi} \kappa^{\phi u} \left\{ \kappa_{\phi\phi}^{(u)} - \frac{1}{2}\kappa_{\phi\phi u} \right\} = \underline{i}_a^\top \, K^{\beta\phi} \mathrm{diagonal}(\Delta W_5) X K^{\beta\phi},$$

$$\sum_{r,\phi,\phi} \kappa^{ar} \kappa^{\phi\phi} \left\{ \kappa_{r\phi}^{(\phi)} - \frac{1}{2}\kappa_{r\phi\phi} \right\} = \underline{i}_a^\top \, K^{\beta\beta} X^\top \mathrm{diagonal}(W_4)^\top K^{\phi\phi}.$$

Analogously, if $S = \mathrm{diag}\{-s_1/2, \ldots, -s_n/2\}$ we have

$$\sum_{\phi,\phi,\phi} \kappa^{a\phi} \kappa^{\phi\phi} \left\{ \kappa_{\phi\phi}^{(\phi)} - \frac{1}{2}\kappa_{\phi\phi\phi} \right\} = \underline{i}_a^\top \, K^{\beta\phi} K^{\phi\phi} \mathrm{tr}(\Delta S).$$

Acknowledgements

References

Botter, D. A., & Cordeiro, G. M. (1998). Improved estimators for generalized linear models with dispersion covariates. *Communications in Statistics, Simulation and Computation, 59*, 1-14. http://dx.doi.org/10.1080/00949659808811926

Cook, R. D., Tsai, C., & Wei, B. (1986). Bias in nonlinear regression. *Biometrika, 73*, 615-623. http://dx.doi.org/10.1093/biomet/73.3.615

Cordeiro, G. M., & Cribari–Neto, F. (1998). On bias reduction in exponential and non-exponential family regression models. *Communications in Statistics, Simulation and Computation, 27*, 485-500. http://dx.doi.org/10.1080/03610919808813491

Cordeiro, G. M., & McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal-Statististical Society B, 53*, 629-643.

Cordeiro, G. M., Rocha, E. C., Rocha, J. G. C., & Cribari–Neto, F. (1997). Bias-corrected maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation, 58*, 21-35. http://dx.doi.org/10.1080/00949659708811820

Cordeiro, G. M., & Toyama, M. C. (2009). Bias correction in generalized nonlinear models with dispersion covariates. *Communications in Statistics, Theory and Methods, 37*, 2219-2225. http://dx.doi.org/10.1080/03610920801931895

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

Cribari–Neto, F., & Vasconcellos, K. L. P. (2002). Nearly unbiased maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation, 72*, 107-118. http://dx.doi.org/10.1080/00949650212144

Cribari–Neto, F., & Zarkos, S. G. (2003). Econometric and statistical computing using Ox. *Computational Economics, 21*, 277-295. http://dx.doi.org/10.1023/A:1023902027800

Cribari–Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software, 34*(2).

Ferrari, S. L. P., & Cribari–Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics, 7*, 799-815. http://dx.doi.org/10.1080/0266476042000214501

Grün, B., Kosmidis, I., & Zeileis, A. (2012). Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned. *Journal of Statistical Software, 48*(11).

Kieschnick R., & McCullough, B. D. (2003). Regression analysis of variates observed on (0,1): percentages, proportions, and fractions. *Statistical Modelling, 3*, 1-21. http://dx.doi.org/10.1191/1471082X03st053oa

Mittelhammer, R. C., Judge, G. G., & Miller, D. J. (2000). *Econometric Foundations.* New York: Cambridge University Press.

Ospina, R., Cribari–Neto, F., & Vasconcellos, K. L. P. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis, 51*(2), 960-981. http://dx.doi.org/10.1016/j.csda.2005.10.002. Erratum: *55*, 2445-2445. http://dx.doi.org/10.1016/j.csda.2011.02.015

Ospina, R., & Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers, 51*, 111-126. http://dx.doi.org/10.1007/s00362-008-0125-4

Ospina, R., & Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis, 56*, 1609-1623. http://dx.doi.org/10.1016/j.csda.2011.10.005

Pace, L., & Salvan, A. (1997). *Principles of Statistical Inference. Singapore: World Scientific in Advanced Series on Statistical Science & Applied Probability, 4.*

Paolino, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis, 9*, 325-346.

Press, W. H., Teulosky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing* (2nd ed.). Cambridge: Cambridge University Press.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (2nd ed.). New York: Wiley.

Smithson, M., & Verkuilen, J. (2006). A better lemon-squeezer? Maximum likelihood regression with beta-distributed dependent variables. *Psychological Methods, 11*, 54-71. http://dx.doi.org/10.1037/1082-989X.11.1.54

Vasconcellos, K. L. P., & Cribari–Neto, F. (2005). Improved maximum likelihood estimation in a new class of beta regression models. *Brazilian Journal of Probability and Statistics, 19*, 13-31.

Notes

Note 1. Note that c can be fixed at any constant value in $[0, 1]$. In practical applications, the inflation point occurs most frequently in $c = 0$ or $c = 1$.

Note 2. The digamma function is defined as $\psi(x) = \mathrm{d} \log \Gamma(x)/\mathrm{d}x, \; x > 0$.