Brazilian Journal of Probability and Statistics (2005), 19, pp. 13–31.

Improved maximum likelihood estimation in a new class of beta regression models

Klaus L.P. Vasconcellos and **Francisco Cribari–Neto**
Universidade Federal de Pernambuco

Abstract: We propose a class of regression models where the response is beta distributed and the two parameters that index the beta distribution are related to covariates and regression parameters. The proposed class of models is useful for modeling data that are restricted to the $(0, 1)$ interval. We discuss maximum likelihood estimation of the parameters that define the regression structure of the model, and derive closed-form expressions for the second order biases of these estimators. The derived expressions are then used to define bias-corrected maximum likelihood estimators. Simulation results show that the bias correction scheme yields nearly unbiased estimators.

Key words: Beta distribution, bias correction, maximum likelihood estimation, regression.

1 Introduction

An important area of research in statistics is the study of the finite-sample behavior of maximum likelihood estimators (MLEs). We know that MLEs are oftentimes biased, thus displaying systematic error. This is not a serious problem for relatively large sample sizes, since bias is typically of order $O(n^{-1})$, while the asymptotic standard errors are of order $O(n^{-1/2})$. However, for small or even moderate values of the sample size n, bias can constitute a problem. Thus, availability of formulae for its approximate computation is important for good estimation performance of many models that are used in a number of applications. Bias correction of MLEs is particularly important when the sample size, or the total information, is small.

Bias adjustment has been extensively studied in the statistical literature. Box (1971) gives a general expression for the n^{-1} bias in multivariate nonlinear models where covariance matrices are known. Pike, Hill and Smith (1980) investigate estimation bias in logistic linear models. For nonlinear regression models, Cook, Tsai and Wei (1986) relate bias to the position of the explanatory variables in the sample space. Young and Bakir (1987) show that bias correction can improve estimation in generalized log-gamma regression models. Cordeiro and McCullagh (1991) give general matrix formulae for bias correction in generalized linear models. More recently, Cordeiro and Vasconcellos (1997) obtained general matrix formulae for bias correction in multivariate nonlinear regression models with normal errors, while Vasconcellos and Cordeiro (2000) obtained bias correction for multivariate nonlinear Student t regression models. Also, Cordeiro and Vasconcellos (1999) obtain second order biases of the maximum likelihood estimators in von Mises regression models, while Vasconcellos, Cordeiro and Barroso (2000) derive bias corrected estimators for heteroskedastic univariate regression models with Student t error.

Practitioners oftentimes need to model data that are restricted to the $(0, 1)$ interval, such as, e.g., the unemployment rate and income concentration measures, among others. The beta distribution is very flexible for modeling such data since its density can display quite different shapes depending on the parameter values, the uniform distribution being one of the many special cases. Bury (1999) lists applications of the beta distribution in engineering. Johnson, Kotz and Balakrishnan (1995) also present and discuss a number of applications of the beta distribution. According to them (p. 235), "[t]he beta distributions are among the most frequently employed to model theoretical distributions." Krysicki (1999) presents some new properties of the beta distribution.

In this paper we define a class of models based on the beta distribution where the two parameters (p and q) that index the distribution have regression structures defined by sets of explanatory variables. The proposed model therefore allows one to study the relationship between the variable of interest and other variables that affect its behavior, and can be used whenever the response is measured as a rate or as a proportion. The proposed model is quite general. A special case of particular interest is when q is regressor-free and the link function for p is exponential, in which case the regression parameters can be interpreted in terms of mean response effects.

We discuss maximum likelihood estimation of the regression parameters, obtaining the log-likelihood function, the score function and Fisher's information matrix. As is well known, however, maximum likelihood estimators, although consistent, are typically biased in finite samples. In order to overcome this shortcoming, we derive a closed-form expression for the bias of the maximum likelihood estimator, and use it to define a bias-adjusted maximum likelihood estimator. Bias corrections for maximum likelihood estimators of the parameters that index the beta distribution were obtained and studied by Cordeiro, Rocha, Rocha & Cribari–Neto (1997) and Cribari–Neto & Vasconcellos (2002). However, their results do not apply to models with regression structures, as the one proposed in this paper. Our focus is on situations where the variable of interest is related to other variables, and such dependence is exploited when modeling the response.

The plan of the paper is as follows. Section 2 introduces the class of beta regression and discusses the estimation of the model parameters by maximum likelihood. Section 3 derives the second order biases of the maximum likelihood estimators of the beta regression parameters. The result is used to define bias-adjusted maximum likelihood estimators. Monte Carlo simulation results are presented and discussed in Section 4. The numerical results show that the bias correction we derive is effective in small samples; it delivers estimators that are nearly unbiased and display superior finite-sample behavior. Finally, Section 5 concludes the paper.

2 The model and maximum likelihood estimation

Let Y have a beta distribution with parameters p and q, i.e., $Y \sim \mathcal{B}(p,q)$. Then, its density function is

$$f(y;p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}$$

and the log-density is

$$\log f(y;p,q) = \log\Gamma(p+q) - \log\Gamma(p) - \log\Gamma(q) + (p-1)\log y + (q-1)\log(1-y).$$

The rth moment of Y about the origin is given by

$$\mu_r'(Y) = \frac{\Gamma(p+q)\Gamma(p+r)}{\Gamma(p)\Gamma(p+q+r)}.$$

Therefore, the mean of Y is given by $p/(p+q)$ and its variance by $pq/\{(p+q)^2(p+q+1)\}$. The mode exists if $p > 1$ and $q > 1$, in which case it equals $(p-1)/(p+q-2)$.

The first order derivatives of the log-density are

$$\frac{\partial \log f(y;p,q)}{\partial p} = \psi(p+q) - \psi(p) + \log y,$$
$$\frac{\partial \log f(y;p,q)}{\partial q} = \psi(p+q) - \psi(q) + \log(1-y),$$

where ψ is the digamma function. Since the expected score equals zero, it follows that

$$E[\log Y] = \psi(p) - \psi(p+q),$$
$$E[\log(1-Y)] = \psi(q) - \psi(p+q).$$

In what follows we propose a regression structure that allows the modeling of relationships between random variables that follow a beta distribution and a set of explanatory variables. The proposed model is defined by establishing relationships between the parameters that index the beta distribution, p and q, and linear predictors on independent variables.

We consider the model where the observations $Y_1, \ldots, Y_n$ are beta distributed and independent. The distribution of Y_i is $\mathcal{B}(p_i, q_i)$, where p_i and q_i are, for each i, described by sets of explanatory variables $\{x_1, \ldots, x_m\}$ and $\{v_1, \ldots, v_M\}$ as

$$p_i = g(\beta_1 x_{1i} + \cdots + \beta_m x_{mi}),$$
$$q_i = h(\gamma_1 v_{1i} + \cdots + \gamma_M v_{Mi}).$$

Here, g and h are link functions and assumed real, strictly positively valued and continuously differentiable up to the third order. The regression parameters $\beta_1, \ldots, \beta_m$ and $\gamma_1, \ldots, \gamma_M$ are unknown.

Let $\theta = (\beta^T, \gamma^T)^T$ be the joint parameter vector. The log-likelihood function is given, apart from unimportant constants, by

$$\ell(\theta) = \sum_{i=1}^{n} \log \Gamma(p_i + q_i) - \sum_{i=1}^{n} \log \Gamma(p_i) - \sum_{i=1}^{n} \log \Gamma(q_i) + \sum_{i=1}^{n} p_i \log y_i + \sum_{i=1}^{n} q_i \log(1 - y_i).$$

The first order derivatives with respect to the regression parameters are

$$\frac{\partial \ell}{\partial \beta_r} = \sum_{i=1}^{n} g_i' x_{ri} [\psi(p_i + q_i) - \psi(p_i) + \log y_i],$$

$$\frac{\partial \ell}{\partial \gamma_R} = \sum_{i=1}^{n} h_i' v_{Ri} [\psi(p_i + q_i) - \psi(q_i) + \log(1 - y_i)],$$

where β_r and γ_R denote the rth and Rth elements of the m dimensional β and and M dimensional γ vectors, respectively. Also, g_i' is the first derivative of g_i with respect to its argument, this derivative being evaluated at $\eta_i = \beta_1 x_{1i} + \cdots + \beta_m x_{mi}$. Also, h_i' represents the first derivative of h_i with respect to its argument, the derivative being evaluated at $\tau_i = \gamma_1 v_{1i} + \cdots + \gamma_M v_{Mi}$.

Maximum likelihood estimators of the β and γ vectors can be obtained by equating the above derivatives to zero and solving the resulting system. There are no closed form expressions for these estimators, and their computation has to be performed numerically using a nonlinear optimization algorithm. We shall assume that the usual regularity conditions for maximum likelihood estimation hold (Cox and Hinkley, 1974, Chapter 9).

It is oftentimes desirable for the regression parameters to be interpretable in terms of the mean response. The parameters in the proposed model have such interpretation, especially when there is a regression structure for p and not for q, and the link function is exponential. That is, suppose we have

$$p_i = \exp\{\beta_1 x_{1i} + \cdots + \beta_m x_{mi}\}.$$

It is easy to show that the derivative of μ with respect to x_j, $j = 1, \ldots, m$, is, then, given by

$$\frac{\partial \mu}{\partial x_j} = \beta_j \mu (1 - \mu). \tag{2.1}$$

This function is plotted in Figure 1 for three values of β_j, namely 0.8, 1.0 and 1.2. We note that: (i) The impact on μ of a change in one of the independent variables is greatest when $\mu = 0.5$ and smallest when μ is close to 0 or 1. (ii) The sign effect is given by the sign of β_j. This is similar to what occurs in logit and probit regression models for binary responses. (iii) The relative effect,

$$\frac{\partial \mu / \partial x_j}{\partial \mu / \partial x_k} = \frac{\beta_j}{\beta_k}, \quad j \neq k = 1, \ldots, m,$$

does not depend on the covariates; that is, the ratio of partial effects is constant and given by the ratio of the corresponding regression coefficients. These three features are also present in logit and probit regression models for binary responses. For a general link function, it follows that

$$\frac{\partial\mu}{\partial x_j} = \lambda'\beta_j\mu(1-\mu), \tag{2.2}$$

where $\lambda = \log(g)$ and, consequently, $\lambda' = g'/g$, where primes here denote derivatives with respect to x_j. When the link function is exponential, λ becomes the identity function, $\lambda' = 1$, and thus (2.2) reduces to (2.1).

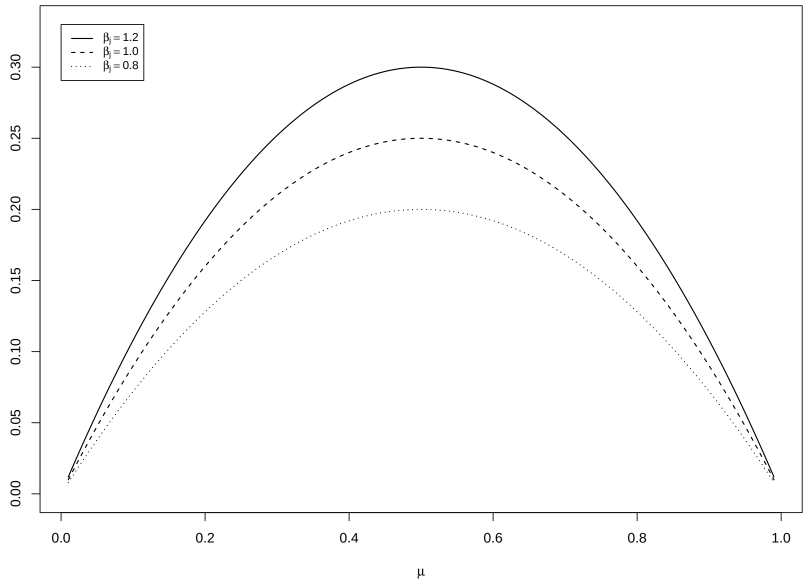

Figure 1 *Derivative of μ with respect to x_j, $j = 1, \ldots, m$.*

It is also desirable to report a measure of the global goodness of fit when estimating regression models. A measure that can be used for the proposed model is the pseudo R^2 defined as

$$R_p^2 = 1 - \frac{\ell_R}{\ell_U},$$

where ℓ_U is the unrestricted (all covariates) maximized log-likelihood and ℓ_R is the restricted (no covariates) maximized log-likelihood. It is clear that R_p^2 lies between 0 and 1, and thus it is a plausible measure of goodness of fit. A similar global measure of fit for binary regression models was proposed by McFadden (1974).

It is well known that maximum likelihood estimators are consistent, asymptotically efficient, and asymptotically normal. However, they can be substantially biased when the sample size is small. Bias is a systematic error in the estimation process, and it is thus undesirable. The next section develops finite-sample adjustments that can be applied to the maximum likelihood estimators of β and γ to greatly reduce small sample bias of these estimators. The main idea is to obtain closed-form expressions for the second order biases of $\widehat{\beta}$ and $\widehat{\gamma}$ and to use these expressions to bias-correct the maximum likelihood estimators.

3 Improved estimation in small samples

At the outset, we shall introduce some notation. The total log-likelihood derivatives with respect to the unknown parameters are indicated by indices, where lower case letters $r, s, t, \ldots$ correspond to derivatives with respect to the β parameters, while upper case letters $R, S, T, \ldots$ correspond to derivatives with respect to the γ parameters. Thus, $U_r = \partial\ell/\partial\beta_r$, $U_R = \partial\ell/\partial\gamma_R$, $U_{Rs} = \partial^2\ell/\partial\gamma_R\partial\beta_s$, $U_{rsT} = \partial^3\ell/\partial\beta_r\partial\beta_s\partial\gamma_T$ and so on. The standard notation for the moments of these derivatives is used here (Lawley, 1956): $\kappa_{rs} = E(U_{rs}), \kappa_{R,S} = E(U_R U_S), \kappa_{rs,T} = E(U_{rs}U_T), \kappa_{rst} = E(U_{rst})$, etc., where all κ's refer to a total over the sample, and are, in general, of order n. Also, their derivatives are denoted by $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t, \kappa_{rS}^{(T)} = \partial\kappa_{rS}/\partial\gamma_T$, etc. Finally, κ^{rs} denotes the (r, s) element of the inverse of Fisher's information matrix.

For simplicity, we shall start with the case where the same explanatory variables are used in the regression structures of p and q. We shall show afterwards that the results obtained can be easily generalized to the more general setting.

The second derivatives of the log-likelihood function are

$$U_{rt} = \sum_{i=1}^{n} g_i^{''} x_{ri}x_{ti}[\psi(p_i+q_i) - \psi(p_i) + \log y_i] + \sum_{i=1}^{n} g_i'^2 x_{ri}x_{ti}[\psi'(p_i+q_i) - \psi'(p_i)],$$

$$U_{rT} = \sum_{i=1}^{n} g_i'h_i'x_{ri}x_{Ti}\psi'(p_i+q_i),$$

$$U_{RT} = \sum_{i=1}^{n} h_i^{''} x_{Ri}x_{Ti}[\psi(p_i+q_i) - \psi(q_i) + \log(1-y_i)] + \sum_{i=1}^{n} h_i'^2 x_{Ri}x_{Ti}[\psi'(p_i+q_i) - \psi'(q_i)].$$

Therefore, their second-order moments can be written as

$$\begin{aligned}
\kappa_{rt} &= \sum_{i=1}^{n} g_i'^2 x_{ri} x_{ti} [\psi'(p_i+q_i) - \psi'(p_i)],\\
\kappa_{rT} &= \sum_{i=1}^{n} g_i' h_i' x_{ri} x_{Ti} \psi'(p_i+q_i),\\
\kappa_{RT} &= \sum_{i=1}^{n} h_i'^2 x_{Ri} x_{Ti} [\psi'(p_i+q_i) - \psi'(q_i)].
\end{aligned}$$

Let $W_{\beta\beta}, W_{\beta\gamma}$ and $W_{\gamma\gamma}$ be $n \times n$ diagonal matrices defined as

$$\begin{aligned}
W_{\beta\beta} &= \operatorname{diag}\{g_i'^2[\psi'(p_i) - \psi'(p_i+q_i)]\},\\
W_{\beta\gamma} &= \operatorname{diag}\{-g_i' h_i' \psi'(p_i+q_i)\},\\
W_{\gamma\gamma} &= \operatorname{diag}\{h_i'^2[\psi'(q_i) - \psi'(p_i+q_i)]\},
\end{aligned}$$

and consider the $(2n) \times (2n)$ matrix $\widetilde{W}$ given by

$$\widetilde{W} = \begin{pmatrix} W_{\beta\beta} & W_{\beta\gamma} \\ W_{\beta\gamma} & W_{\gamma\gamma} \end{pmatrix}.$$

Also let X be the $n \times m$ matrix with the values of the explanatory variables for each observation and define $\widetilde{X} = I_2 \otimes X$, where I_2 denotes the 2×2 identity matrix and $\otimes$ stands for the Kronecker product. Then, we can readily see that the information matrix for θ is $K_\theta = \widetilde{X}^T \widetilde{W} \widetilde{X}$.

Our goal is to obtain a closed-form expression for the bias of the maximum likelihood estimator of θ. This expression will then be used to define a modified maximum likelihood estimator that, unlike the original estimator, is nearly unbiased in small samples. To that end, we shall use Cox and Snell's (1968) general formula, namely:

$$\begin{aligned}
B(\widehat{\beta}_s) = &\sum_{r,t,u} \kappa^{sr}\kappa^{tu}\left\{\kappa_{rt}^{(u)} - \frac{1}{2}\kappa_{rtu}\right\} + \sum_{R,t,u} \kappa^{sR}\kappa^{tu}\left\{\kappa_{Rt}^{(u)} - \frac{1}{2}\kappa_{Rtu}\right\}\\
&+ \sum_{r,T,u} \kappa^{sr}\kappa^{Tu}\left\{\kappa_{rT}^{(u)} - \frac{1}{2}\kappa_{rTu}\right\} + \sum_{r,t,U} \kappa^{sr}\kappa^{tU}\left\{\kappa_{rt}^{(U)} - \frac{1}{2}\kappa_{rtU}\right\}\\
&+ \sum_{R,T,u} \kappa^{sR}\kappa^{Tu}\left\{\kappa_{RT}^{(u)} - \frac{1}{2}\kappa_{RTu}\right\} + \sum_{R,t,U} \kappa^{sR}\kappa^{tU}\left\{\kappa_{Rt}^{(U)} - \frac{1}{2}\kappa_{RtU}\right\}\\
&+ \sum_{r,T,U} \kappa^{sr}\kappa^{TU}\left\{\kappa_{rT}^{(U)} - \frac{1}{2}\kappa_{rTU}\right\} + \sum_{R,T,U} \kappa^{sR}\kappa^{TU}\left\{\kappa_{RT}^{(U)} - \frac{1}{2}\kappa_{RTU}\right\},
\end{aligned}$$

where $B(\widehat{\beta}_s)$ denotes the second order bias of the maximum likelihood estimator of β_s. We need to calculate the necessary cumulants so that the second order bias, following Cox and Snell, can be obtained. It is important to observe that the entries of the matrix $W_{\beta\gamma}$ defined before will, in general, not be zero, which means that the β and γ vectors will not be orthogonal. This introduces a complication, since none of the eight terms in the above expansion will vanish and all of them have to be obtained.

It can be shown (see the Appendix) that the second order bias for the maximum likelihood estimator of β is given by

$$\begin{aligned} B(\widehat{\beta}) &= \frac{1}{2}K^{\beta\beta}X^T(W_1\delta_{\beta\beta} + W_2\delta_{\beta\gamma} + W_3\delta_{\beta\gamma} + W_5\delta_{\gamma\gamma}) \\ &+ \frac{1}{2}K^{\beta\gamma}X^T(W_2\delta_{\beta\beta} + W_4\delta_{\beta\gamma} + W_5\delta_{\beta\gamma} + W_6\delta_{\gamma\gamma}). \end{aligned}$$

(The matrices $W_1, \ldots, W_6$ and the vectors $\delta_{\beta\beta}$, $\delta_{\beta\gamma}$ and $\delta_{\gamma\gamma}$ are defined in the Appendix.) The matrices $K^{\beta\beta}$ and $K^{\beta\gamma}$ are the corresponding blocks of K_θ^{-1}. Now, define

$$K^{\beta *} = \begin{pmatrix} K^{\beta\beta} & K^{\beta\gamma} \end{pmatrix},$$

the upper $m \times 2m$ block of the matrix K_θ^{-1}. We can thus write

$$B(\widehat{\beta}) = \frac{1}{2}K^{\beta *}\widetilde{X}^T\widetilde{\delta},$$

where $\widetilde{\delta}$ is the $2n \times 1$ vector

$$\begin{pmatrix} W_1\delta_{\beta\beta} + W_2\delta_{\beta\gamma} + W_3\delta_{\beta\gamma} + W_5\delta_{\gamma\gamma} \\ W_2\delta_{\beta\beta} + W_4\delta_{\beta\gamma} + W_5\delta_{\beta\gamma} + W_6\delta_{\gamma\gamma} \end{pmatrix}.$$

Note that the matrix $\widetilde{X}K_\theta^{-1}\widetilde{X}^T$ is given by

$$\begin{pmatrix} XK^{\beta\beta}X^T & XK^{\beta\gamma}X^T \\ XK^{\gamma\beta}X^T & XK^{\gamma\gamma}X^T \end{pmatrix}.$$

Now define the $2n \times 2n$ matrices Z_1 and Z_2, respectively, as

$$Z_1 = \begin{pmatrix} W_1 & W_2 \\ W_3 & W_5 \end{pmatrix} \odot \begin{pmatrix} XK^{\beta\beta}X^T & XK^{\beta\gamma}X^T \\ XK^{\gamma\beta}X^T & XK^{\gamma\gamma}X^T \end{pmatrix}$$

and

$$Z_2 = \begin{pmatrix} W_2 & W_4 \\ W_5 & W_6 \end{pmatrix} \odot \begin{pmatrix} XK^{\beta\beta}X^T & XK^{\beta\gamma}X^T \\ XK^{\gamma\beta}X^T & XK^{\gamma\gamma}X^T \end{pmatrix},$$

where $A \odot B$ represents the Hadamard product, defined by $(A \odot B)_{ij} = (A)_{ij}(B)_{ij}$, if A and B are two matrices of the same dimension. Now let $\widetilde{Z}$ be the $4n \times 4n$ matrix

$$\widetilde{Z} = \begin{pmatrix} Z_1 & 0 \\ 0 & Z_2 \end{pmatrix},$$

where the 0's represent $2n \times 2n$ matrices of zeros. Also, let $\widetilde{J}$ be the $4n \times 2n$ matrix given by

$$\widetilde{J} = \left(\begin{array}{cc} I_n & 0 \\ I_n & 0 \\ 0 & I_n \\ 0 & I_n \end{array} \right) = I_2 \otimes \left(\begin{array}{c} 1 \\ 1 \end{array} \right) \otimes I_n,$$

the 0's representing $n \times n$ matrices of zeros. We note that $\widetilde{\delta}$ is the $2n \times 1$ vector containing the diagonal elements of $\widetilde{J}^T \widetilde{Z} \widetilde{J}$.

We now move to the second order bias of the maximum likelihood estimator of γ. It is not difficult to conclude that the second order bias of $\widehat{\gamma}$ is given by

$$B(\widehat{\gamma}) = \frac{1}{2} K^{\gamma *} \widetilde{X}^T \widetilde{\delta},$$

where

$$K^{\gamma *} = \left(\begin{array}{cc} K^{\gamma\beta} & K^{\gamma\gamma} \end{array} \right)$$

is the lower $m \times 2m$ block of the matrix K_θ^{-1}.

Therefore, if we consider the joint parameter vector $\theta = (\beta^T, \gamma^T)^T$, we can write the expression for the second order bias of the maximum likelihood estimator of θ as

$$B(\widehat{\theta}) = \frac{1}{2} K_\theta^{-1} \widetilde{X}^T \widetilde{\delta},$$

or, alternatively, as

$$B(\widehat{\theta}) = \frac{1}{2} \left(\widetilde{X}^T \widetilde{W} \widetilde{X} \right)^{-1} \widetilde{X}^T \widetilde{\delta}.$$

Also, if we define the vector $\widetilde{\varphi} = \frac{1}{2} \widetilde{W}^{-1} \widetilde{\delta}$, it is possible to write

$$B(\widehat{\theta}) = \left(\widetilde{X}^T \widetilde{W} \widetilde{X} \right)^{-1} \widetilde{X}^T \widetilde{W} \widetilde{\varphi},$$

which means that the bias vector can be given by the estimated coefficients of a generalized least squares regression.

The above result can be used to define a bias-adjusted maximum likelihood estimator. The improved estimator $\widetilde{\theta}$ is defined as

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

where $\widehat{B}(\widehat{\theta})$ denotes the maximum likelihood estimator of $B(\widehat{\theta})$, that is, the unknown parameters in $B(\widehat{\theta})$ are replaced by their maximum likelihood estimates. The corrected estimator is expected to have better finite-sample behavior than the original maximum likelihood estimator since $E(\widetilde{\theta} - \theta) = O(n^{-2})$. It is, nonetheless, important to note that the bias of the maximum likelihood estimator and, hence, its bias correction are not invariant with respect to reparameterizations of the model.

The results presented here, as pointed out earlier, can be easily extended to a more general model where p_i and q_i are described by different sets of explanatory variables. That is, suppose p_i and q_i are, for each i, modeled as

$$p_i = g(\beta_1 x_{1i} + \cdots + \beta_m x_{mi}),$$
$$q_i = h(\gamma_1 v_{1i} + \cdots + \gamma_M v_{Mi}),$$

where $\{x_1, \ldots, x_m\}$ and $\{v_1, \ldots, v_M\}$ are the sets of explanatory variables for the p and q parameters, respectively, of the beta distribution. Consider the $n \times m$ matrix X with the values of $x_1, \ldots, x_m$ and the $n \times M$ matrix V with the values of $v_1, \ldots, v_M$. It is not difficult to see that our previous expression for the second order bias is still valid provided that the matrix $\widetilde{X}$ is redefined as the following $2n \times (m + M)$ block-diagonal matrix:

$$\begin{pmatrix} X & 0 \\ 0 & V \end{pmatrix}.$$

In short, the extension of the proposed model to handle different sets of covariates in the two regression structures is straightforward.

Finally, it is also of importance to obtain the second order bias of the maximum likelihood estimator of μ_i, $i = 1, \ldots, n$. It can be shown that

$$(p_i + q_i)^2 B(\widehat{\mu}_i) = q_i g_i' \mathbf{x}_i' B(\widehat{\beta}) - p_i h_i' \mathbf{v}_i' B(\widehat{\gamma}) + \Big\{(1-\mu_i) g_i'^2 - \frac{q_i}{2} g_i''\Big\} \mathbf{x}_i' K^{\beta\beta} \mathbf{x}_i$$
$$+ (q_i - p_i) g_i' h_i' \mathbf{x}_i' K^{\beta\gamma} \mathbf{v}_i + \Big\{\frac{p_i}{2} h_i'' - \mu_i h_i'^2\Big\} \mathbf{v}_i' K^{\gamma\gamma} \mathbf{v}_i,$$

where $\mathbf{v}_i$ is the ith row of matrix V.

4 Numerical results

This section presents Monte Carlo simulation results for the model where the response follows a $\mathcal{B}(p_i, q_i)$ distribution, with

$$p_i = \exp(\beta_1 + \beta_2 x_i),$$
$$q_i = \exp(\gamma_1 + \gamma_2 x_i).$$

Hence, $m = 4$, i.e., the model is defined by four regression parameters. The sample sizes used in the experiment were $n = 20, 40, 60$, and the values of x were selected as random draws of a standard normal distribution; these values were kept constant throughout the experiment. The true values of the four regression parameters were set equal to one, and the results were obtained using 100,000 Monte Carlo replications. The pseudo-random number generator employed was George Marsaglia's multiply-with-carry random number generator (Marsaglia, 1997); it has period approximately equal to 2^{60} and passes stringent randomness tests. Log-likelihood maximizations were carried out using the quasi-Newton method known as BFGS, due to Broyden, Fletcher, Goldfarb and Shanno; see Press et al. (1992, §10.7) for details. See also Mittelhammer, Judge and Miller (2000, §8.13); according to them, "[t]he BFGS algorithm is generally regarded as the best performing method" (p. 199). All simulations were performed using the matrix programming language `Ox` (Cribari–Neto and Zarkos, 2003; Doornik, 2001).

Table 1 reports the estimated relative biases of the maximum likelihood ('MLE') and bias-corrected maximum likelihood ('BC') estimators of the regression parameters $\beta_1, \beta_2, \gamma_1, \gamma_2$. The relative bias of an estimator $\widehat{\alpha}$ of a scalar parameter α is defined as $\{E(\widehat{\alpha}) - \alpha\}/\alpha$. The estimated relative bias is obtained by replacing $E(\widehat{\alpha})$ by a Monte Carlo estimate. The root mean squared errors for the unmodified and corrected estimators are also presented. (Hats denote unmodified estimators and tildes denote corrected estimators.) The figures in Table 1 reveal that the maximum likelihood estimators of the beta regression parameters can be substantially biased when the sample size is small, and that the bias correction we derived in the previous section is very effective. For instance, when $n = 20$ the biases of $\beta_1, \beta_2, \gamma_1, \gamma_2$ average 0.1825 whereas the biases of the four corresponding bias-adjusted estimators average 0.0390; that is, the average bias of the MLEs is almost five times greater than that of the corrected estimators. It is also noteworthy that the root mean squared errors of the corrected estimators are also smaller than the corresponding root mean squared errors of the MLEs.

Table 2 displays simulation results for the situation where the parameter values are set at $\beta_1 = \beta_2 = 0.75$ and $\gamma_1 = \gamma_2 = -0.50$. We note that the bias-adjusted estimator again displays smaller bias than the standard maximum likelihood estimator. In particular, the maximum likelihood estimators of β_1 and γ_1 display substantial bias, and the bias correction proves to be quite effective when applied to these estimators. For instance, when $n = 20$, the estimated relative bias of $\widehat{\beta}_1$ equals 32% whereas the corresponding measure for the adjusted estimator of β_1 equals 3.64%; that is, the estimated relative bias for the corrected estimator is nearly 9 times smaller than that of the maximum likelihood estimator. The root mean squared errors of the bias-corrected estimators are also smaller than those of the unmodified maximum likelihood estimators.

Next, we compute an aggregate bias measure for each estimator and each sample size. To that end, we report the square root of the squared estimated relative bias for the eight estimates considered (four for each simulation design). The aggregate bias of each estimator is presented in Table 3 for $n = 20, 40, 60$. We see from the figures in Table 3 that the bias-corrected estimator displays aggregate biases for the three sample sizes that are quite small. Indeed, for samples of 40 and 60 observations the aggregate biases of the adjusted estimator are negligible. The maximum likelihood estimator, on the other hand, displays large aggregate biases. For instance, for $n = 40$ the aggregate bias for the maximum likelihood estimator is almost 15 times larger than that of the improved estimator derived in the previous section.

Table 1 *Simulation results,* $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 1$.

n	estimator	rel. bias	RMSE
20	$\widehat{\beta}_1$	0.2067	0.4016
	$\widehat{\beta}_2$	0.1582	0.5143
	$\widehat{\gamma}_1$	0.2065	0.4021
	$\widehat{\gamma}_2$	0.1585	0.5141
	$\widetilde{\beta}_1$	0.0293	0.3428
	$\widetilde{\beta}_2$	0.0485	0.4694
	$\widetilde{\gamma}_1$	0.0292	0.3435
	$\widetilde{\gamma}_2$	0.0489	0.4691
40	$\widehat{\beta}_1$	0.1077	0.2652
	$\widehat{\beta}_2$	0.0338	0.2384
	$\widehat{\gamma}_1$	0.1075	0.2653
	$\widehat{\gamma}_2$	0.0343	0.2383
	$\widetilde{\beta}_1$	0.0070	0.2401
	$\widetilde{\beta}_2$	0.0053	0.2311
	$\widetilde{\gamma}_1$	0.0068	0.2402
	$\widetilde{\gamma}_2$	0.0058	0.2309
60	$\widehat{\beta}_1$	0.0677	0.1951
	$\widehat{\beta}_2$	0.0275	0.1885
	$\widehat{\gamma}_1$	0.0677	0.1948
	$\widehat{\gamma}_2$	0.0274	0.1884
	$\widetilde{\beta}_1$	0.0034	0.1822
	$\widetilde{\beta}_2$	0.0041	0.1841
	$\widetilde{\gamma}_1$	0.0034	0.1818
	$\widetilde{\gamma}_2$	0.0040	0.1840

Table 2 *Simulation results,* $\beta_1 = \beta_2 = 0.75$, $\gamma_1 = \gamma_2 = -0.5$.

n	estimator	rel. bias	RMSE
20	$\widehat{\beta}_1$	0.3200	0.5074
	$\widehat{\beta}_2$	0.0643	0.3730
	$\widehat{\gamma}_1$	−0.3678	0.3753
	$\widehat{\gamma}_2$	−0.0040	0.3001
	$\widetilde{\beta}_1$	0.0364	0.4464
	$\widetilde{\beta}_2$	0.0084	0.3602
	$\widetilde{\gamma}_1$	−0.0388	0.3176
	$\widetilde{\gamma}_2$	0.0008	0.2855
40	$\widehat{\beta}_1$	0.1434	0.2999
	$\widehat{\beta}_2$	0.0132	0.2183
	$\widehat{\gamma}_1$	−0.1655	0.2237
	$\widehat{\gamma}_2$	0.0211	0.1832
	$\widetilde{\beta}_1$	0.0097	0.2794
	$\widetilde{\beta}_2$	0.0019	0.2147
	$\widetilde{\gamma}_1$	−0.0093	0.2050
	$\widetilde{\gamma}_2$	0.0009	0.1789
60	$\widehat{\beta}_1$	0.0935	0.2372
	$\widehat{\beta}_2$	0.0101	0.1930
	$\widehat{\gamma}_1$	−0.1085	0.1767
	$\widehat{\gamma}_2$	0.0086	0.1592
	$\widetilde{\beta}_1$	0.0040	0.2262
	$\widetilde{\beta}_2$	0.0004	0.1908
	$\widetilde{\gamma}_1$	−0.0049	0.1665
	$\widetilde{\gamma}_2$	−0.0008	0.1566

Table 3 *Aggregate biases.*

n	MLE	BC
20	0.6143	0.0967
40	0.2721	0.0185
60	0.1771	0.0098

5 Concluding remarks

The purpose of the present paper was twofold. First, we proposed a regression structure for modeling random variables that are restricted to the $(0,1)$ interval. To that end, the response was assumed to follow a beta distribution where the two parameters that index such distribution (p and q) are linked to regression equations involving covariates and regression parameters. We have shown that when q is regressor-free and the link function for p is exponential, the regression parameters can be interpreted in terms of mean response effects. The analysis, however, is developed under a more general setting. Maximum likelihood estimation of the unknown parameters was discussed. The estimators do not have closed-form expressions but can be computed by numerically maximizing the log-likelihood function. Second, we derived a bias-adjustment scheme that nearly eliminates the bias of the maximum likelihood estimator in small samples. The simulation results presented showed that the bias correction we derived is very effective, even when the sample size is not large. Indeed, the bias correction mechanism proposed in this paper yields modified maximum likelihood estimators that are nearly unbiased.

Appendix

After some algebra, we obtain the following cumulants for the proposed class of beta regression models:

$$\begin{aligned}
\kappa_{rtu} &= 3\sum_{i=1}^{n} g_i' g_i'' x_{ri} x_{ti} x_{ui} [\psi'(p_i+q_i) - \psi'(p_i)] \\
&\quad + \sum_{i=1}^{n} g_i'^3 x_{ri} x_{ti} x_{ui} [\psi''(p_i+q_i) - \psi''(p_i)], \\
\kappa_{rtU} &= \sum_{i=1}^{n} h_i' x_{ri} x_{ti} x_{Ui} [g_i'' \psi'(p_i+q_i) + g_i'^2 \psi''(p_i+q_i)], \\
\kappa_{rTU} &= \sum_{i=1}^{n} g_i' x_{ri} x_{Ti} x_{Ui} [h_i'' \psi'(p_i+q_i) + h_i'^2 \psi''(p_i+q_i)], \\
\kappa_{RTU} &= 3\sum_{i=1}^{n} h_i' h_i'' x_{Ri} x_{Ti} x_{Ui} [\psi'(p_i+q_i) - \psi'(q_i)] \\
&\quad + \sum_{i=1}^{n} h_i'^3 x_{Ri} x_{Ti} x_{Ui} [\psi''(p_i+q_i) - \psi''(q_i)].
\end{aligned}$$

Using the results above we then obtain the following:

$$\kappa_{rt}^{(u)} - \frac{1}{2}\kappa_{rtu} = \frac{1}{2}\sum_{i=1}^{n} g_i' g_i'' x_{ri} x_{ti} x_{ui} [\psi'(p_i+q_i) - \psi'(p_i)] + \frac{1}{2}\sum_{i=1}^{n} g_i'^3 x_{ri} x_{ti} x_{ui} [\psi''(p_i+q_i) - \psi''(p_i)],$$

$$\kappa_{Rt}^{(u)} - \frac{1}{2}\kappa_{Rtu} = \frac{1}{2}\sum_{i=1}^{n} h_i' x_{Ri} x_{ti} x_{ui} [g_i''\psi'(p_i+q_i) + g_i'^2\psi''(p_i+q_i)],$$

$$\kappa_{rT}^{(u)} - \frac{1}{2}\kappa_{rTu} = \frac{1}{2}\sum_{i=1}^{n} h_i' x_{ri} x_{Ti} x_{ui} [g_i''\psi'(p_i+q_i) + g_i'^2\psi''(p_i+q_i)],$$

$$\kappa_{rt}^{(U)} - \frac{1}{2}\kappa_{rtU} = \frac{1}{2}\sum_{i=1}^{n} h_i' x_{ri} x_{ti} x_{Ui} [g_i'^2\psi''(p_i+q_i) - g_i''\psi'(p_i+q_i)],$$

$$\kappa_{RT}^{(u)} - \frac{1}{2}\kappa_{RTu} = \frac{1}{2}\sum_{i=1}^{n} g_i' x_{Ri} x_{Ti} x_{ui} [h_i'^2\psi''(p_i+q_i) - h_i''\psi'(p_i+q_i)],$$

$$\kappa_{Rt}^{(U)} - \frac{1}{2}\kappa_{RtU} = \frac{1}{2}\sum_{i=1}^{n} g_i' x_{Ri} x_{ti} x_{Ui} [h_i'^2\psi''(p_i+q_i) + h_i''\psi'(p_i+q_i)],$$

$$\kappa_{rT}^{(U)} - \frac{1}{2}\kappa_{rTU} = \frac{1}{2}\sum_{i=1}^{n} g_i' x_{ri} x_{Ti} x_{Ui} [h_i'^2\psi''(p_i+q_i) + h_i''\psi'(p_i+q_i)],$$

$$\kappa_{RT}^{(U)} - \frac{1}{2}\kappa_{RTU} = \frac{1}{2}\sum_{i=1}^{n} h_i' h_i'' x_{Ri} x_{Ti} x_{Ui} [\psi'(p_i+q_i) - \psi'(q_i)] + \frac{1}{2}\sum_{i=1}^{n} h_i'^3 x_{Ri} x_{Ti} x_{Ui} [\psi''(p_i+q_i) - \psi''(q_i)].$$

Next, we define the following matrices:

$$\begin{aligned}
W_1 &= \mathrm{diag}\{g_i' g_i''[\psi'(p_i+q_i) - \psi'(p_i)] + g_i'^3[\psi''(p_i+q_i) - \psi''(p_i)]\},\\
W_2 &= \mathrm{diag}\{h_i'[g_i''\psi'(p_i+q_i) + g_i'^2\psi''(p_i+q_i)]\},\\
W_3 &= \mathrm{diag}\{h_i'[g_i'^2\psi''(p_i+q_i) - g_i''\psi'(p_i+q_i)]\},\\
W_4 &= \mathrm{diag}\{g_i'[h_i'^2\psi''(p_i+q_i) - h_i''\psi'(p_i+q_i)]\},\\
W_5 &= \mathrm{diag}\{g_i'[h_i'^2\psi''(p_i+q_i) + h_i''\psi'(p_i+q_i)]\},\\
W_6 &= \mathrm{diag}\{h_i' h_i''[\psi'(p_i+q_i) - \psi'(q_i)] + h_i'^3[\psi''(p_i+q_i) - \psi''(q_i)]\}.
\end{aligned}$$

We can now obtain a closed-form expression for the second order bias. For instance, the first term in Cox and Snell expansion is

$$\sum_{r,t,u} \kappa^{sr}\kappa^{tu}\left\{\kappa_{rt}^{(u)} - \frac{1}{2}\kappa_{rtu}\right\} = \frac{1}{2}\sum_{r,t,u}\kappa^{sr}\kappa^{tu}\sum_{i=1}^{n} w_{1i}x_{ri}x_{ti}x_{ui},$$

where w_{1i} denotes the ith diagonal element of W_1. Hence,

$$\begin{aligned}\sum_{r,t,u} \kappa^{sr}\kappa^{tu}\left\{\kappa_{rt}^{(u)} - \frac{1}{2}\kappa_{rtu}\right\} &= \frac{1}{2}\sum_{i=1}^{n} w_{1i}\sum_{r}\kappa^{sr}x_{ri}\sum_{t,u}x_{ti}\kappa^{tu}x_{ui}\\ &= -\frac{1}{2}\sum_{i=1}^{n} w_{1i}\sum_{r}\kappa^{sr}x_{ri}\mathbf{x}_i K^{\beta\beta}\mathbf{x}_i^T,\end{aligned}$$

where $\mathbf{x}_i$ is the row vector defining the ith row of X and $K^{\beta\beta}$ is the upper left $m \times m$ block of K_θ^{-1}. We have

$$\begin{aligned}\sum_{r,t,u} \kappa^{sr}\kappa^{tu}\left\{\kappa_{rt}^{(u)} - \frac{1}{2}\kappa_{rtu}\right\} &= -\frac{1}{2}\sum_{i=1}^{n} w_{1i}(\mathbf{x}_i K^{\beta\beta}\mathbf{x}_i^T)\sum_{r}\kappa^{sr}x_{ri}\\ &= \frac{1}{2}\sum_{i=1}^{n} w_{1i}(\mathbf{x}_i K^{\beta\beta}\mathbf{x}_i^T)\rho_s^T K^{\beta\beta}\mathbf{x}_i^T,\end{aligned}$$

where ρ_s stands for the sth column of the $m \times m$ identity matrix. Therefore,

$$\sum_{r,t,u} \kappa^{sr}\kappa^{tu}\left\{\kappa_{rt}^{(u)} - \frac{1}{2}\kappa_{rtu}\right\} = \frac{1}{2}\rho_s^T K^{\beta\beta}\sum_{i=1}^{n} w_{1i}(\mathbf{x}_i K^{\beta\beta}\mathbf{x}_i^T)\mathbf{x}_i^T.$$

Let $\delta_{\beta\beta}$ be the $n\times 1$ vector consisting of the diagonal elements of $XK^{\beta\beta}X^T$. Then, it can be easily seen that the previous expression can be written as

$$\sum_{r,t,u} \kappa^{sr}\kappa^{tu}\left\{\kappa_{rt}^{(u)} - \frac{1}{2}\kappa_{rtu}\right\} = \frac{1}{2}\rho_s^T K^{\beta\beta}X^T W_1\delta_{\beta\beta}.$$

Similarly, we can define the $n\times 1$ vectors $\delta_{\beta\gamma}$ and $\delta_{\gamma\gamma}$ using the blocks $K^{\beta\gamma}$ and $K^{\gamma\gamma}$, respectively, of K^{-1}. We obtain, for the remaining terms in the expansion that

$$\sum_{R,t,u} \kappa^{sR}\kappa^{tu}\left\{\kappa_{Rt}^{(u)} - \frac{1}{2}\kappa_{Rtu}\right\} = \frac{1}{2}\rho_s^T K^{\beta\gamma}X^T W_2\delta_{\beta\beta},$$

$$\sum_{r,T,u} \kappa^{sr}\kappa^{Tu}\left\{\kappa_{rT}^{(u)} - \frac{1}{2}\kappa_{rTu}\right\} = \frac{1}{2}\rho_s^T K^{\beta\beta}X^T W_2\delta_{\beta\gamma},$$

$$\sum_{r,t,U} \kappa^{sr}\kappa^{tU}\left\{\kappa_{rt}^{(U)} - \frac{1}{2}\kappa_{rtU}\right\} = \frac{1}{2}\rho_s^T K^{\beta\beta}X^T W_3\delta_{\beta\gamma},$$

$$\sum_{R,T,u} \kappa^{sR}\kappa^{Tu}\left\{\kappa_{RT}^{(u)} - \frac{1}{2}\kappa_{RTu}\right\} = \frac{1}{2}\rho_s^T K^{\beta\gamma} X^T W_4 \delta_{\beta\gamma},$$

$$\sum_{R,t,U} \kappa^{sR}\kappa^{tU}\left\{\kappa_{Rt}^{(U)} - \frac{1}{2}\kappa_{RtU}\right\} = \frac{1}{2}\rho_s^T K^{\beta\gamma} X^T W_5 \delta_{\beta\gamma},$$

$$\sum_{r,T,U} \kappa^{sr}\kappa^{TU}\left\{\kappa_{rT}^{(U)} - \frac{1}{2}\kappa_{rTU}\right\} = \frac{1}{2}\rho_s^T K^{\beta\beta} X^T W_5 \delta_{\gamma\gamma},$$

$$\sum_{R,T,U} \kappa^{sR}\kappa^{TU}\left\{\kappa_{RT}^{(U)} - \frac{1}{2}\kappa_{RTU}\right\} = \frac{1}{2}\rho_s^T K^{\beta\gamma} X^T W_6 \delta_{\gamma\gamma}.$$

Therefore, the second order bias for the maximum likelihood estimator of β is as given in Section 3.

Acknowledgements

We gratefully acknowledge partial financial support from CNPq. We also thank two anonymous referees for comments and suggestions.

(Received April, 2004. Accepted November, 2004.)

References

Box, M. (1971). Bias in nonlinear estimation (with discussion). *Journal of the Royal Statistical Society B*, **33**, 171–201.

Bury, K. (1999). *Statistical Distributions in Engineering*. New York: Cambridge University Press.

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

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

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

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.

Cordeiro, G. M. and Vasconcellos, K. L. P. (1999). Second-order biases of the maximum likelihood estimates in von Mises regression models. *Australian and New Zealand Journal of Statistics*, **41**, 901–910.

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

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

Cribari–Neto, F. and Vasconcellos, K. L. P. (2002). Nearly unbiased maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, **72**, 107–118.

Cribari–Neto, F. and Zarkos, S. G. (2003). Econometric and statistical computing using Ox. *Computational Economics*, **21**, 277–295.

Doornik, J. A. (2001). *Ox: an Object-oriented Matrix Programming Language*, 4th ed. London: Timberlake Consultants and Oxford: `http: //www.nuff.ox.ac.uk/Users/Doornik`.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, vol.2, 2nd ed. New York: Wiley.

Krysicki, W. (1999). On some new properties of the beta distribution. *Statistics and Probability Letters*, **42**, 131–137.

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

Marsaglia, G. (1997). A random number generator for C. Posting to the usenet newsgroup `sci.stat.math`.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zabembka, F. (ed.), *Frontiers in Econometrics*. New York: Academic Press.

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

Pike, M., Hill, A. and Smith, P. (1980). Bias and efficiency in logisitc analysis of stratified case-control studies. *International Journal of Epidemiology*, **9**, 89–95.

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

Vasconcellos, K. L. P. and Cordeiro, G. M. (2000). Bias corrected estimates in multivariate Student t regression models. *Communications in Statistics, Theory and Methods*, **29**, 797–822.

Vasconcellos, K. L. P., Cordeiro, G. M. and Barroso, L. P. (2000). Improved estimation for robust econometric regression models. *Brazilian Journal of Probability and Statistics*, **14**, 141–157.

Young, D. and Bakir, S. (1987). Bias correction for a generalized log-gamma regression model. *Technometrics*, **29**, 183–191.

Klaus L.P. Vasconcellos and **Francisco Cribari–Neto**
Departamento de Estatística
Universidade Federal de Pernambuco
Cidade Universitária
50740–540, Recife/PE, Brazil
E-mails: klaus@de.ufpe.br and cribari@de.ufpe.br