SEMIPARAMETRIC ADDITIVE BETA REGRESSION MODELS: INFERENCE AND LOCAL INFLUENCE DIAGNOSTICS

Authors: GERMÁN IBACACHE-PULGAR
– Institute of Statistics, Universidad de Valparaíso, Chile

JORGE FIGUEROA-ZUÑIGA
– Department of Statistics, Universidad de Concepción, Concepción, Chile

CAROLINA MARCHANT
– Faculty of Basic Sciences, Universidad Católica del Maule, Talca, Chile

Abstract:

- In this paper, we study a semiparametric additive beta regression model using a parameterization based on the mean and a dispersion parameter. This model is useful for situations where the response variable is continuous and restricted to the unit interval, in addition to being related to other variables through a semiparametric regression structure. First, we formulate the model and then estimation of its parameters is discussed. A back-fitting algorithm is derived to attain the maximum penalized likelihood estimates by using natural cubic smoothing splines. We provide closed-form expressions for the score function, Fisher information matrix and its inverse. Local influence methods are derived as diagnostic tools. Finally, a practical illustration based real data is presented and discussed.

Key-Words:

- *Beta distribution; Diagnostic techniques; Maximum penalized likelihood estimates; Penalized likelihood function; Semiparametric additive models.*

AMS Subject Classification:

- 62D05; 62F99; 62J99;

1. INTRODUCTION

Since the beta regression (BR) model was introduced by Ferrari and Cribari-Neto (2004), it has become an excellent tool for modeling continuous data in the unit interval. This is mainly because of the flexibility of its probability density function (PDF), which can cover a wide range of shapes (symmetric, asymmetric, unimodal and bimodal) depending on different values of its parameters. In addition, the beta distribution can be parameterized in terms of its mean; see more details in Cribari-Neto and Zeileis (2010), Figueroa-Zuñiga et al. (2013), Zhao et al. (2014), Queiroz da-Silva and Migon (2016) and Huerta et al. (2018).

Inclusion of nonparametric functions enhances the modeling flexibility for accommodating nonlinear effects of covariates. Semiparametric models have been successfully used to describe nonlinear components. Semiparametric additive beta regression (SABR) models emerge as a useful tool to describe situations where the response variable is continuous, restricted to the unit interval and related to covariates through a semiparametric regression structure (Zhu and Lee, 2003).

Note that parameter estimation in BR models, and consequently in SABR models, can be influenced by outlying observations. For this reason, diagnostic is a fundamental stage in the modeling of data. Diagnostic techniques used in a regression model can be divided into global influence (elimination of cases) and local influence. The main idea of the local influence technique, proposed by Cook (1986), is to evaluate the sensitivity of the parameter estimators when small perturbations are introduced in the assumptions of the model or in the data (for example, in the response and covariates). This technique has the advantage, with respect to elimination of cases, that does not needs to calculate the parameters estimates for each case eliminated. The following works are related to the local influence technique: Zhu et al. (2003) considered it in generalized linear mixed models; Zhu et al. (2003) and Ibacache-Pulgar and Paula (2011) provided local influence measures to evaluate the sensitivity of the maximum penalized likelihood (MPL) estimates in normal and Student-t partially linear models, respectively; Osorio et al. (2007) derived it in elliptical linear models for longitudinal data; Cao and Lin (2011) applied it to elliptical linear mixed models with first-order autoregressive errors; Ibacache-Pulgar et al. (2012, 2013) analyzed it in elliptical semiparamteric mixed and symmetric semiparametric additive models, respectively; Uribe-Opazo et al. (2012) and Garcia-Papani et al. (2018) used it to evaluate sensitivity in spatial models; Zhang et al. (2015) and Ibacache-Pulgar and Reyes (2018) developed it for normal and elliptical partially varying-coefficient models, respectively; Emami (2017) utilized it in Liu penalized least squares estimators; Marchant et al. (2016) considered it in multivariate regression models; Ferreira and Paula (2017) extended it for different perturbation schemes considering a skew-normal partially linear model; Leao et al. (2018) derived it in cure rate models with frailties; Cysneiros et al. (2019) implemented it in Cobb-Douglas type models; and Tapia et al. (2019a,b) applied it to mixed effects logistic and longitudinal count regression models. In the case of BR models, Espinheria et al. (2008) derived the local influence technique under different perturbation schemes; Ferrari (2011) derived it in BR models with varying dispersion; and Rocha and Simas (2011) applied it to a general class of the BR models. Note that the SABR model is a particular case of the GAMLSSS models (Stasinopoulos and Rigby, 2007). To the best of our knowledge, local influence diagnostics in SABR models have been no analyzed to the date. Therefore, the aim of this paper is to study the parameters estimation and to apply the approach of local influence in the SABR model.

This paper is organized as follows. In Section 2, the SABR model is presented and a penalized log-likelihood function is considered for the parameters estimation. In this section, we present a weighted back-fitting algorithm to obtain MPL estimates and parameters smoothing selection. Section 3 discusses and derives the local influence curvatures and 4 illustrates the proposed methodology with a real data set. Finally, in Section 5 some concluding remarks are mentioned.

2. THE SEMIPARAMETRIC ADDITIVE BETA REGRESSION MODEL

2.1. Formulation

Let $Y_1, \ldots, Y_n$ be independent random variables following a beta distribution, where each Y_i has a PDF given by

$$f_Y(y_i; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu_i\phi)\Gamma([1-\mu_i]\phi)} y_i^{\mu_i\phi-1}[1-y_i]^{[1-\mu_i]\phi-1}, \quad i = 1, \ldots, n, \tag{2.1}$$

with $0 < y_i < 1$, $0 < \mu_i < 1$ and $\phi > 0$. Then, the mean of Y_i can be written as

$$g(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \tag{2.2}$$

where $g(\cdot)$ is a strictly monotonic and twice differentiable link function that maps $(0,1)$ into real numbers set; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of unknown regression parameters; and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is a vector of observed covariates $(p < n)$.

The SABR models are often used in research related to longitudinal, clustered and spatial sampling schemes. The mean of this model can be obtained from (2.2) as

$$g(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + f_1(t_{1_i}) + \cdots + f_s(t_{s_i}), \tag{2.3}$$

or alternatively as $g(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{n}_{1_i}^\top \mathrm{f}_1 + \cdots + \boldsymbol{n}_{s_i}^\top \mathrm{f}_s$, where $g(\cdot)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ are such as in (2.2), but now we add the nonparametric structure by f_ks, which are unknown smooth arbitrary functions on covariates t'_ks, for $k = 1, \ldots, s$; where $\boldsymbol{n}_{k_i}^\top$ denotes the ith row of the incidence matrix $\boldsymbol{N}_k$ whose (i, l)th element corresponds to the indicator function $I(t_{k_i} = t_{k_l}^0)$, with $t_{k_l}^0$, for $l = 1, \ldots, r_k$, denoting the distinct and ordered values of the covariate t_k and $\mathrm{f}_k = (f_k(t_{k_1}), \cdots, f_k(t_{k_{r_k}}))^\top$. There are several possible choices for the link function $g(\cdot)$. For instance, one can use the logit specification $g(\mu) = \log\{\mu/(1-\mu)\}$, the probit function $g(\mu) = \Phi^{-1}(\mu)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function, the complementary log-log link $g(\mu) = \log\{-\log(1-\mu)\}$, and the log-log link

$g(\mu) = -\log\{-\log(\mu)\}$, among others. A particularly useful link function is the logit link, in which case we write

$$\mu_i = \frac{\exp\left[\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{k=1}^{s} \boldsymbol{n}_{k_i}^\top \mathrm{f}_k\right]}{1 + \exp\left[\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{k=1}^{s} \boldsymbol{n}_{k_i}^\top \mathrm{f}_k\right]}.$$

Since the functions f_ks belong to the infinite dimensional space and are considered parameters with respect to the expected value of Y_i, some restricted subspace should be defined for the nonparametric functions to ensure identifiability of the parameters associated with the model. Therefore, we assume that the function f_k belongs to the Sobolev function space (Adams and Fournier, 2003) defined as $\mathcal{W}_2^{(2)} = \{f_k : f_k, f_k^{(1)}\text{abs. cont.}, f_k^{(2)} \in \mathcal{L}^2[a_k, b_k]\}$, where $f_k^{(2)}(t_k) = \partial^2/\partial t_k^2 f_k(t_k)$, with $t_k^0 \in [a_k, b_k]$. Then, the log-likelihood function of the model defined in (2.1) and (2.3) is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{2.4}$$

where

$$\ell_i(\boldsymbol{\theta}) = \log(\Gamma(\phi)) - \log(\Gamma(\mu_i\phi)) - \log(\Gamma([1-\mu_i]\phi)) + [\mu_i\phi - 1]\log(y_i) + \{[1-\mu_i]\phi - 1\}\log[1-y_i],$$

with μ_i defined in (2.2) and $\boldsymbol{\theta} = (\beta^\top, \mathrm{f}_\mathbf{1}^\top, \ldots, \mathrm{f}_\mathbf{s}^\top, \phi)^\top \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{\mathbf{p}^*}$, with $p^* = p + r + 1$ and $r = \sum_{k=1}^{s} r_k$. Incorporating a penalty function over each function f_k, we have that the penalized log-likelihood function can be expressed as

$$\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \ell(\boldsymbol{\theta}) - \sum_{k=1}^{s} \frac{\alpha_k}{2}\, \mathrm{f}_k^\top \boldsymbol{K}_k\, \mathrm{f}_k, \tag{2.5}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_s)^\top$ denotes an $(s \times 1)$ vector of smoothing parameters and $\boldsymbol{K}_k$ is a $(r_k \times r_k)$ nonnegative definite matrix that depends only on the knots $t_{k_l}^0$, for $l = 1, \ldots, r_k$. Details about the construction of this matrix can be found in Green and Silverman (1994). Note that direct maximization of the log-likelihood function, without imposing restrictions on smooth functions, can generate problems of identifiability or over-fitting. To correct these problems, it is suggested to incorporate a penalty term for each smooth function in the log-likelihood function. Then, the MPL estimates are obtained by maximizing this function. As the resulting estimation equations are non-linear, an iterative process is required to obtain the parameter estimates. Therefore, in the analysis of local influence presented in Section 3, the penalized maximum likelihood estimate is replaced by an estimate obtained in the last iteration of the process, after reaching convergence.

2.2. Estimation

In order to define the penalized score functions, consider $\boldsymbol{X}$ being an $n \times p$ matrix whose ith row is $\boldsymbol{x}_i^\top$, $\boldsymbol{N}_k$ being an $n \times r_k$ matrix which ith row is $\boldsymbol{n}_{k_i}^\top$, $\boldsymbol{T} = \mathrm{diag}(1/g'(\mu_1), \ldots, 1/g'(\mu_n))$,

$yn^* = (y_1^*, \ldots, y_n^*)^\top$, $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_n^*)^\top$, $y_i^* = \log(y_i/[1-y_i])$ and $\mu_i^* = \psi(\mu_i\phi) - \psi[(1-\mu_i)\phi]$, for $i = 1, \ldots, n$, with $\psi(\cdot)$ denoting the digamma function, this is, $\psi(z) = \mathrm{d}\log\Gamma(z)/\mathrm{d}z$, for $z > 0$. Then, assuming that (2.5) is regular with respect to $\boldsymbol{\beta}$, $\mathrm{f}_1, \ldots, \mathrm{f}_s$ and ϕ, the penalized score function of $\boldsymbol{\theta}$ is defined as

$$\boldsymbol{U}_{\mathrm{p}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\partial \ell_{\mathrm{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta}}.$$

After some algebraic manipulations we have in matrix form that

$$\begin{aligned}
\frac{\partial \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} &= \phi \boldsymbol{X}^\top \boldsymbol{T}[\boldsymbol{y}_i^* - \boldsymbol{\mu}^*], \\
\frac{\partial \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial\, \mathrm{f}_k} &= \phi \boldsymbol{N}_k^\top \boldsymbol{T}[\boldsymbol{y}_i^* - \boldsymbol{\mu}^*] - \alpha_k \boldsymbol{K}_k\, \mathrm{f}_k, k = 1, \ldots, s, \\
\frac{\partial \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \phi} &= \sum_{i=1}^{n} \Big\{ \mu_i [y_i^* - \mu_i^*] + \log(1-y_i) - \psi[(1-\mu_i)\phi] + \psi(\phi) \Big\}.
\end{aligned}$$

2.3. Weighted back-fitting algorithm

To estimate $\boldsymbol{\theta}$ by the MPL method, we have to solve the $\mathrm{U}_{\mathrm{p}}(\boldsymbol{\theta}) = \mathbf{0}$. However, the estimating equations are nonlinear and require an iterative method. For example, the determination of the MPL estimates $\widehat{\boldsymbol{\theta}}$ can be performed by using the Fisher scoring algorithm. Let $f_0 = \boldsymbol{\beta}$ and $\boldsymbol{N}_0 = \boldsymbol{X}$, and consider for simplicity $\boldsymbol{\alpha}$ and $\boldsymbol{W}$ fixed, with $\boldsymbol{W}$ defined in the appendix. Then, the Fisher scoring algorithm is given by (see Ibacache-Pulgar et al., 2013)

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{S}_0^{(u)} \boldsymbol{N}_1 & \ldots & \boldsymbol{S}_0^{(u)} \boldsymbol{N}_s \\ \boldsymbol{S}_1^{(u)} \boldsymbol{N}_0 & \boldsymbol{I} & \ldots & \boldsymbol{S}_1^{(u)} \boldsymbol{N}_s \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{S}_s^{(u)} \boldsymbol{N}_0 & \boldsymbol{S}_s^{(u)} \boldsymbol{N}_1 & \ldots & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \mathrm{f}_0^{(u+1)} \\ \mathrm{f}_1^{(u+1)} \\ \vdots \\ \mathrm{f}_s^{(u+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_0^{(u)} \boldsymbol{z}^{(u)} \\ \boldsymbol{S}_1^{(u)} \boldsymbol{z}^{(u)} \\ \vdots \\ \boldsymbol{S}_s^{(u)} \boldsymbol{z}^{(u)} \end{pmatrix}, \tag{2.6}$$

where $\boldsymbol{z}^{(u)} = \boldsymbol{\eta} + \boldsymbol{W}^{-1}\boldsymbol{T}[yn^* - \boldsymbol{\mu}^*]\big|_{\boldsymbol{\theta}^{(u)}}$ with $\boldsymbol{\eta} = (g(\mu_1), \ldots, g(\mu_n))^\top$ and $g(\cdot)$ as given in (**??**), and $\boldsymbol{S}_k^{(u)} = \boldsymbol{S}_k\big|_{\boldsymbol{\theta}^{(u)}}$, with

$$\boldsymbol{S}_k = \begin{cases} (\boldsymbol{N}_0^\top \boldsymbol{W} \boldsymbol{N}_0)^{-1} \boldsymbol{N}_0^\top \boldsymbol{W}, & k = 0, \\ (\boldsymbol{N}_k^\top \boldsymbol{W} \boldsymbol{N}_k + \alpha_k \boldsymbol{K}_k)^{-1} \boldsymbol{N}_k^\top \boldsymbol{W}, & k = 1, \ldots, s. \end{cases}$$

As it is known, the back-fitting algorithm is a simple iterative procedure used to fit a generalized additive model; see Hastie et al. (2001). Then, in our case, the back-fitting (Gauss-Seidel) iterations (Hastie and Tibshirani, 1990) that are used to solve the equations system (2.6) take the form

$$\mathrm{f}_k^{(u+1)} = \boldsymbol{S}_k^{(u)} \Big[\boldsymbol{z}^{(u)} - \sum_{l=0, l\neq k}^{s} \boldsymbol{N}_l\, \mathrm{f}_l^{(u)} \Big]. \tag{2.7}$$

Note that the system of equations given in (2.6) is consistent and the back-fitting algorithm given in (2.7) converges to a solution for any starting values if the weight matrix involved is symmetric and defined positive; see Berhane and Tibshirani (1998). In addition, we have that this solution is unique under concurvity in the data. In particular, for a model with smooth terms f_1 and f_2 but without the constant terms, $\boldsymbol{\beta}$, we have the following considerations:

(i) If $\|\boldsymbol{S}_1\boldsymbol{S}_2\| < 1$, the estimating equations are consistent and have a unique solution, and the final iterations from the back-fitting algorithm are independent of the starting values and starting order.

(ii) If $\|\boldsymbol{S}_1\boldsymbol{S}_2\| = 1$, this gives an indication of concurvity in the data (strict collinearity), and therefore the back-fitting algorithm converges to one of the solutions of estimating equations system, and the starting functions determine the final solutions.

(iii) If the $\boldsymbol{S}_k$ smoothers are not centered ($\boldsymbol{S}_k^\top\mathbf{1} = \mathbf{1}$), typically $\|\boldsymbol{S}_1\boldsymbol{S}_2\| = 1$, we can consider a centered smoother such that $\boldsymbol{S}_k^\top\mathbf{1} = \mathbf{0}$, with $\mathbf{1}$ denoting a $(r_k \times 1)$ vector of ones, which is defined as

$$\boldsymbol{S}_k = \left(\boldsymbol{I}_{(r_k,r_k)} - \frac{\mathbf{1}\mathbf{1}^\top}{r_k}\right)\left(\boldsymbol{N}_k^\top\boldsymbol{W}\boldsymbol{N}_k + \alpha_k\boldsymbol{K}_k\right)^{-1}\boldsymbol{N}_k^\top\boldsymbol{W}.$$

The MPL estimate of the scale parameter, $\widehat{\phi}$, can be obtained by solving the following iterative process (see Ibacache-Pulgar and Reyes, 2018):

$$\phi^{(u+1)} = \phi^{(u)} - \mathrm{E}\left\{\frac{\partial^2\ell_\mathrm{p}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\phi\partial\phi}\right\}^{-1}\frac{\partial\ell_\mathrm{p}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\phi}\bigg|_{\boldsymbol{\theta}}^{(u)}.$$

The covariance matrix of $\widehat{\boldsymbol{\theta}}$ is obtained from the inverse of the expected information matrix $\boldsymbol{\mathcal{I}}_\mathrm{p}$ defined in the appendix. Therefore, the approximate covariance matrix of $\widehat{\boldsymbol{\theta}}$ is given as $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\theta}}) \approx \boldsymbol{\mathcal{I}}_\mathrm{p}^{-1}\big|_{\widehat{\boldsymbol{\theta}}}$, where

$$\boldsymbol{\mathcal{I}}_\mathrm{p}^{-1} = \begin{pmatrix} \boldsymbol{\mathcal{J}}_1^{-1} & -\boldsymbol{\mathcal{I}}_{11}^{-1}\boldsymbol{\mathcal{I}}_{12}\boldsymbol{\mathcal{J}}_2^{-1} \\ -\boldsymbol{\mathcal{I}}_{22}\boldsymbol{\mathcal{I}}_{21}\boldsymbol{\mathcal{J}}_1^{-1} & \boldsymbol{\mathcal{J}}_2^{-1} \end{pmatrix},$$

with $\boldsymbol{\mathcal{J}}_1 = \boldsymbol{\mathcal{I}}_{11} - \boldsymbol{\mathcal{I}}_{12}\boldsymbol{\mathcal{I}}_{22}^{-1}\boldsymbol{\mathcal{I}}_{21}$ and $\boldsymbol{\mathcal{J}}_2 = \boldsymbol{\mathcal{I}}_{22} - \boldsymbol{\mathcal{I}}_{21}\boldsymbol{\mathcal{I}}_{11}^{-1}\boldsymbol{\mathcal{I}}_{12}$. An approximate pointwise standard error band ($\mathrm{SEB}_{\mathrm{approx}}$) for $f_k(\cdot)$, that allows us to assess how accurate the estimator $\widehat{f_k}(\cdot)$, can be defined as

$$\mathrm{SEB}_{\mathrm{approx}}(f_k(t_l^0)) = \widehat{f_k}(t_l^0) \pm 2\sqrt{\widehat{\mathrm{Var}}(\widehat{f_k}(t_l^0))}, \quad l = 1, \ldots, r,$$

where $\mathrm{Var}(\widehat{f_k}(t_l))$ is the lth principal diagonal element of the corresponding block-diagonal matrix from $\boldsymbol{\mathcal{I}}_\mathrm{p}^{-1}$.

Note that Equation (2.7), which involves a diagonal matrix denoted by $\boldsymbol{W}$, leads to an iterative weighted backfitting solution. The convergence of the iterative process is guaranteed by the diagonal structure of $\boldsymbol{W}$. Note also that this matrix must be updated in each iteration of the backftting iterative process and in each stage of the Fischer scoring algorithm.

Observe that the joint iterative procedure proposed to estimate the parameters of the model is based on the Fisher scoring and backfitting algorithms. First, note that Equation (2.6) corresponds to the matrix equation of the Fisher scoring algorithm. Then, after algebraic manipulations, the solutions to this system correspond precisely to the backfitting iterations. In addition, note also that the scale parameter is estimated by a Fisher scoring algorithm. In summary, the Fisher scoring algorithm allows us to estimate the parameter vector associated with our model and the backfitting algorithm to update the estimates of the parameters associated with the parametric and nonparametric components of the model for each stage of the Fisher score algorithm.

2.4. Degrees of freedom and smoothing parameters

The determination of the parameter vector $\boldsymbol{\alpha}$ is a crucial aspect in the estimation process. Different choice methods are available in the literature for this purpose. For example, an alternative to select smoothing parameters under the SABR model is to consider the Akaike information criterion –AIC– (see details in Ferreira et al, 2012; Ventura et al., 2019) defined by

$$\text{AIC}(\boldsymbol{\alpha}) = -2\ell_{\text{p}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha}) + 2\Big[p + 1 + \text{df}(\boldsymbol{\alpha})\Big],$$

where $\ell_{\text{p}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha})$ denotes the penalized log-likelihood function available at $\widehat{\boldsymbol{\theta}}$ for a fixed $\boldsymbol{\alpha}$ and $\text{df}(\boldsymbol{\alpha}) = \sum_{k=1}^{s} \text{df}(\alpha_k)$ denotes approximately the number of effective parameters involved in modeling of the nonparametric effects. The idea is to minimize the function $\text{AIC}(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$. Following Hastie and Tibshirani (1990) and Eilers and Marx (1996), the degrees of freedom (DF) associated with the kth smooth function are given by

$$\text{df}(\alpha_k) = \text{tr}\big\{\widetilde{\boldsymbol{S}}_k\big\} = \sum_{j=1}^{r_k} \frac{1}{1 + \alpha_k \phi L_j}, \quad j = 1, \ldots, r_k,$$

which measures the individual effect contribution of the kth component, with $\widetilde{\boldsymbol{S}}_k = \boldsymbol{N}_k \boldsymbol{S}_k$ and $\boldsymbol{S}_k$ defined previously, and L_j are the eigenvalues of the matrix $\boldsymbol{Q}_{\boldsymbol{N}_k}^{-1/2} \boldsymbol{Q}_{\alpha_k} \boldsymbol{Q}_{\boldsymbol{N}_k}^{-1/2}$, with $\boldsymbol{Q}_{\boldsymbol{N}_k} = \boldsymbol{N}_k^\top \boldsymbol{W} \boldsymbol{N}_k$ and $\boldsymbol{Q}_{\alpha_k} = \alpha_k \phi \boldsymbol{K}_k$. Note that the AIC is based on information theory and is useful for selecting an appropriate model and smoothing parameters given data with adequate sample size; see Ferreira et al (2012) and Ventura et al. (2019).

3. LOCAL INFLUENCE DERIVATION

3.1. General context

Note that, in general, local influence analysis can be developed jointly for the entire parameter vector. However, it is important to know the influence that the observations exert separately on the estimates of the parametric components, nonparametric components and the dispersion parameter. Some works related to the application of the method of local influence in semiparametric models have revealed empirical evidence that the observations that exert an influence on the estimates of the parametric component are not necessarily influential on the estimates of the non parametric component and viceversa.

To assess the influence of perturbations on the MPL estimates $\widehat{\boldsymbol{\theta}}$, we can consider the likelihood displacement defined by $\mathrm{LD}(\boldsymbol{\omega}) = 2[\ell_{\mathrm{p}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha}) - \ell_{\mathrm{p}}(\widehat{\boldsymbol{\theta}}_\omega, \boldsymbol{\alpha})] \geq 0$, where $\widehat{\boldsymbol{\theta}}_\omega$ is the MPL estimates of $\boldsymbol{\theta}$ for a perturbed model, whose perturbed penalized log-likelihood function is denoted by $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})$, and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ is an n-dimensional vector of perturbations restricted to some open subset $\Omega \in \mathbb{R}^n$. It is assumed that exists $\boldsymbol{\omega}_0 \in \Omega$, a vector of no perturbation, such that $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega}_0) = \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha})$. Cook (1986) suggested to study the local behavior of $\mathrm{LD}(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$. The normal curvature at the arbitrary direction $\boldsymbol{l}$, with $\|\boldsymbol{l}\| = 1$ is given by $C_{\boldsymbol{l}}(\widehat{\boldsymbol{\theta}}) = -2\{\boldsymbol{l}^\top \boldsymbol{\Delta}_{\mathrm{p}}^\top \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{-1} \boldsymbol{\Delta}_{\mathrm{p}} \boldsymbol{l}\}$, which is the objective function of the normal curvature, where $\ddot{\boldsymbol{\ell}}_{\mathrm{p}}$ is the penalized Hessian matrix of $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ evaluated at $\widehat{\boldsymbol{\theta}}$, and $\boldsymbol{\Delta}_{\mathrm{p}}$ is a penalized perturbation matrix, with elements

$$\boldsymbol{\Delta}_{\mathrm{p}} = \frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial\theta_l \partial\omega_j}\Big|_{\theta=\widehat{\theta}, \omega=\omega_0}, \quad l = 1, \ldots, p^*; j = 1, \ldots, n,$$

and $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ being the penalized log-likelihood function corresponding to the model perturbed by $\boldsymbol{\omega}$. For the SABR model, the elements of $\ddot{\boldsymbol{\ell}}_{\mathrm{p}}$ are given in the appendix. We consider the direction $\boldsymbol{l} = \boldsymbol{e}_i$, called total local influence of the ith individual, where $\boldsymbol{e}_i$ is an n-dimensional vector with a one at the ith position and zeros at the remaining positions. In this case, the normal curvature takes the form $C_{\boldsymbol{e}_i}(\boldsymbol{\theta}) = 2|\,\mathrm{c}_{ii}\,|$, where c_{ii} is the ith principal diagonal element of the matrix $\boldsymbol{C} = \boldsymbol{\Delta}_{\mathrm{p}}^\top \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{-1} \boldsymbol{\Delta}_{\mathrm{p}}$. The index plot of $\boldsymbol{l}$ may reveal those observations that under small perturbations exert a notable influence on $\widehat{\boldsymbol{\theta}}$. In order to have a curvature invariant under uniform change of scale, we consider the conformal normal curvature proposed by Poon and Poon (1999). This normal curvature is defined as $B_\ell(\boldsymbol{\theta}) = -[\boldsymbol{l}^\top \boldsymbol{\Delta}_{\mathrm{p}}^\top \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{-1} \boldsymbol{\Delta}_{\mathrm{p}} \boldsymbol{l}]/[(\mathrm{tr}(\boldsymbol{\Delta}_{\mathrm{p}}^\top \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{-1} \boldsymbol{\Delta}_{\mathrm{p}})^2)^{1/2}]$, and is characterized to allow for any unitary direction $\boldsymbol{l}$, with $0 \leq B_\ell(\theta) \leq 1$. A suggestion is to consider, for example, the direction $\boldsymbol{l} = \boldsymbol{e}_i$ and observing the index plot of $B_{\boldsymbol{e}_i}(\boldsymbol{\theta})$. If our interest lies in studying the local influence on a subvector of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}_1$, the normal curvature for $\boldsymbol{\theta}_1$ at the unitary direction $\boldsymbol{l}$ is given by $C_\ell(\widehat{\boldsymbol{\theta}}_1) = -2[\boldsymbol{l}^\top \boldsymbol{\Delta}_{\mathrm{p}}^\top (\ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{-1} - \boldsymbol{G}_{22}) \boldsymbol{\Delta}_{\mathrm{p}} \boldsymbol{l}]$, where

$$\boldsymbol{G}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\boldsymbol{\ell}}_{\mathrm{p}_{22}}^{-1} \end{pmatrix},$$

with $\ddot{\boldsymbol{\ell}}_{\mathrm{p}_{22}}$ obtained from the partition of $\ddot{\boldsymbol{\ell}}_{\mathrm{p}}$ according to the partition of $\boldsymbol{\theta}$. In this case, the index plot of the eigenvector $\boldsymbol{l} = \boldsymbol{l}_{\max}$, which corresponds to the largest absolute eigenvalue of the matrix $\boldsymbol{G} = \boldsymbol{\Delta}_{\mathrm{p}}^{\top}(\ddot{\boldsymbol{L}}_{\mathrm{p}}^{-1} - \boldsymbol{G}_{22})\boldsymbol{\Delta}_{\mathrm{p}}$, may indicate the points with large influence on $\widehat{\boldsymbol{\theta}}_1$.

3.2. Cases-weight perturbation

Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^{\top}$ be a weight vector. In this case, perturbed penalized log-likelihood function is given by

$$\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega}) = \ell(\boldsymbol{\theta}|\boldsymbol{\omega}) - \sum_{k=1}^{s} \frac{\alpha_k}{2}\,\mathrm{f}_k^{\top}\,\boldsymbol{K}_k\,\mathrm{f}_k,$$

where $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \ell_i(\boldsymbol{\theta})$, with $0 \leq \omega_i \leq 1$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^{\top}$. Hence, the elements of the penalized perturbation matrix are expressed as

$$\begin{aligned}
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^{\top}}\Big|_{\theta=\hat{\theta},\omega=\omega_0} &= \widehat{\phi}\boldsymbol{X}^{\top}\widehat{\boldsymbol{T}}\widehat{\boldsymbol{E}}, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial\,\mathrm{f}_k\,\partial\boldsymbol{\omega}^{\top}}\Big|_{\theta=\hat{\theta},\omega=\omega_0} &= \widehat{\phi}\boldsymbol{N}_k^{\top}\widehat{\boldsymbol{T}}\widehat{\boldsymbol{E}}, \quad k = 1, \ldots, s, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial\phi\partial\boldsymbol{\omega}^{\top}}\Big|_{\theta=\hat{\theta},\omega=\omega_0} &= \widehat{\boldsymbol{u}}^{\top},
\end{aligned}$$

where $\widehat{\boldsymbol{E}} = \mathrm{diag}(y_i^* - \widehat{\mu}_i^*)$, for $i = 1, \ldots, n$, and $\widehat{\boldsymbol{u}} = (\widehat{u}_1, \ldots, \widehat{u}_n)^{\top}$, with $u_i = \mu_i[y_i^* - \mu_i^*] + \ln(1 - y_i) - \psi[(1 - \mu_i)\phi] + \psi(\phi)$.

3.3. Response perturbation

Consider now an additive perturbation on the ith response by making $y_{i\omega} = y_i + \omega_i$, where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^{\top}$ is the vector of perturbations. Then, under the scheme of response perturbation, the perturbed penalized log-likelihood function is constructed from (2.5) with y_i replaced by $y_{i\omega}$ and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^{\top}$. Hence, the elements of the penalized perturbation matrix take the form

$$\begin{aligned}
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^{\top}}\Big|_{\theta=\hat{\theta},\omega=\omega_0} &= \widehat{\phi}\boldsymbol{X}^{\top}\widehat{\boldsymbol{T}}\boldsymbol{M}, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial\,\mathrm{f}_k\,\partial\boldsymbol{\omega}^{\top}}\Big|_{\theta=\hat{\theta},\omega=\omega_0} &= \widehat{\phi}\boldsymbol{N}_k^{\top}\widehat{\boldsymbol{T}}\boldsymbol{M}, \quad k = 1, \ldots, s, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial\phi\partial\boldsymbol{\omega}^{\top}}\Big|_{\theta=\hat{\theta},\omega=\omega_0} &= \widehat{\boldsymbol{a}}^{\top},
\end{aligned}$$

where $\boldsymbol{M} = \mathrm{diag}_{1\leq i\leq n}(m_i)$ and $\boldsymbol{a} = (a_1, \ldots, a_n)^{\top}$, with $m_i = 1/[y_i(1 - y_i)]$ and $a_i = -[y_i - \mu_i]/[y_i(1 - y_i)]$.

Note the perturbation $\boldsymbol{\omega}$ must be generated in the support $[-y, 1-y]$ in order to guarantee that the perturbed response variable retains the original distribution support. It is important to mention that the space of $\boldsymbol{\omega}$ depends on the type of perturbation that is introduced in the response variable. In our case, we consider a perturbation of additive type, but eventually we could consider a perturbation of the multiplicative type.

3.4. Continuos covariate perturbation

Consider now an additive perturbation on a continuous covariate, namely $x_{id\omega}$, by making $x_{id\omega} = x_{id} + \omega_i$, with $\omega_i \in \mathbb{R}$. Then, under the scheme of covariate perturbation, the perturbed penalized log-likelihood function is constructed from (2.5) with x_{id} replaced by $x_{id\omega}$, $\mu_{i\omega} = g^{-1}(\eta_{i\omega})$ in the place of μ_i, for $\eta_{i\omega} = \boldsymbol{x}_{i\omega}^\top \boldsymbol{\beta} + \boldsymbol{n}_{1_i}^\top \mathrm{f}_1 + \ldots + \boldsymbol{n}_{s_i}^\top \mathrm{f}_s$, and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$. Hence, the elements of the penalized perturbation matrix assumes the form

$$
\begin{aligned}
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\Big|_{\theta=\hat{\theta}, \omega=\omega_0} &= -\widehat{\phi}\widehat{\beta}_d \boldsymbol{X}^\top \widehat{\boldsymbol{Q}} + \widehat{\phi} \boldsymbol{P} \widehat{\boldsymbol{T}} \widehat{\boldsymbol{E}}, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial \, \mathrm{f}_k \, \partial \boldsymbol{\omega}^\top}\Big|_{\theta=\hat{\theta}, \omega=\omega_0} &= -\widehat{\phi}\widehat{\beta}_d \boldsymbol{N}^\top \widehat{\boldsymbol{Q}} + \widehat{\phi} \boldsymbol{P} \widehat{\boldsymbol{T}} \widehat{\boldsymbol{E}}, \quad k = 1, \ldots, s, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^\top}\Big|_{\theta=\hat{\theta}, \omega=\omega_0} &= -\widehat{\beta}_d \widehat{\boldsymbol{h}}^\top \widehat{\boldsymbol{T}},
\end{aligned}
$$

where $\boldsymbol{P}$ is a $p \times n$ matrix of zeros except for the pth line, which contains ones, and $\boldsymbol{h} = (h_1, \ldots, h_n)^\top$, with $h_i = c_i - (y_i^* - \mu_i^*)$ and c_i as defined in appendix.

3.5. Scale perturbation

Under scale parameter perturbation scheme, it is assumed that $\phi_{i\omega} = \omega_i^{-1}\phi$, with $\omega_i > 0$. Then, the perturbed penalized log-likelihood function is constructed from (2.5) with ϕ replaced by $\phi_{i\omega}$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. Hence, the elements of the penalized perturbation matrix take the form

$$
\begin{aligned}
\frac{\partial^2 \ell_{\mathrm{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\Big|_{\theta=\hat{\theta}, \omega=\omega_0} &= \widehat{\phi} \boldsymbol{X}^\top \widehat{\boldsymbol{T}} \widehat{\boldsymbol{F}}, \\
\frac{\partial^2 \ell_{\mathrm{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial \, \mathrm{f}_k \, \partial \boldsymbol{\omega}^\top}\Big|_{\theta=\hat{\theta}, \omega=\omega_0} &= \widehat{\phi} \boldsymbol{N}_k^\top \widehat{\boldsymbol{T}} \widehat{\boldsymbol{F}}, \quad k = 1, \ldots, s, \\
\frac{\partial^2 \ell_{\mathrm{p}_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^\top}\Big|_{\theta=\hat{\theta}, \omega=\omega_0} &= \widehat{\phi} \widehat{\boldsymbol{d}}^\top - \widehat{\boldsymbol{u}}^\top,
\end{aligned}
$$

where $\boldsymbol{F} = \mathrm{diag}_{1 \le i \le n}\big(F_i\big)$, $\boldsymbol{d} = (d_1, \ldots, d_n)^\top$, $\boldsymbol{u} = (u_1, \ldots, u_n)^\top$, with $u_i = \mu_i[y_i^* - \mu_i^*] + \log(1 - y_i) - \psi[(1 - \mu_i)\phi] + \psi(\phi)$, $F_i = [c_i - (y_i^* - \mu_i^*)]$ and c_i as defined in appendix.

4. EMPIRICAL ILLUSTRATION

4.1. Data and exploratory analysis

To illustrate our methodology, we consider the Australian athletes data set that has been reported by Telford and Cunningham (1991). The purpose of this study is to investigate the relationships of hematological measures with various covariates, such as height and mass, among others, for a sample of 202 elite Australian athletes who trained at the Australian Institute of Sport. The objetive of the present data analysis is to model the percent body fat through the SABR model. We consider as covariates: (i) sum of skin folds (SSF), (ii) hemaglobin concentration (HG), and (iii) lean body mass (LBM), whereas the percent body fat (BFAT) is the response variable. Figure 1 contains the scatter plots between the response and each covariate. From Figure 1 (a), we observe that a linear relationship between BFAT and SSF, while that Figures 1 (b)-(c) show no evidence of linear relationships between BFAT and the covariates HG and LBM.

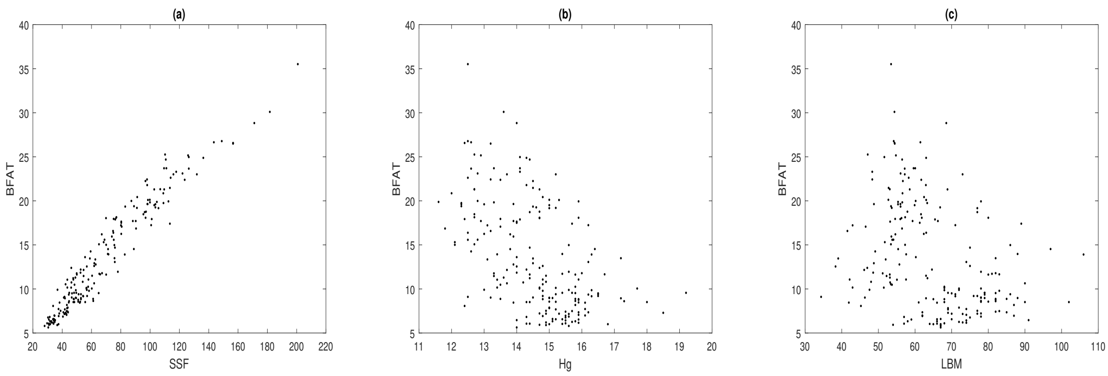

Figure 1: Scatter plots of BFAT versus SSF (a), HG (b) and LBM (c) with Australian athletes data.

4.2. Model fitting

First, we consider an BR model given by

$$g(\mu_i) = \beta_0 + \beta_1 \mathrm{SSF}_i + \beta_2 \mathrm{HG}_i + \beta_3 \mathrm{LBM}_i, \quad i = 1, \ldots, 202,$$

with logit link function. The maximum likelihood estimates and the corresponding approximate estimated standard error (in parenthesis) are reported in Table 1, with AIC $= -1031.317$. Note that under this model, we are assuming that the relationship of each covariate with the response is linear. However, as previously commented, we observe in Figures 1(b) and 1(c) that the relationships between BFAT and the covariates HG and LBM seem to be nonlinear, which suggests a SABR model with link function

$$g(\mu_i) = \beta_0 + \beta_1 \mathrm{SSF}_i + f_1(\mathrm{HG}_i) + f_2(\mathrm{LBM}_i).$$

The MPL estimates and the corresponding approximate estimated standard error associated with the parametric component and scale parameter are also reported in Table 1.

Table 1: Maximum likelihood and MPL estimates and the standard error (in parenthesis) for indicated model with Australian athletes data.

	Parameters				
Model	β_0	β_1	β_2	β_3	ϕ
BR	-2.020(0.1591)	0.012(0.0003)	-0.027(0.0122)	-0.005(0.0012)	307.180(30.5601)
SABR	-2.788(0.0420)	0.012(0.0004)	–	–	361.768(35.9955)

Comparing the results reported in Table 1, we note a similarity between the estimates for $\widehat{\beta}_0$ and $\widehat{\beta}_1$ under both models, but the estimated standard error of $\widehat{\beta}_0$ is smaller under the SABR model. However, the estimate $\widehat{\phi}$ under the SABR model is larger (including its estimated standard error) than that obtained for the BR model. The estimates of the smoothing parameters α_1 and α_2, as well as the corresponding DFs, are reported in Table 2.

Table 2: Smoothing components fitted under the SABR model to Australian athletes data.

	Smoothing function	
	f_1(HG)	f_2(LBM)
DF(α_k)	4.662	4.988
α_k	0.0014	0.9012

Figures 2(a)-(b) show the estimated smooth functions under the SABR model and the corresponding approximate SEB (dashed curves). The estimated smooth functions are computed using the smoothing parameters obtained by the AIC as described in Subsection 2.4. The graphical plots suggest nonlinear tendencies for HG and LBM. Then, we find a value of AIC$(\alpha_1, \alpha_2) = -1055.466$, which is less than that obtained under the BR model, indicating a superiority of the model that includes a nonparametric additive component.

4.3. Diagnostics

In this illustration, we consider residual proposed by Ferrari and Cribari-Neto (2004) given by

$$r_i = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mathrm{Var}}(y_i)}}, \quad i = 1, \ldots, 202,$$

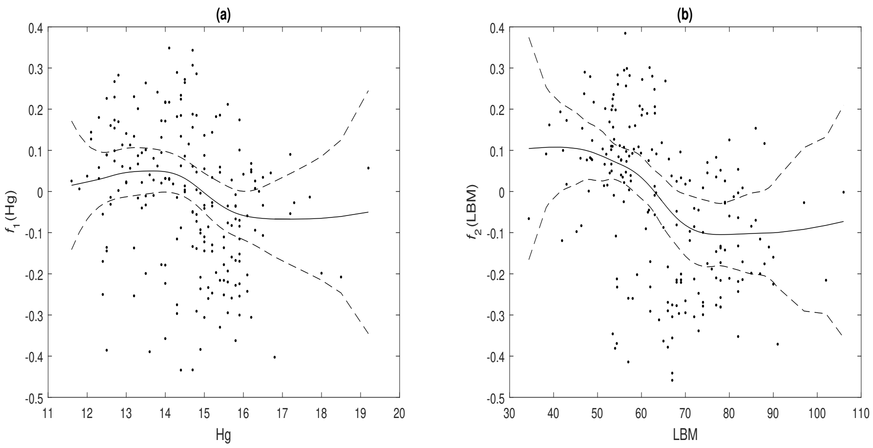

Figure 2: Plots estimated smooth functions f_1 (a) and f_2 (b) for the Australian athletes data and their approximate pointwise SEB denoted by the dashed lines.

where $\widehat{\mathrm{Var}}(y_i) = \widehat{\mu}_i[1-\widehat{\mu}_i]/[1+\widehat{\phi}]$ and $\widehat{\mu}_i = g^{-1}(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}} + \boldsymbol{n}_{1_i}^\top \widehat{\mathrm{f}}_1 + \boldsymbol{n}_{2_i}^\top \widehat{\mathrm{f}}_2)$, with $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathrm{f}}_k$, for $k = 1, 2$, and $\widehat{\phi}$ denoting the MPL estimates.

Figure 3 displays the graphical plot of the standardized ordinary residuals against the indices of the observations. We note that the residual are randomly scattered around zero and that observations #51, #53 and #56 are indicated as atypical cases. Note that a residual analysis permits us to detect deviations from the model assumptions, but also atypical cases. An atypical case can be potentially influential or not, but from a scope of global influence. However, potentially influential cases detected by the displacement of likelihood functions are evaluated from a scope of local influence. In any case, this potential influence (global or local) must be studied by means of relative changes (RC) when the potentially influential case is removed from the full data set. This allows us to know whether inferential changes are generated or not. Now, in

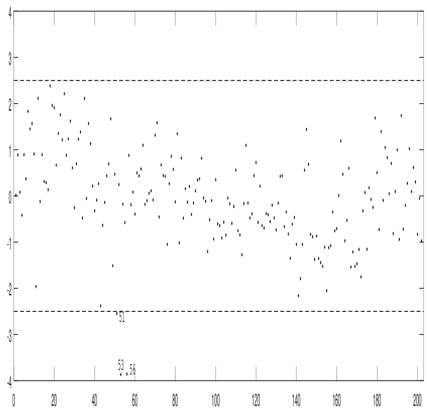

Figure 3: Plot standardized ordinary residuals versus the index of the observation.

order to identify potentially influential observations under the fitted model to Australian athletes data, we present index plots of $B_i = B_{e_i}(\boldsymbol{\lambda})$, for $\boldsymbol{\lambda} = \boldsymbol{\beta}, \mathrm{f}_k, \phi$, with $k = 1, 2$. Figure 4 shows the index plot of B_i for the case-weight perturbation scheme under the fitted model. Looking at Figure 4, note that observations #51, #53 and #56 are more influential on the MPL estimate $\widehat{\boldsymbol{\lambda}}$.

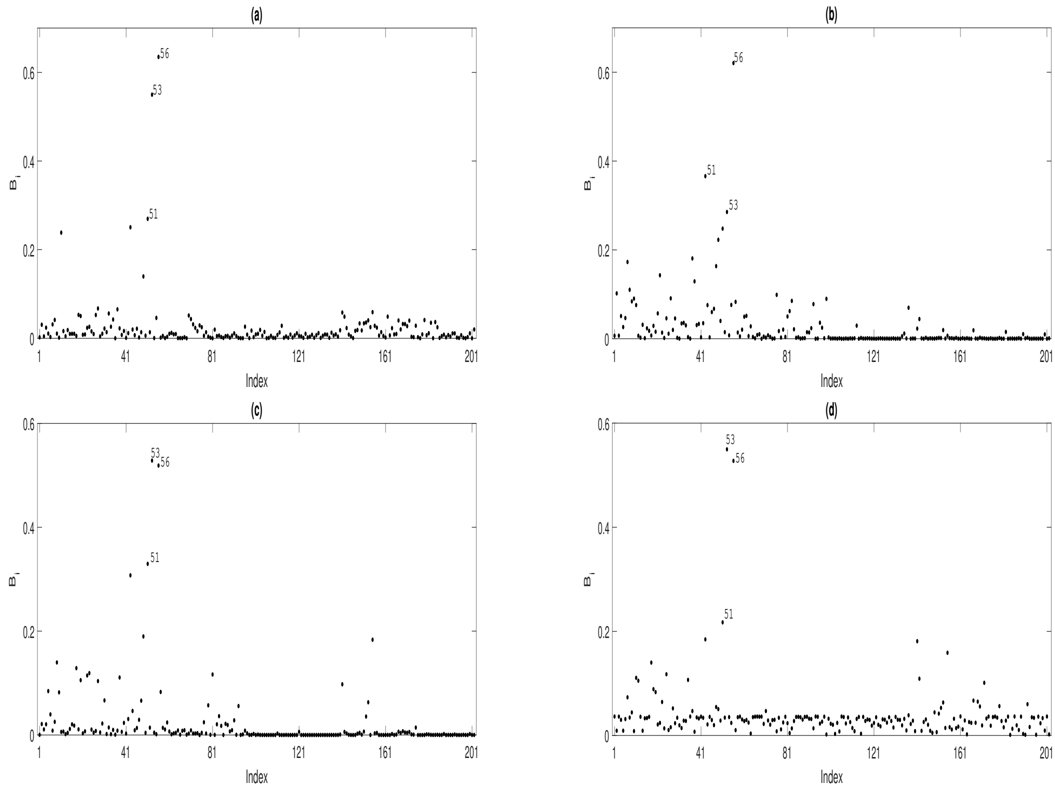

Figure 4: Index plots of B_i for assessing local influence on $\widehat{\boldsymbol{\beta}}$ (a), $\widehat{f_1}$ (b), $\widehat{f_2}$ (c) and $\widehat{\phi}$ (d) considering case-weight perturbation under model fitted to Australian athletes data.

Figure 5 presents the index plots of B_i, considering the response perturbation scheme under the fitted model. Considering Figure 5, observe that observations #160, #166, and #181 are more influential on MPL estimate $\widehat{\boldsymbol{\lambda}} = \widehat{f_1}$, whereas none observation is pointed out on the estimates remaining. The index plots of B_i for the scale parameter and covariate perturbation are omitted because the results are similar to those obtained under case-weight perturbation scheme. Note that observations #51, #53 and #56 are also detected as atypical according to the residual analysis.

We now investigate the impact on the model inference when the observations detected as potentially influential in the diagnostic analysis are removed. Table 3 presents the RCs in % of the MPL estimates of β_j, for $j = 1, 2$, and ϕ after removing from the data set the pointed out observations in the local influence graphical plots under the SABR model. The RCs of each estimated parameter are defined as $\mathrm{RC}_\psi = |(\widehat{\psi} - \widehat{\psi}_{(\mathrm{I})})/\widehat{\psi}| \times 100\%$, where $\widehat{\psi}_{(\mathrm{I})}$ denotes the MPL estimate of ψ, with $\psi = \beta_j, \phi$, after the corresponding observation(s) are removed according to the set I. Note that, although some RC are large, inferential changes are not detected. It is interesting to notice from Table 3 the coherence with the diagnostic graphical plots shown pre-

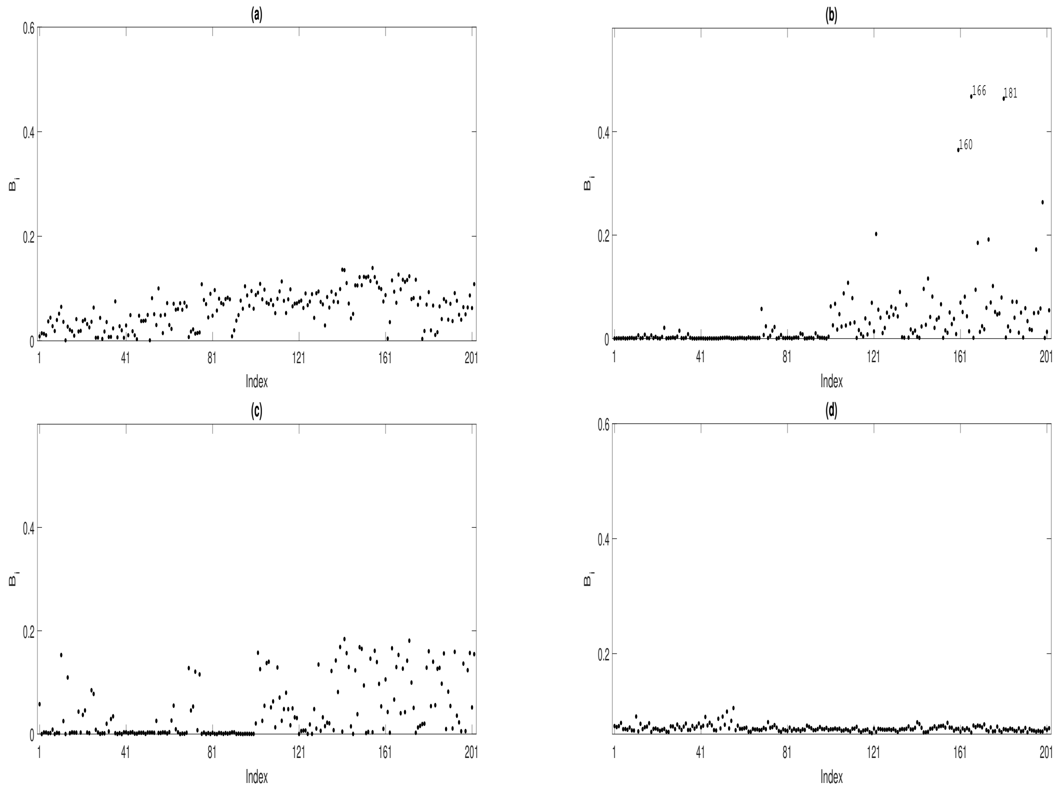

Figure 5: Index plots of B_i for assessing local influence on $\widehat{\boldsymbol{\beta}}$ (a), $\widehat{f}_1$ (b), $\widehat{f}_2$ (c) and $\widehat{\phi}$ (d) considering response perturbation under the fitted model to Australian athletes data.

viously. For instance, elimination of the observations #51, #53 and #56, detected as potentially influential observations by local influence technique, leads to significant changes in the MPL estimate, mainly in $\widehat{\beta}_1$ and $\widehat{\phi}$. This indicates the need of a diagnostic examination.

Table 3: RC (in %) on the MPL estimate of β_j and ϕ under the SABR model fitted to Australian athletes data after removing the indicated cases.

	Parameters			Relative changes		
Removed case	β_0	β_1	ϕ	RC$_{\beta_0}$	RC$_{\beta_1}$	RC$_{\phi}$
51	-2.8004	0.0126	374.3139	0.45	5.00	3.46
53	-2.8216	0.0129	393.3938	1.21	7.50	8.74
56	-2.8230	0.0130	395.2551	1.25	8.31	9.29
51-53	-2.8007	0.0126	372.6768	0.45	5.01	3.02
51-56	-2.8031	0.0127	374.6170	0.54	5.83	3.55
53-56	-2.8245	0.0130	394.0092	1.31	8.32	8.91
51-53-56	-2.8013	0.0126	371.6199	0.47	5.02	2.72

5. CONCLUDING REMARKS

In this paper, we have proposed a methodology of inference and diagnostics for the semiparametric additive beta regression model. Specifically, we have derived a weighted back-fitting iterative process to estimate the parameters of the additive component of the model, that is, of regression coefficients and smooth functions. We have estimated the approximate variance-covariance matrix of maximum penalized likelihood estimates based on the Fisher information matrix obtained from the penalized log-likelihood function. Moreover, we have derived diagnostics for this model using the local influence technique to evaluate the sensitivity of the maximum penalized likelihood estimates by using several perturbation schemes in the model and data. Finally, we have performed a statistical modeling with real data set. The study has provided evidences on the advantage of incorporating a semiparametric additive term in those situations where there are covariates that contribute nonlinearly to the model. Thus, we recommend semiparametric additive beta regression models as an option to fit continuous data sets in the unit interval when covariates are present and that contribute nonlinearly to the model. The computational codes used in the illustration are available under request from the authors.

ACKNOWLEDGEMENTS

The authors would like to thank the editors and four reviewers very much for their constructive comments on an earlier version of this manuscript which resulted in this improved version.This work was supported by Project FONDECYT 11130704, Chile. J.I. Figueroa-Zúñiga acknowledges funding support by grant VRID Enlace N0 217.014.027-1.0, from the Universidad de Concepción, Chile.

APPENDIX

Hessian matrix

Let $\ddot{\boldsymbol{\ell}}_{\rm p}$ $(p^* \times p^*)$ be the Hessian matrix with the (j^*, ℓ^*)-element given by $\partial^2 \ell_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\alpha})/\partial\theta_{j^*}\theta_{\ell^*}$, for $j^*, \ell^* = 1, \ldots, p^*$. After some algebraic manipulations, we find

$$\begin{aligned}
\frac{\partial^2 \ell_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -\phi \boldsymbol{X}^\top \boldsymbol{Q} \boldsymbol{X}, \\
\frac{\partial^2 \ell_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial\,\mathrm{f}_k\,\partial\,\mathrm{f}_{k'}^\top} &= \begin{cases} -\phi \boldsymbol{N}_k^\top \boldsymbol{Q} \boldsymbol{N}_k - \alpha_k \boldsymbol{K}_k, & k = k', \\ -\phi \boldsymbol{N}_k^\top \boldsymbol{Q} \boldsymbol{N}_{k'}, & k \neq k' \end{cases} \\
\frac{\partial^2 \ell_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\,\mathrm{f}_k^\top} &= -\phi \boldsymbol{X}^\top \boldsymbol{Q} \boldsymbol{N}_k, \quad k = 1, \ldots, s,
\end{aligned}$$

where $\boldsymbol{Q} = \text{diag}_{1\leq i\leq n}\big(q_i\big)$, with

$$q_i \;=\; \left[\phi\{\psi'(\mu_i\phi) + \psi'[(1-\mu_i)\phi]\} + [y_i^* - \mu_i^*]\frac{g''(\mu_i)}{g'(\mu_i)}\right]\frac{1}{[g'(\mu_i)]^2}.$$

In addition, we have that the second derivative de $\ell_{\text{p}}(\boldsymbol{\theta},\boldsymbol{\alpha})$ with respect to ϕ, and $\boldsymbol{\beta}$ and f_k, respectively, can be written by

$$\begin{aligned}\frac{\partial^2\ell_{\text{p}}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\phi} &= \frac{2}{\phi^2}\boldsymbol{X}^\top\boldsymbol{b},\\ \frac{\partial^2\ell_{\text{p}}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\,\text{f}_k\,\partial\phi} &= \frac{2}{\phi^2}\boldsymbol{N}_k^\top\boldsymbol{b}, \quad k = 1,\ldots,s,\end{aligned}$$

where $\boldsymbol{b} = (b_1,\ldots,b_n)^\top$, with

$$b_i \;=\; \left\{[y_i^* - \mu_i^*] - \phi\frac{\partial\mu_i^*}{\partial\phi}\right\}\frac{1}{[g'(\mu_i)]}.$$

Furthermore, the second derivative de $\ell_{\text{p}}(\boldsymbol{\theta},\boldsymbol{\alpha})$ with respect to ϕ is given by

$$\frac{\partial^2\ell_{\text{p}}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\phi^2} \;=\; \text{trace}(\boldsymbol{D}),$$

where $\boldsymbol{D} = \text{diag}_{1\leq i\leq n}\big(d_i\big)$, with

$$d_i \;=\; -\Big[\psi'[\mu_i\phi]\mu_i^2 + \psi'[(1-\mu_i)\phi][1-\mu_i]^2 - \psi'(\phi)\Big].$$

Expected information matrix

In general, by calculating the expectation of the matrix $-\ddot{\boldsymbol{\ell}}_{\text{p}}$, we obtain the $p^*\times p^*$ penalized expected information matrix denoted by

$$\boldsymbol{\mathcal{I}}_{\text{p}} \;=\; -\text{E}\left(\frac{\partial^2\ell_{\text{p}}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right).$$

Let $\boldsymbol{W} = \text{block diag}_{1\leq i\leq n}\big(w_i\big)$ and $\boldsymbol{c} = (c_1,\ldots,c_n)^\top$, with

$$w_i = \phi\Big[\psi'[\mu_i\phi] + \psi'[1-\mu_i]\phi\Big]\frac{1}{[g'(\mu_i)]^2}, \quad c_i = \phi\Big[\psi'[\mu_i\phi]\mu_i - \psi'[(1-\mu_i)\phi][1-\mu_i]\Big].$$

After some algebraic manipulations, we find

$$\boldsymbol{\mathcal{I}}_{\text{p}} \;=\; \begin{pmatrix}\boldsymbol{\mathcal{I}}_{11} & \boldsymbol{\mathcal{I}}_{12}\\ \boldsymbol{\mathcal{I}}_{21} & \boldsymbol{\mathcal{I}}_{22}\end{pmatrix},$$

where

$$
\begin{aligned}
\boldsymbol{\mathcal{I}}_{11} &= \begin{pmatrix} \phi \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X} & \phi \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{N}_1 & \dots & \phi \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{N}_s \\ \phi \boldsymbol{N}_1^\top \boldsymbol{W} \boldsymbol{X} & \phi \boldsymbol{N}_1^\top \boldsymbol{W} \boldsymbol{N}_1 + \lambda_1 \boldsymbol{K}_1 & \dots & \phi \boldsymbol{N}_1^\top \boldsymbol{W} \boldsymbol{N}_s \\ \vdots & \vdots & \ddots & \vdots \\ \phi \boldsymbol{N}_s^\top \boldsymbol{W} \boldsymbol{X} & \phi \boldsymbol{N}_s^\top \boldsymbol{W} \boldsymbol{N}_1 & \dots & \phi \boldsymbol{N}_s^\top \boldsymbol{W} \boldsymbol{N}_s + \lambda_s \boldsymbol{K}_s \end{pmatrix}, \\
\boldsymbol{\mathcal{I}}_{12} &= \begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{T} \boldsymbol{c} \\ \boldsymbol{N}_1^\top \boldsymbol{T} \boldsymbol{c} \\ \vdots \\ \boldsymbol{N}_1^\top \boldsymbol{T} \boldsymbol{c} \end{pmatrix} = \boldsymbol{\mathcal{I}}_{21}^\top,
\end{aligned}
$$

and $\boldsymbol{\mathcal{I}}_{22} = \text{trace}(\boldsymbol{D})$. Note that the parameters $\boldsymbol{\beta}$, f_k, with $k = 1, \dots, n$, and ϕ are not orthogonal, in contrast to what is verified in the class of generalized linear regression models.

Iterative process

The solution of the estimating equation system given in (2.6) to obtain the MPL estimate of $\boldsymbol{\theta}$ may be attained by iterating between a weighted back-fitting algorithm with weight matrix $\boldsymbol{W}$ and a Fisher score algorithm to obtain maximum likelihood estimation of the parameter ϕ, which is equivalent to the following iterative process:

(i) Initialize:

(a) By fitting a beta regression model considering $\mathrm{f}_0^{(0)} = \boldsymbol{\beta}^{(0)}$ and $\boldsymbol{N}_0 = \boldsymbol{X}$.

(b) By getting a starting value for ϕ by using the fitted values from (a).

(c) From the current value $\mathrm{f}^{(0)} = (\mathrm{f}_0^\top, \mathrm{f}_1^{(0)\top}, \dots, \mathrm{f}_s^{(0)\top}, \phi^{(0)})^\top$ by obtaining the weight matrix $\boldsymbol{W}^{(0)}$ and $\boldsymbol{T}^{(0)}$, with $w_i^{(0)} = w_i\big|_{\boldsymbol{\theta}^{(0)}}$, and then by getting

$$
\begin{aligned}
\boldsymbol{z}^{(0)} &= \boldsymbol{\eta}^{(0)} + \boldsymbol{W}^{(0)^{-1}} \boldsymbol{T}^{(0)} (yn^* - \boldsymbol{\mu}^{*(0)}), \\
\boldsymbol{S}_0^{(0)} &= (\boldsymbol{N}_0^\top \boldsymbol{W}^{(0)} \boldsymbol{N}_0)^{-1} \boldsymbol{N}_0^\top \boldsymbol{W}^{(0)}, \\
\boldsymbol{S}_k^{(0)} &= (\boldsymbol{N}_k^\top \boldsymbol{W}^{(0)} \boldsymbol{N}_k + \alpha_k \boldsymbol{K}_k)^{-1} \boldsymbol{N}_k^\top \boldsymbol{W}^{(0)}, k = 1, \dots, s.
\end{aligned}
$$

(ii) Iterate repeatedly by cycling between the equations

$$
\begin{aligned}
\mathrm{f}_0^{(u+1)} &= \boldsymbol{S}_0^{(u)} \left(\boldsymbol{z}^{(u)} - \sum_{l=1}^{s} \boldsymbol{N}_l \, \mathrm{f}_l^{(u)} \right), \\
\mathrm{f}_1^{(u+1)} &= \boldsymbol{S}_1^{(u)} \left(\boldsymbol{z}^{(u)} - \boldsymbol{N}_0 \, \mathrm{f}_0^{(u+1)} - \sum_{l=2}^{s} \boldsymbol{N}_l \, \mathrm{f}_l^{(u)} \right), \\
&\vdots \\
\mathrm{f}_s^{(u+1)} &= \boldsymbol{S}_s^{(u)} \left(\boldsymbol{z}^{(u)} - \sum_{l=0}^{s-1} \boldsymbol{N}_l \, \mathrm{f}_l^{(u+1)} \right),
\end{aligned}
$$

for $u = 0, 1, \ldots$. Repeat step (ii) replacing $\mathrm{f}_j^{(u)}$ by $\mathrm{f}_j^{(u+1)}$ until convergence criterion $\Delta_u(\mathrm{f}_j^{(u+1)}, \mathrm{f}_j^{(u)})= \sum_{j=0}^{s} \parallel \mathrm{f}_j^{(u+1)} - \mathrm{f}_j^{(u)} \parallel / \sum_{j=0}^{s} \parallel f_j^{(u)} \parallel$ is reached for a threshold value; see Hastie and Tibshirani (1990).

(iii) For current values $\mathrm{f}_j^{(u+1)}$, with $j = 0, 1, \ldots, s$, obtain $\phi^{(u+1)}$ by using

$$\phi^{(u+1)} = \phi^{(u)} - \mathrm{E}\left\{\frac{\partial^2 \ell_{\mathrm{p}}^{c}(\phi, \boldsymbol{\alpha})}{\partial\phi\partial\phi}\right\}^{-1} \frac{\partial \ell_{\mathrm{p}}^{c}(\phi, \boldsymbol{\alpha})}{\partial\phi}\Big|_{\boldsymbol{\theta}}^{(u)}.$$

(iv) Iterate between steps (ii) and (iii) by replacing $\mathrm{f}_j^{(0)}$, with $j = 0, 1, \ldots, s$, and $\phi^{(0)}$ by $\mathrm{f}_j^{(u+1)}$ and $\phi^{(u+1)}$, respectively, until reaching convergence.

REFERENCES

Adams, R.A. and Fournier, J. (2003). Sobolev Spaces. Pure and Applied Mathematics. Academic Press, Boston.

Berhane, K. and Tibshirani, J. (1998). Generalized additive models for longitudinal data. The Canadian Journal of Statistics, 26, 517–535.

Cao, C.Z. and Lin, J.G. (2011). Diagnostics for elliptical linear mixed models with first-order autoregressive errors. Journal of Statistical Computation and Simulation, 81, 1281–1296.

Cysneiros, F.J.A., Leiva, V., Liu, S., Marchant, C., and Scalco, P. (2019) A Cobb-Douglas type model with stochastic restrictions: formulation, local influence diagnostics and data analytics in economics. Quality and Quantity, 53, 1693–1719.

Cook, R.D. (1986). Assessment of local influence (with discussion). Journal of the Royal Statistical Society B, 48, 133–169.

Cribari-Neto, F. and Zeileis, A. (2010). Beta regression in R. Journal of Statistical Software, 34, Issue 2.

Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. Statistical Science, 11, 89–121.

Emmami, H. (2017). Local influence for Liu estimators in semiparametric linear models. Statistical Papers, 19, 529–544.

Espinheria, P., Ferrari, S., and Cribari-Neto, F. (2008). On beta regression residuals. Journal of Applied Statistics, 35, 407–419.

Espinheria, P., Ferrari, S., and Cribari-Neto, F. (2008). Influence diagnostics in beta regression. Computational Statistics and Data Analysis, 52, 4417–4431.

Fang, K., Yao, Z. (2017). Semi-parametric additive beta regression model with its application CMS, 25, 116–124. `10.16381/j.cnki.issn1003-207x.2017.09.013`.

Ferrari, S. and Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions. Journal of Applied Statistics, 31, 799–815.

Ferrari, S. (2011). Diagnostics tools in beta regression with varying dispersion. Statistica Neerlandica, 65, 337–351.

Ferreira, M., Gomes, M.I., and Leiva, V. (2012) On an extreme value version of the Birnbaum-Saunders distribution. REVSTAT, 10, 181–210.

Ferreira, C.S. and Paula, G.A. (2017). Estimation and diagnostic for skew-normal partially linear models. Journal of Applied Statistics, 44, 3033–3053.

Figueroa-Zuñiga, J., Arellano-Valle, R., and Ferrari, S. (2013). Mixed beta regression: A Bayesian perspective. Computational Statistics and Data Analysis, 61, 137–147.

Garcia-Papani, F., Leiva, V., Uribe-Opazo, M.A., and Aykroyd, R.G. (2018) Birnbaum-Saunders spatial regression models: Diagnostics and application to chemical data. Chemometrics and Intelligent Laboratory Systems, 177, 114–128.

Green, P.J. and Silverman, B.W. (1994). Nonparametric Regression and Generalized Linear Models. Chapman and Hall, Boca Raton.

Hastie, T. and Tibshirani, R. (1990). Generalized Additive Models. Chapman and Hall, London.

Hastie, T., Tibshirani, R, and Friedman, J. (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer,

Huerta, M., Leiva, V., Lillo, C., and Rodriguez, M. (2018) A beta partial least squares regression model: Diagnostics and application to mining industry data. Applied Stochastic Models in Business and Industry, 34, 305–321.

Ibacache-Pulgar, G. and Paula, G.A. (2011). Local influence for Student-t partially linear models. Computational Statistics and Data Analysis, 55, 1462–1478.

Ibacache-Pulgar, G., Paula, G.A., and Galea, M. (2012). Influence diagnostics for elliptical semiparametric mixed models. Statistical Modelling, 12, 165–193.

Ibacache-Pulgar, G., Paula, G.A., and Cysneiros, F.J.A. (2013). Semiparametric additive models under symmetric distributions. Test, 22, 103–121.

Ibacache-Pulgar, G. and Reyes, S. (2018). Local influence for elliptical partially varying-coefficient model. Statistical Modelling, 18, 149–174.

Leao, J., Leiva, V., Saulo, H., and Tomazella, V. (2018) Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. Statistics in Medicine, 37, 4421–4440.

Marchant, C., Leiva, V., Cysneiros, F.J.A., and Vivanco, J.F. (2016) Diagnostics in multivariate Birnbaum-Saunders regression models. Journal of Applied Statistics, 43, 2829–2849.

Osorio, F., Paula, G.A. and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. Computational Statistics and Data Analysis, 51, 4354–4368.

Poon, W. and Poon, Y.S. (1999). Conformal normal curvature and assessment of local influence. Journal of the Royal Statistical Society B, 61, 51–61.

Queiroz da-Silva, C. and Migon, H. (2016) Hierarchical Beta model. REVSTAT, 14, 49–73.

Rocha, A. and Simas, A. (2011). Influence diagnostics in a general class of beta regression models. TEST, 20, 95–119.

Stasinopoulos, D.M. and Rigby, R.A. (2007). Generalized additive models for location scale and shape (GAMLSS) in R. Journal of Statistical Software, 23, 1–46.

Tapia, H., Leiva, V., Diaz, M.P., and Giampaoli, V. (2019a) Influence diagnostics in mixed effects logistic regression models. TEST, pages in press.

Tapia, H., Giampaoli, V., Diaz, M.P., and Leiva, V. (2019b) Sensitivity analysis of longitudinal count responses: A local influence approach and application to medical data. Journal of Applied Statistics, 46, 1021–1042.

Telford, R.D. and Cunningham, R.B. (1991). Sex, sport and body-size dependency of hematology in highly trained athletes. Medicine and Science in Sports and Exercise, 23, 788–794.

Uribe-Opazo, M.A., Borssoi, J.A. and Galea, M. (2012). Influence diagnostics in Gaussian spatial linear models. Journal of Applied Statistics, 3, 615–630.

Ventura, M., Saulo, H., Leiva, V., and Monsueto, S. (2019). Log-symmetric regression models: information criteria, application to movie business and industry data with economic implications. Applied Stochastic Models in Business and Industry, pages in press.

Zhu, H.T. and Lee, S.Y. (2003). Local influence for generalized linear mixed models. The Canadian Journal of Statistics, 31, 293–309.

Zhu, Z.Y., He, X., and Fung, W.K. (2003). Local influence analysis for penalized Gaussian likelihood estimators in partially linear models. Scandinavian Journal of Statistics, 30, 767–780.

Zhang, J., Zhang, X., Ma, H., and Zhiya, C. (2015). Local influence analysis of varying coefficient linear model. Journal of Interdisciplinary Mathematics, 3, 293–306.

Zhao, W., Zhang, R., Lv, Y., and Liu, J. (2014). Variable selection for varying dispersion beta regression model. Journal of Applied Statistics, 41, 95–108.