REVSTAT – Statistical Journal
Volume 22, Number 2, April 2024, 211–237
https://doi.org/10.57805/revstat.v22i2.483

Birnbaum–Saunders Semi-Parametric Additive Modeling: Estimation, Smoothing, Diagnostics, and Application

Authors: Esteban Cárcamo
– Institute of Statistics, Universidad de Valparaíso, Chile
esteban.carcamo@alumnos.uv.cl

Carolina Marchant
– Faculty of Basic Sciences, Universidad Católica del Maule, Talca, Chile
– ANID – Millennium Science Initiative Program – Millennium Nucleus Center for the Discovery of Structures in Complex Data, Santiago, Chile
carolina.marchant.fuentes@gmail.com

Germán Ibacache-Pulgar
– Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile
– Interdisciplinary Center for Atmospheric and Astro-Statistical Studies, Universidad de Valparaíso, Valparaíso, Chile
gibacachepulgar@gmail.com

Víctor Leiva ✉
– School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
victorleivasanchez@gmail.com

Received: January 2022 Revised: May 2022 Accepted: May 2022

Abstract:

- Inclusion of nonparametric functions enhances the modeling when accommodating non-linear effects of covariates. Semi-parametric models have been successfully used for describing non-linear structures by means of parametric and nonparametric components. In this work, we formulate a semi-parametric additive regression model based on a Birnbaum–Saunders distribution and carry out influence diagnostics for such a model. This semi-parametric structure permits us to model the mean and variance simultaneously. We employ a back-fitting algorithm to get the penalized maximum likelihood estimates by utilizing cubic smoothing splines. We derive methods of local influence by calculating the normal curvatures under different perturbation schemes. The obtained results are computationally implemented in the R software so that diverse users have available this model computationally to be applied in practice. Finally, an application of the proposed model with real data from one of the most polluted cities in the world is presented.

Keywords:

- *local influence; penalized maximum likelihood estimators; R software; splines; weighted back-fitting algorithm.*

AMS Subject Classification:

- 62D05, 62F99, 62J99.

✉ Corresponding author

1. INTRODUCTION

Semi-parametric structures are a powerful tool in statistical modeling when incorporating covariates that can contribute parametrically or nonparametrically in the model [20, 21]. The Birnbaum–Saunders (BS) distribution has received considerable attention in recent years, due to its theoretical arguments associated with cumulative damage processes, its properties, and its relation with the normal distribution. Specifically, the amount of cumulative damage that allows the BS distribution to be generated is assumed to follow a normal distribution. The BS model corresponds to a unimodal, positively skewed, two-parameter distribution and with support in the positive real numbers. The BS distribution has its genesis from engineering, but it has been also mathematically justified and applied to solve environmental problems [7, 8, 28, 31, 35, 40, 48].

Extensive work has been done on different aspects with regard to the BS distribution during the last two decades. The main addressed topics are related to mathematical and statistical properties of this distribution, point and interval estimation with classical or Bayesian methods, hypothesis testing, log-linear and non-linear regression models, random number generators, development of packages in the `R` software, generalizations and reparameterizations, multivariate versions, and extensions to temporal and spatial modeling; for a recent and detailed review on these issues, see [3, 30].

A new parameterization of the BS (RBS) distribution stated in [46, 47] is based on its mean and precision parameters. By using this parametrization, a type of generalized linear model (GLM) was formulated in [27] based on the BS distribution although this does not belong to the exponential family. Such a formulation permits us to model the mean and variance simultaneously, maintaining the original scale of the data with no transformations on the modeled variable, because it is well known that transformations reduce interpretability. Statistical modeling based on BS distributions has considerably attracted the attention of a number of researchers; see, for example, [10, 32, 42, 49]. To the best of our knowledge, no semi-parametric models based on the RBS distribution have been derived until now in the literature.

Diagnostic analytics is an important step to be considered in all data modeling. Diagnostics can be conducted by goodness-of-fit techniques, residual analysis, and global/local influence methods. Goodness-of-fit methods allow us to assess the adequacy of a model to a data set. Residual analysis is a helpful tool for evaluating the fit of a statistical model. Several types of residuals for BS regressions were proposed in [27, 29], stating by simulations which of them has better performance. Global influence techniques remove cases and evaluate their effect on the fitted model. Local influence was proposed in [9] and permits us to detect the effect of perturbations on the estimates of model parameters. Local influence methods for RBS regression models were derived in [27] by calculating the normal curvatures under different perturbation schemes. To the best of our knowledge, no diagnostic analytics in semi-parametric additive models (SAM) based on BS distributions have been considered in the literature at present.

Our main objective is to formulate a novel RBS semi-parametric additive regression model (RBS-SAM). This semi-parametric model enables us to describe the mean and variance

simultaneously. Our secondary objectives are:

- (**i**) to estimate the model parameters with the maximum penalized likelihood (MPL) method and a back-fitting algorithm;
- (**ii**) to describe the nonparametric structure with cubic smoothing splines;
- (**iii**) to derive local influence for model diagnostics by calculating the normal curvatures under different perturbations;
- (**iv**) to implement the obtained results in the `R` software;
- (**v**) to apply the results to real data associated with pollutant contents in Santiago of Chile — one of the most polluted cities in the world [33, 8, 40].

These data were secured from the website of the Chilean Ministry of Environment. Our numerical illustrations were developed with the `R` software [41].

The application presented in this study is motivated by the fact that inclusion of nonparametric functions greatly enhances the modeling when accommodating non-linear effects of covariates [21, 20]. These covariates correspond in our case to contents of pollutants and meteorological variables as atmospheric pressure, precipitation, relative humidity, temperature, and wind speed [40]. Semi-parametric estructures have been successfully used to model non-linear components [23]. In addition, the application considered in the present study is supported by theoretical arguments that permit us to justify the use of BS distributions for describing environmental data. This argumentation was formalized in [28] employing a novel mathematical model for environmental sciences based on physical laws.

This article is organized as follows. Section 2 provides an overview of BS distributions, introduces the RBS-SAM, and considers a penalized log-likelihood function for parameter estimation. In Section 3, we obtain the MPL estimators, use a back-fitting algorithm, derive the penalized score vector and Hessian matrix, determine the degrees of freedom, select the smoothing parameter, and conduct inference for the corresponding parameters. In Section 4, the main concepts of local influence and the derivation of normal curvatures for some perturbation schemes are presented. Section 5 introduces the empirical application of the proposed model to an environmental data set. In Section 6, some concluding remarks and ideas for future research are given.

2. PARAMETRIC SPECIFICATION AND NONPARAMETRIC COMPONENT

In this section, an overview of BS distributions, the RBS-SAM structure, and the corresponding penalized log-likelihood function for parameter estimation are stated.

2.1. The BS distribution and related models

Next, we present the traditional BS distribution and related models. If a random variable Y^* is BS distributed with shape ($\alpha > 0$) and scale ($\beta > 0$) parameters, we employ the notation $Y^* \sim \mathrm{BS}(\alpha, \beta)$. Note that $Z = (1/\alpha)((Y^*/\beta)^{1/2} - (\beta/Y^*)^{1/2}) \sim \mathrm{N}(0,1)$ if $Y^* \sim \mathrm{BS}(\alpha, \beta)$.

Hence, each BS distributed random variable Y^* may be obtained as a transformation of a standard normal distributed random variable Z by means of

$$Y^* = \beta\left(\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1}\right)^2. \tag{2.1}$$

Note from the formula stated in (2.1) that the quantile function or $100q$-th quantile of the BS distribution is directly generated as

$$y^*(q;\alpha,\beta) = \beta\left(\frac{\alpha z(q)}{2} + \sqrt{\left(\frac{\alpha z(q)}{2}\right)^2 + 1}\right)^2, \quad 0 < q \leq 1, \tag{2.2}$$

where $z(q)$ is the $100q$-th quantile of the standard normal distribution. We observe that, if $q = 0.5$, then $z(0.5) = 0$, which corresponds to the mean, median and mode of $Z \sim \mathrm{N}(0,1)$. Thus, from the quantile function defined in (2.2), we have that $y^*(0.5;\alpha,\beta) = \beta$, which confirms that the scale parameter $\beta > 0$ is also the BS median. The expression given in (2.1) is used for generating random numbers of the BS distribution and also for deriving goodness-of-fit tools associated with this distribution.

Let $Y^* \sim \mathrm{BS}(\alpha,\beta)$. Then, the probability density function of Y^* is stated as

$$f_{Y^*}(y;\alpha,\beta) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2\alpha^2}\left(\frac{y}{\beta} + \frac{\beta}{y} - 2\right)\right)\frac{1}{2\alpha\beta}\left(\left(\frac{y}{\beta}\right)^{-1/2} + \left(\frac{y}{\beta}\right)^{-3/2}\right), \tag{2.3}$$

for $y,\alpha,\beta > 0$. Properties of $Y^* \sim \mathrm{BS}(\alpha,\beta)$ are: $bY^* \sim \mathrm{BS}(b\alpha,\beta)$, with $b > 0$; $1/Y^* \sim \mathrm{BS}(\alpha,1/\beta)$. Also, the mean and variance of Y^* are $\mathrm{E}(Y^*) = \beta(1+\alpha^2/2)$ and $\mathrm{Var}(Y^*) = \beta^2\alpha^2(1+5\alpha^2/4)$.

It is worth noting that diverse models related to the BS distribution have been derived. The probability density function formulated in (2.3) corresponds to the traditional BS distribution and this is quite flexible. However, new versions of the BS distribution have been proposed with different properties and shapes even more flexible. For some recent works on these new versions and their mathematical and statistical features, the interested reader is referred to [2] for a mixture Birnbaum–Saunders distribution with applications in biomedicine; to [1] for skew BS distributions with illustrations in fatigue of materials; to [5] for a transmuted BS distribution with examples in problems of engineering and medicine; to [36, 37, 38, 45] for unit and quantile BS distributions considering applications in economy, medicine and politics; to [43] for a bimodal BS distribution with illustrations using environmental and medical data; and to [13] for a BS-gamma distributed claim size applied to insurance.

2.2. RBS distribution

The RBS distribution is a novel model related to the traditional BS case, based on the parameters $\mu,\delta > 0$, where μ is a scale parameter and the distribution mean, whereas δ is a shape and precision parameter. In this case, the notation $Y \sim \mathrm{RBS}(\mu,\delta)$ is used. Based on this reformulation of the BS distribution, the probability density function of the random variable $Y \sim \mathrm{RBS}(\mu,\delta)$ is given by

$$f_Y(y;\mu,\delta) = \frac{\exp(\delta/2)\sqrt{\delta+1}}{4\,y^{3/2}\sqrt{\pi\mu}}\left(y + \frac{\delta\mu}{\delta+1}\right)\exp\left(-\frac{\delta}{4}\left(\frac{y(\delta+1)}{\delta\mu} + \frac{\delta\mu}{y(\delta+1)}\right)\right), \tag{2.4}$$

where $y, \mu, \delta > 0$. The mean and variance of Y are stated as $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \mu^2/\phi$, respectively, with $\phi = (\delta+1)^2/(2\delta+5)$ so that, as mentioned, δ is a precision parameter. Note that, for fixed values of μ, $\mathrm{Var}(Y) \to 0$ as $\delta \to \infty$, but if $\delta \to 0$, then $\mathrm{Var}(Y) \to 5\mu^2$. In addition, $\mathrm{Var}(Y)$ is similar to the variance of the gamma distribution, which has a quadratic relation with its mean. It is also possible to show that $bY \sim \mathrm{RBS}(b\,\mu, \delta)$, with $b > 0$, and $1/Y \sim \mathrm{RBS}(\mu^*, \delta)$, where $\mu^* = (\delta+1)/(\delta\mu)$.

2.3. Modeling

Let $\boldsymbol{Y} = (Y_1, ..., Y_n)^\top$ be independent random variables, where $Y_i \sim \mathrm{RBS}(\mu_i, \delta)$, for $i \in \{1, ..., n\}$, and $\boldsymbol{y} = (y_1, ..., y_n)^\top$ are the corresponding observations of $\boldsymbol{Y}$. Then, we define the RBS-SAM structure based on (2.4) by the systematic component expressed as

$$h(\mu_i) = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + f_1(t_{1_i}) + \cdots + f_s(t_{s_i}), \quad i \in \{1, ..., n\}, \tag{2.5}$$

or, equivalently,

$$h(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{n}_{1_i}^\top \boldsymbol{f}_1 + \cdots + \boldsymbol{n}_{s_i}^\top \boldsymbol{f}_s, \tag{2.6}$$

where $\boldsymbol{x}_i^\top = (1, x_{i_2}, ..., x_{i_p})$ is the i-th row of observed values for the covariates matrix $\boldsymbol{X}$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$ is a vector of regression parameters to be estimated. In addition, $\boldsymbol{n}_{k_i}^\top$ denotes the i-th row of the incidence matrix $\boldsymbol{N}_k$, whose (i,l)-th element is equal to the indicator function $\mathbb{1}\,(t_{k_i} = t_{k_l}^0)$;$\boldsymbol{f}_k = (\xi_{k_1}, ..., \xi_{k_{r_k}})^\top$ is an $r_k \times 1$ vector such that $\xi_{k_j} = f_k(t_{k_j}^0)$, where f_k is an arbitrary unidimensional (scalar) smooth function that quantifies the effect of the k-th covariate t_k for $k \in \{1, ..., s\}$ (note that f_k is different from the vector $\boldsymbol{f}_k$ above defined); and $t_{k_j}^0$, for $l \in \{1, ..., r_k\}$, are the distinct and ordered values of the covariate t_k. Note that the model stated in (2.5) is formed by both parametric and nonparametric components. In effect, for $p < n$, $\boldsymbol{x}_i^\top \boldsymbol{\beta}$ is the parametric component of the model and $\boldsymbol{n}_{1_i}^\top \boldsymbol{f}_1 + \cdots + \boldsymbol{n}_{s_i}^\top \boldsymbol{f}_s$ is the corresponding nonparametric component. In matrix terms, (2.6) can be written as $h(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1\boldsymbol{f}_1 + \cdots + \boldsymbol{N}_s\boldsymbol{f}_s$, where $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)^\top$ and $\mu_i = h^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{n}_{1_i}^\top \boldsymbol{f}_1 + \cdots + \boldsymbol{n}_{s_i}^\top \boldsymbol{f}_s)$, for $i \in \{1, ..., n\}$, with h^{-1} being the inverse of the link function h: $\mathbb{R} \rightsquigarrow \mathbb{R}^+$, which is strictly monotone, positive, and at least twice differentiable; for example, $h(\mu) = \log(\mu)$ or $h(\mu) = \sqrt{\mu}$.

Formally, in the RBS-SAM, we have that $\mathrm{Var}(Y_i)$ is a function of μ_i and, consequently, of the covariates $\boldsymbol{x}_i$. Then, because we are modeling the mean based on a particular structure, we are also modeling the variance due to $\mathrm{Var}(Y_i) = \mu_i^2/\phi$. Therefore, problems where a non-constant variance is present could be analyzed by using this model as well.

2.4. Penalized likelihood function

The RBS-SAM log-likelihood function defined in (2.5) for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{f}_1^\top, ..., \boldsymbol{f}_s^\top, \delta)^\top$ is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\mu_i, \delta; y_i), \tag{2.7}$$

where

$$\ell_i(\mu_i, \delta; y_i) = \frac{\delta}{2} - \frac{\log(16\pi)}{2} - \frac{1}{2}\log\left(\frac{(\delta+1)y_i^3\mu_i}{(\delta\, y_i + y_i + \delta\mu_i)^2}\right) - \frac{y_i(\delta+1)}{4\mu_i} - \frac{\delta^2\mu_i}{4(\delta+1)y_i}.$$

To avoid the problem of overfitting and non-identification of the parameter $\boldsymbol{\beta}$, we can incorporate a penalty term in the log-likelihood function of the model, denoted here by $J(f_k)$, on each smooth function f_k, with f_k belonging to the Sobolev function space defined as $\mathcal{W}_2^{(2)} \equiv \{f_k\colon f_k^{(1)}, f_k^{(2)} \in \mathcal{L}^2[a_k, b_k]\}$, where f_k is an absolutely continuous function and $f_k^{(2)}(t_k) = \mathrm{d}^2 f_k(t_k)/\mathrm{d}t_k^2$. Therefore, in our case, the log-likelihood function given in (2.7) is now expressed as a penalized function stated by

$$\ell_{\mathrm{p}}(\boldsymbol{\theta}, \lambda_1, ..., \lambda_s) \;=\; \ell(\boldsymbol{\theta}) + \sum_{k=1}^{s} \lambda_k^* J(f_k), \tag{2.8}$$

where λ_k is a constant that depends on the smoothing parameter $\lambda_k \geq 0$ that controls the trade-off between goodness-of-fit and the selected smoothness functions.

Different types of penalties have been proposed depending on the method to fit the nonparametric curves. We consider, as a measure of the curvature of the functions, the formula established as

$$J(f_k) = \int_{a_k}^{b_k} f_k^{(2)}(t_k)^2 \mathrm{d}t_k. \tag{2.9}$$

The first term on the right side of (2.8) measures the goodness of fit, whereas the second term, defined by (2.9), penalizes the roughness of each f_k with a fixed parameter λ_k. In this case, the selection of f_k leads to a cubic spline with knots at the points $t_{k_l}^0$, that is, this is a third degree polynomial partitioned on each interval $[t_{k_l}, t_{k_{l+1}}]$, for $l \in \{1, ..., r_k - 1\}$. According to [14], $J(f_k)$ can be written as $J(f_k) = \boldsymbol{f}_k^\top \boldsymbol{K}_k \boldsymbol{f}_k$, where $\boldsymbol{K}_k$ is an $r_k \times r_k$ non-negative-definite matrix that depends only on the knots t_k^0. Then, if we consider $\lambda_k^* = -\lambda_k/2$, the penalized log-likelihood function given in (2.8) may be expressed as

$$\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \ell(\boldsymbol{\theta}) - \sum_{k=1}^{s} \frac{\lambda_k}{2} \boldsymbol{f}_k^\top \boldsymbol{K}_k \boldsymbol{f}_k, \tag{2.10}$$

where $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_s)^\top$ denotes an $s \times 1$ vector of smoothing parameters. Note that an essential aspect of the semi-parametric modeling process is related to the selection of smoothing parameters. In the literature there are several efficient methods of selection, among which cross validation, generalized cross validation, Akaike information criterion (AIC) and mean average squared error can be mentioned.

3. PARAMETERS ESTIMATION AND INFERENCE

In this section, we discuss the process of obtaining MPL estimators, as well as the derivation of a back-fitting algorithm, the penalized score vector, the penalized Hessian matrix, the determination of degrees of freedom, the selection of the smoothing parameter, and the corresponding statistical inference.

3.1. Penalized score function

Assuming that the function given in (2.10) is regular with respect to $\boldsymbol{\beta}$, $\boldsymbol{f}_1, ..., \boldsymbol{f}_s$ and δ, we have that the penalized score function vector of $\boldsymbol{\theta}$ is stated as

$$\boldsymbol{U}_{\mathrm{p}}(\boldsymbol{\theta}) = \frac{\partial \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{U}_{\mathrm{p}}^{\boldsymbol{\beta}}(\boldsymbol{\theta}) \\ \boldsymbol{U}_{\mathrm{p}}^{\boldsymbol{f}_1}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{U}_{\mathrm{p}}^{\boldsymbol{f}_s}(\boldsymbol{\theta}) \\ U_{\mathrm{p}}^{\delta}(\boldsymbol{\theta}) \end{pmatrix},$$

whose elements of the vector have the form $\boldsymbol{U}_{\mathrm{p}}^{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{X}^{\top}\boldsymbol{D}_a\boldsymbol{z}, \boldsymbol{U}_{\mathrm{p}}^{\boldsymbol{f}_k}(\boldsymbol{\theta}) = \boldsymbol{N}_k^{\top}\boldsymbol{D}_a\boldsymbol{z} - \lambda_k\boldsymbol{K}_k\boldsymbol{f}_k$, for $k \in \{1, ..., s\}$, and $U_{\mathrm{p}}^{\delta}(\boldsymbol{\theta}) = \mathrm{tr}\big(\boldsymbol{D}_b\big)$, where $\boldsymbol{D}_a = \mathrm{diag}\{a_1, ..., a_n\}$ and $\boldsymbol{D}_b = \mathrm{diag}\{b_1, ..., b_n\}$ are $n \times n$ matrices, while $\boldsymbol{z} = (z_1, ..., z_n)^{\top}$, with

$$\begin{aligned} z_i &= -\frac{1}{2\mu_i} + \frac{\delta}{(\delta\, y_i + y_i + \delta\mu_i)} + \frac{y_i(\delta+1)}{4{\mu_i}^2} - \frac{\delta^2}{4y_i(\delta+1)}, \\ b_i &= \frac{1}{2} - \frac{1}{2(\delta+1)} + \frac{(y_i+\mu_i)}{(\delta\, y_i + y_i + \delta\mu_i)} - \frac{y_i}{4\mu_i} - \frac{\delta(\delta+2)\mu_i}{4(\delta+1)^2 y_i}, \quad a_i = \frac{1}{h'(\mu_i)}, \end{aligned} \tag{3.1}$$

where h' is the derivative of h. In the RBS-SAM context, the MPL estimators of the model parameters cannot be obtained in an explicit form and need to be calculated by solving a non-linear equation. The Fisher scoring method was used in [27] to estimate the RBS regression parameters. We propose to adjust the RBS-SAM by combining the proposals given in [21, 27] to jointly determine the regression coefficients, the smooth functions, and the precision parameter based on penalized likelihood criterion.

3.2. Penalized Hessian and information matrix

Let $\ddot{\boldsymbol{\ell}}_{\mathrm{p}}(\boldsymbol{\theta})$ be the $p^* \times p^*$ penalized Hessian matrix with its (j^*, l^*)-th element being given by $\partial^2\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})/\partial\theta_{j^*}\theta_{l^*}$, for $j^*, l^* \in \{1, ..., p^*\}$ and $p^* = 1 + p + \sum_{k=1}^{s} r_k$. After algebraic manipulation, we find that the corresponding penalized Hessian matrix has the form

$$\ddot{\boldsymbol{\ell}}_{\mathrm{p}}(\boldsymbol{\theta}) = \frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}} = \begin{pmatrix} \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_1}(\boldsymbol{\theta}) & \cdots & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_s}(\boldsymbol{\theta}) & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_1^{\top}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_1}(\boldsymbol{\theta}) & \cdots & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_s}(\boldsymbol{\theta}) & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{f}_1\delta}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_s^{\top}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_s^{\top}}(\boldsymbol{\theta}) & \cdots & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{f}_s\boldsymbol{f}_s}(\boldsymbol{\theta}) & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{f}_s\delta}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{f}_1\delta}(\boldsymbol{\theta}) & \cdots & \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{f}_s\delta}(\boldsymbol{\theta}) & \ddot{\ell}_{\mathrm{p}}^{\delta\delta}(\boldsymbol{\theta}) \end{pmatrix}, \tag{3.2}$$

whose elements of the matrix can be written as

$$\begin{aligned} \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \boldsymbol{X}^{\top}\boldsymbol{D}_c\boldsymbol{X}, \quad \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_k}(\boldsymbol{\theta}) = \boldsymbol{X}^{\top}\boldsymbol{D}_c\boldsymbol{N}_k, \quad k \in \{1, ..., s\}, \\ \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) &= \boldsymbol{X}^{\top}\boldsymbol{D}_a\boldsymbol{m}, \quad \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{f}_k\delta}(\boldsymbol{\theta}) = \boldsymbol{N}_k^{\top}\boldsymbol{D}_a\boldsymbol{m}, \quad \ddot{\ell}_{\mathrm{p}}^{\delta\delta}(\boldsymbol{\theta}) = \mathrm{tr}(\boldsymbol{D}_d), \\ \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{\boldsymbol{f}_k\boldsymbol{f}_{k'}}(\boldsymbol{\theta}) &= \begin{cases} -\boldsymbol{N}_k^{\top}\boldsymbol{D}_c\boldsymbol{N}_k - \lambda_k\boldsymbol{K}_k, & k = k', \\ \boldsymbol{N}_k^{\top}\boldsymbol{D}_c\boldsymbol{N}_{k'}, & k \neq k', \end{cases} \end{aligned}$$

where $\boldsymbol{D}_a = \text{diag}\{a_1, ..., a_n\}$, $\boldsymbol{D}_c = \text{diag}\{c_1, ..., c_n\}$, $\boldsymbol{D}_d = \text{diag}\{d_1, ..., d_n\}$, and $\boldsymbol{m} = (m_1, ..., m_n)^\top$, with a_i being stated as in (3.1) and

$$\begin{aligned} c_i &= \frac{\partial^2 \ell_i(\mu_i, \delta)}{\partial \mu_i^2}\left(\frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i}\right)^2 + \frac{\partial \ell_i(\mu_i, \delta)}{\partial \mu_i}\left(\frac{\partial}{\partial \mu_i}\frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i}\right)\frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i}, \\ d_i &= \frac{1}{2(\delta+1)^2} - \frac{(y_i+\mu_i)^2}{(\delta\, y_i + y_i + \delta\mu_i)^2} - \frac{\mu_i}{2(\delta+1)^3 y_i}, \\ m_i &= \frac{y_i}{(\delta\, y_i + y_i + \delta\mu_i)^2} + \frac{y_i}{4\mu_i^2} - \frac{\delta(\delta+2)}{4(\delta+1)^2 y_i}, \quad i \in \{1, ..., n\}. \end{aligned}$$

In addition, calculating the expectation of the matrix $-\ddot{\boldsymbol{\ell}}_{\mathrm{p}}(\boldsymbol{\theta})$ given in (3.2), we obtain the $p^* \times p^*$ penalized expected information matrix expressed as

$$\boldsymbol{\mathcal{I}}_{\mathrm{p}}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_1}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_s}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) \\ \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_1^\top}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_1}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_s}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_1\delta}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_s^\top}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_s^\top}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_s\boldsymbol{f}_s}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_s\delta}(\boldsymbol{\theta}) \\ \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_1\delta}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_s\delta}(\boldsymbol{\theta}) & \mathcal{I}_{\mathrm{p}}^{\delta\delta}(\boldsymbol{\theta}) \end{pmatrix}, \tag{3.3}$$

where each element of the matrix can be written as

$$\begin{aligned} \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \boldsymbol{X}^\top \boldsymbol{D}_v \boldsymbol{X}, \quad \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_k}(\boldsymbol{\theta}) = \boldsymbol{X}^\top \boldsymbol{D}_v \boldsymbol{N}_k, \quad k \in \{1, ..., s\}, \\ \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) &= \boldsymbol{X}^\top \boldsymbol{D}_a \boldsymbol{s}, \quad \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_k\delta}(\boldsymbol{\theta}) = \boldsymbol{N}^\top \boldsymbol{D}_a \boldsymbol{s}, \quad \mathcal{I}_{\mathrm{p}}^{\delta\delta}(\boldsymbol{\theta}) = \text{tr}(\boldsymbol{D}_u), \\ \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_k\boldsymbol{f}_{k'}}(\boldsymbol{\theta}) &= \begin{cases} \boldsymbol{N}_k^\top \boldsymbol{D}_v \boldsymbol{N}_k + \lambda_k \boldsymbol{K}_k, & k = k', \\ \boldsymbol{N}_k^\top \boldsymbol{D}_v \boldsymbol{N}_{k'} & k \neq k', \end{cases} \end{aligned}$$

with $\boldsymbol{D}_v = \text{diag}\{v_1, ..., v_n\}$, $\boldsymbol{D}_u = \text{diag}\{u_1, ..., u_n\}$ and $\boldsymbol{s} = (s_1, ..., s_n)^\top$, considering

$$v_i = \frac{\delta a_i^2}{2\mu_i^2} + \frac{\delta^2 a_i^2}{(\delta+1)^2}\mathcal{J}(\boldsymbol{\theta}), \quad u_i = \frac{(\delta^2+3\delta+1)}{2\delta^2(\delta+1)^2} + \frac{\mu_i^2}{(\delta+1)^4}\mathcal{J}(\boldsymbol{\theta}), \quad s_i = \frac{1}{2\mu_i(\delta+1)} + \frac{\delta\mu_i}{(\delta+1)^3}\mathcal{J}(\boldsymbol{\theta}),$$

for $i \in \{1, ..., n\}$, where

$$\begin{aligned} \mathcal{J}(\boldsymbol{\theta}) &= \mathrm{E}\left(\left(Y + \frac{\mu\delta}{(\delta+1)}\right)^{-2}\right) \\ &= \int_0^\infty \frac{\sqrt{\delta+1}\exp(\delta/2)}{4\sqrt{\pi\mu}y^{3/2}}\left(y + \frac{\delta\mu}{\delta+1}\right)^{-2} \exp\left(-\frac{\delta}{4}\left(\frac{(\delta+1)y}{\delta\mu} + \frac{\delta\mu}{(\delta+1)y}\right)\right)\mathrm{d}y. \end{aligned}$$

Note that, in the RBS-SAM, the property of orthogonality between parameters vectors $(\boldsymbol{\beta}, \boldsymbol{f}_k)$ and δ is not verified, unlike what is observed in other models, as the class of GLM, among others. In general, the orthogonality property simplifies the estimation process, in the sense that it allows the parameters to be estimated separately. More details about this issue in the semi-parametric context can be found in [21].

3.3. Finding the solution in practice: Iterative process

To estimate the model parameters by the MPL method, we solve the equation $\boldsymbol{U}_{\mathrm{p}}(\boldsymbol{\theta}) = \boldsymbol{0}$. However, as mentioned, no closed-form expressions for the MPL estimate of $\boldsymbol{\theta}$ are available.

Then, an iterative method for non-linear optimization is needed, such as the Fisher scoring or Newton or quasi-Newton algorithms, where the Broyden–Fletcher–Goldfarb–Shanno method provides often good results. Considering that the matrix $-\ddot{\boldsymbol{\ell}}_{\mathrm{p}}(\boldsymbol{\theta})$ can be non-positive definite, we suggest replacing it with the matrix $-\boldsymbol{\mathcal{I}}_{\mathrm{p}}(\boldsymbol{\theta})$ and using the Fisher scoring method. Then, the algorithm for estimating $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + (\boldsymbol{\mathcal{I}}_{\mathrm{p}}(\boldsymbol{\theta})^{-1})^{(m)}\boldsymbol{U}_{\mathrm{p}}(\boldsymbol{\theta})^{(m)}, \quad m \in \{0, 1, ...\},$$

which is equivalent to solving the matrix equation

$$\begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{D}_v \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{D}_v \boldsymbol{N}_1 & \cdots & \boldsymbol{X}^\top \boldsymbol{D}_v \boldsymbol{N}_s & \boldsymbol{X}^\top \boldsymbol{D}_a \boldsymbol{s} \\ \boldsymbol{N}_1^\top \boldsymbol{D}_v \boldsymbol{X} & \boldsymbol{N}_1^\top \boldsymbol{D}_v \boldsymbol{N}_1 + \lambda_1 \boldsymbol{K}_1 & \cdots & \boldsymbol{N}_1^\top \boldsymbol{D}_v \boldsymbol{N}_s & \boldsymbol{N}_1^\top \boldsymbol{D}_a \boldsymbol{s} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{N}_s^\top \boldsymbol{D}_v \boldsymbol{X} & \boldsymbol{N}_s^\top \boldsymbol{D}_v \boldsymbol{N}_1 & \cdots & \boldsymbol{N}_s^\top \boldsymbol{D}_v \boldsymbol{N}_s + \lambda_s \boldsymbol{K}_s & \boldsymbol{N}_s^\top \boldsymbol{D}_a \boldsymbol{s} \\ \boldsymbol{s}^\top \boldsymbol{D}_a \boldsymbol{X} & \boldsymbol{s}^\top \boldsymbol{D}_a \boldsymbol{N}_1 & \cdots & \boldsymbol{s}^\top \boldsymbol{D}_a \boldsymbol{N}_s & \mathrm{tr}(\boldsymbol{D}_u) \end{pmatrix}^{(m)} \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(m+1,m)} \\ \boldsymbol{\Delta}_{\boldsymbol{f}_1}^{(m+1,m)} \\ \vdots \\ \boldsymbol{\Delta}_{\boldsymbol{f}_s}^{(m+1,m)} \\ \Delta_{\delta}^{(m+1,m)} \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{D}_a \boldsymbol{z} \\ \boldsymbol{N}_1^\top \boldsymbol{D}_a \boldsymbol{z} - \lambda_1 \boldsymbol{K}_1 \boldsymbol{f}_1 \\ \vdots \\ \boldsymbol{N}_s^\top \boldsymbol{D}_a \boldsymbol{z} - \lambda_s \boldsymbol{K}_s \boldsymbol{f}_s \\ \mathrm{tr}(\boldsymbol{D}_b) \end{pmatrix}^{(m)} \tag{3.4}$$

where $\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(m+1,m)} = \boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)}$, $\boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} = \boldsymbol{f}_k^{(m+1)} - \boldsymbol{f}_k^{(m)}$ and $\Delta_{\delta}^{(m+1,m)} = \delta^{(m+1)} - \delta^{(m)}$. Then, after algebraic manipulation, we obtain expressions for the iterative solutions stated as

$$\begin{aligned}
\boldsymbol{\beta}^{(m+1)} &= (\boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \Big(\boldsymbol{\psi}_{\boldsymbol{\beta}}^{(m)} - \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{s} \Delta_{\delta}^{(m+1,m)} - \sum_{k=1}^{s} \boldsymbol{N}_k \boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} \Big), \\
\boldsymbol{f}_k^{(m+1)} &= (\boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N}_k + \lambda_k \boldsymbol{K})^{-1} \boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \\
&\quad \times \Big(\boldsymbol{\psi}_{\boldsymbol{f}_l}^{(m)} - \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{s} \Delta_{\delta}^{(m+1,m)} - \boldsymbol{X} \boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(m+1,m)} - \sum_{k=1, k \neq \ell}^{s} \boldsymbol{N}_k \boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} \Big), \quad k \in \{1, ..., s\}, \\
\delta^{(m+1)} &= \mathrm{tr}^{-1}(\boldsymbol{D}_u^{(m)}) \\
&\quad \times \Big(\mathrm{tr}(\boldsymbol{D}_b^{(m)}) + \mathrm{tr}(\boldsymbol{D}_u^{(m)}) \delta^{(m)} - \boldsymbol{s}^\top \boldsymbol{D}_a^{(m)} \boldsymbol{X} \boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(m+1,m)} - \boldsymbol{s}^\top \boldsymbol{D}_a^{(m)} \sum_{k=1}^{s} \boldsymbol{N}_k \boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} \Big),
\end{aligned}$$

where $\boldsymbol{\psi}_{\boldsymbol{\beta}}^{(m)} = \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{X} \boldsymbol{\beta}^{(m)}$ and $\boldsymbol{\psi}_{\boldsymbol{f}_l}^{(m)} = \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{N}_l \boldsymbol{f}_l^{(m)}$, with $\boldsymbol{D}_{v,a}^{(m)} = \boldsymbol{D}_v^{(m)^{-1}} \boldsymbol{D}_a^{(m)}$.

In general, the system of equations given in (3.4) is consistent and the back-fitting algorithm stated as in (3.5) converges to a solution for any starting values if the $\boldsymbol{D}_v$ weight matrix is symmetric and positive-definite. Additionally, this solution is unique under no concurvity in the data [4, 21].

When δ is known, it is possible to obtain simplified expressions for the iterative solutions of $\boldsymbol{\beta}^{(m+1)}$ and $\boldsymbol{f}_k^{(m+1)}$. Indeed, after some algebraic manipulation, we get that

$$\begin{aligned}
\boldsymbol{\beta}^{(m+1)} &= (\boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \Big(\boldsymbol{\psi}_{\boldsymbol{\beta}}^{(m)} - \sum_{k=1}^{s} \boldsymbol{N}_k \boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} \Big), \\
\boldsymbol{f}_k^{(m+1)} &= (\boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N}_k + \lambda_k \boldsymbol{K})^{-1} \boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \Big(\boldsymbol{\psi}_{\boldsymbol{f}_l}^{(m)} - \boldsymbol{X} \boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(m+1,m)} - \sum_{k=1, k\neq \ell}^{s} \boldsymbol{N}_k \boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} \Big),
\end{aligned}$$

for $k \in \{1, ..., s\}$, or, equivalently,

$$\begin{aligned}
\boldsymbol{\beta}^{(m+1)} &= (\boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \Big(\boldsymbol{r}_{v,a}^{(m)} - \sum_{k=1}^{s} \boldsymbol{N}_k \boldsymbol{f}_k^{(m+1)} \Big), \\
\boldsymbol{f}_k^{(m+1)} &= (\boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N}_k + \lambda_k \boldsymbol{K})^{-1} \boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \Big(\boldsymbol{r}_{v,a}^{(m)} - \boldsymbol{X} \boldsymbol{\beta}^{(m+1)} - \sum_{k=1, k\neq \ell}^{s} \boldsymbol{N}_k \boldsymbol{f}_k^{(m+1)} \Big),
\end{aligned}$$

where $\boldsymbol{r}_{v,a}^{(m)} = \boldsymbol{D}_{(v,a)}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{\eta}^{(m)}$, with $\boldsymbol{\eta}^{(m)} = \boldsymbol{X}\boldsymbol{\beta}^{(m)} + \sum_{k=1}^{s} \boldsymbol{N}_k \boldsymbol{f}_k^{(m)}$. It is possible to prove that these expressions correspond to the weighted back-fitting (Gauss–Seidel) iterations considering $\boldsymbol{r}_{v,a}^{(m)}$ as a dependent modified variable and $\boldsymbol{D}_v$ as a matrix of weights that changes with each iteration of the process. A general formula for these iterations is stated as

$$\boldsymbol{f}_l^{(m+1)} = \boldsymbol{S}_l^{(m)} \Big(\boldsymbol{r}_{v,a}^{(m)} - \sum_{k=0, k\neq l}^{s} \boldsymbol{N}_k \boldsymbol{f}_k^{(m+1)} \Big), \quad l \in \{0, 1, ..., s\}, \tag{3.5}$$

where $\boldsymbol{r}_{v,a}^{(m)} = \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{\eta}^{(m)}$, with $\boldsymbol{\eta}^{(m)} = \sum_{k=0}^{s} \boldsymbol{N}_k \boldsymbol{f}_k^{(m)}$, $\boldsymbol{N}_0 = \boldsymbol{X}$, $\boldsymbol{f}_0 = \boldsymbol{\beta}$, and

$$\boldsymbol{S}_0^{(m)} = (\boldsymbol{N}_0^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N}_0)^{-1} \boldsymbol{N}_0^\top \boldsymbol{D}_v^{(m)}, \boldsymbol{S}_k^{(m)} = (\boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N}_k + \lambda_k \boldsymbol{K}_k)^{-1} \boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)}, \quad k \in \{1, ..., s\}.$$

Note that if the nonparametric component is not present in the model formulated in (2.5), the iterative process presented in (3.5) reduces to the expressions proposed in [27], that is, to $\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{z}_{\boldsymbol{\beta}}^{(m)}$, where $\boldsymbol{z}_{\boldsymbol{\beta}}^{(m)} = \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{X}\boldsymbol{\beta}^{(m)}$. A similar formula can be obtained for the case where the parametric component is absent in the model and only a smooth function is considered as linear predictor. Specifically, the iterative process would reduce to $\boldsymbol{f}^{(m+1)} = (\boldsymbol{N}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N})^{-1} \boldsymbol{N}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{z}_{\boldsymbol{f}}^{(m)}$, where $\boldsymbol{z}_f^{(m)} = \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{N}\boldsymbol{f}^{(m)}$. Thus, at convergence, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{f}}$ can be interpreted as least squares estimators of $\boldsymbol{D}_{\widehat{v}}^{1/2} \widehat{\boldsymbol{z}}_{\boldsymbol{\beta}}$ and $\boldsymbol{D}_{\widehat{v}}^{1/2} \widehat{\boldsymbol{z}}_f$ against the columns of $\boldsymbol{D}_{\widehat{v}}^{1/2} \boldsymbol{X}$ and $\boldsymbol{D}_{\widehat{v}}^{1/2} \boldsymbol{N}$, respectively, where $\widehat{\boldsymbol{z}}_{\boldsymbol{\beta}} = \boldsymbol{D}_{\widehat{v},\widehat{a}} \widehat{\boldsymbol{z}} + \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{z}}_{\boldsymbol{f}} = \boldsymbol{D}_{\widehat{v},\widehat{a}} \widehat{\boldsymbol{z}} + \boldsymbol{N}\widehat{\boldsymbol{f}}$.

3.4. Approximate standard errors

Next, we consider the problem of estimating the variance-covariance matrix of the MPL estimator $\widehat{\boldsymbol{\theta}}$. Considering the fact that we obtain the MPL estimate of $\boldsymbol{\theta}$ through the Fisher scoring algorithm, it is reasonable to derive the corresponding variance-covariance matrix by using the inverse of the penalized Fisher information matrix [20, 50, 51]. To compute the

inverse matrix of $\boldsymbol{\mathcal{I}}_{\mathrm{p}}(\boldsymbol{\theta})$ given in (3.3), consider

$$\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{11} = \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_1}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_s}(\boldsymbol{\theta}) \\ \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_1^\top}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_1}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_s}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_s^\top}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_s^\top}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_s\boldsymbol{f}_s}(\boldsymbol{\theta}) \end{pmatrix}, \quad \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{12} = \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) \\ \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_1\delta}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{\boldsymbol{f}_s\delta}(\boldsymbol{\theta}) \end{pmatrix}, \quad \mathcal{I}_{\mathrm{p}}^{22} = \mathcal{I}_{\mathrm{p}}^{\delta\delta}.$$

Thus, the matrix $\boldsymbol{\mathcal{I}}_{\mathrm{p}}(\boldsymbol{\theta})$ can be written as

$$\boldsymbol{\mathcal{I}}_{\mathrm{p}}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{11} & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{12} \\ \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{12^\top} & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{22} \end{pmatrix}. \tag{3.6}$$

Assuming that all the necessary inverses exist, some algebraic manipulations on the expression stated in (3.6) show that the inverse matrix of $\boldsymbol{\mathcal{I}}_{\mathrm{p}}(\boldsymbol{\theta})$ assumes a block form given by

$$\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{11.1} & -\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{11.1}\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{12}\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{22^{-1}} \\ -\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{22^{-1}}\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{12^\top}\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{11.1} & \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{22.1} \end{pmatrix},$$

where $\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{11.1} = (\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{11} - \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{12}\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{22^{-1}}\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{12^\top})^{-1}$ and $\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{22.1} = \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{22^{-1}} + \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{22^{-1}}\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{12^\top}\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{11.1}\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{12}\boldsymbol{\mathcal{I}}_{\mathrm{p}}^{22^{-1}}$. Therefore, the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\theta}}$ is defined as

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\theta}}) \approx \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{-1}(\boldsymbol{\theta})\big|_{\widehat{\boldsymbol{\theta}}}. \tag{3.7}$$

In particular, we have that $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{f}}_1, ..., \widehat{\boldsymbol{f}}_s) \approx \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{11.1}\big|_{\widehat{\boldsymbol{\theta}}}$ and $\widehat{\mathrm{Cov}}(\widehat{\delta}) \approx \boldsymbol{\mathcal{I}}_{\mathrm{p}}^{22.1}\big|_{\widehat{\boldsymbol{\theta}}}$. To approximate the pointwise standard errors (SE) bands for nonparametric functions f_k, one can evaluate the accuracy of the estimators $\widehat{f}_k$ for different locations within the range of interest [17]. In our case, these SE bands are constructed using the corresponding diagonal elements of the matrix $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{f}}_1, ..., \widehat{\boldsymbol{f}}_s)$ as estimators of the SE of f_k. Indeed, we can consider as approximate pointwise SE bands the expression defined as $\mathrm{SE}(f_k(t_l^0)) = \widehat{f}_k(t_l^0) \pm 2(\widehat{\mathrm{Var}}(\widehat{f}_k(t_l^0)))^{1/2}$, where $\mathrm{Var}(\widehat{f}_k(t_l))$ is the l-th principal diagonal element of the matrix stated in (3.7), for $l \in \{1, ..., r\}$. Note that t_l^0 corresponds to the knots associated with each variable whose contribution to the model is nonparametric.

Observe that the asymptotic approach based on the penalized Fisher information matrix to approximate the variance-covariance matrix of $\widehat{\boldsymbol{\theta}}$ has been used by several authors, and has resulted in an efficient tool to approximate the SEs of $\widehat{\boldsymbol{\theta}}$ and construct SE bands for smooth functions [22, 23, 50].

3.5. Degrees of freedom

Next, we define the degrees of freedom associated with the parametric and nonparametric components. In the latter case, they correspond to the number of effective parameters that are being considered in the nonparametric modeling, and that are used in the selection of the smoothing parameters. Our definition of degrees of freedom is based on the convergence of the iterative process given in (3.5) to select $\widehat{\boldsymbol{f}}_l$, for $l \in \{0, 1, ..., s\}$. Indeed, fixing δ

and λ_k, we obtain $\widehat{\boldsymbol{f}}_l = \widehat{\boldsymbol{S}}_l \widehat{\boldsymbol{r}}^*_{v,a}$, for $l \in \{0, 1, ..., s\}$, where $\widehat{\boldsymbol{r}}^*_{v,a} = \widehat{\boldsymbol{r}}_{v,a} - \sum_{k=0, k\neq l}^{s} \boldsymbol{N}_k \widehat{\boldsymbol{f}}_k$, with $\widehat{\boldsymbol{r}}_{v,a} = \widehat{\boldsymbol{D}}_{v,a}\widehat{\boldsymbol{z}} + \widehat{\boldsymbol{\eta}}$, $\widehat{\boldsymbol{\eta}} = \sum_{k=0}^{s} \boldsymbol{N}_k \widehat{\boldsymbol{f}}_k$ and

$$\begin{aligned} \widehat{\boldsymbol{S}}_0 &= (\boldsymbol{N}_0^\top \widehat{\boldsymbol{D}}_v \boldsymbol{N}_0)^{-1} \boldsymbol{N}_0^\top \widehat{\boldsymbol{D}}_v, \\ \widehat{\boldsymbol{S}}_k &= (\boldsymbol{N}_k^\top \widehat{\boldsymbol{D}}_v \boldsymbol{N}_k + \lambda_k \boldsymbol{K}_k)^{-1} \boldsymbol{N}_k^\top \widehat{\boldsymbol{D}}_v, \quad k \in \{1, ..., s\}. \end{aligned}$$

In the literature concerning to additive models, there are different definitions for the degrees of freedom, depending on the context in which they are used [6]. In our case, the degrees of freedom associated with the parametric component, $\boldsymbol{N}_0 \widehat{\boldsymbol{f}_0}$ namely, are defined as

$$\mathrm{df}_{\boldsymbol{N}_0} = \mathrm{tr}\big\{\boldsymbol{N}_0 (\boldsymbol{N}_0^\top \widehat{\boldsymbol{D}}_v \boldsymbol{N}_0)^{-1} \boldsymbol{N}_0^\top \widehat{\boldsymbol{D}}_v\big\} = p,$$

where p is the rank of $\boldsymbol{N}_0 = \boldsymbol{X}$. In addition, the degrees of freedom for the nonparametric component, $\boldsymbol{N}_k \widehat{\boldsymbol{f}}_k$ namely, are stated as

$$\mathrm{df}(\lambda_k) = \mathrm{tr}\big\{\boldsymbol{N}_k (\boldsymbol{N}_k^\top \widehat{\boldsymbol{D}}_v \boldsymbol{N}_k + \lambda_k \boldsymbol{K}_k)^{-1} \boldsymbol{N}_k^\top \widehat{\boldsymbol{D}}_v\big\}, \tag{3.8}$$

which measure the individual effect contribution of the k-th nonparametric component.

3.6. Selection of smoothing parameters

Note that, based on [23], we consider a grid of values for the smoothing parameter and select one of them for which the AIC is minimized. Alternatively, we can consider the inverse relationship between the degrees of freedom and the smoothing parameter [23], and select the value of this parameter associated with a specific value of the degrees of freedom.

In the previous sections, the smoothing parameters λ_k were assumed fixed. However, in practice situations, the smoothing parameters should be selected from the data. A criterion to select the smoothing parameters based on the AIC is described below. Before proposing a criterion to select the smoothing parameters, remember that, under the RBS-SAM, we have a total of $1 + p + \mathrm{df}(\boldsymbol{\lambda})$ parameters to be determined, with $\mathrm{df}(\boldsymbol{\lambda}) = \sum_{k=1}^{s} \mathrm{df}(\lambda_k)$ denoting the number of effective parameters involved in the modeling of the smooth functions. In this case, the AIC (or alternatively the Bayesian information criterion) can be used for selecting the smoothing parameters λ_k. The idea is to minimize a function with respect to $\boldsymbol{\lambda}$ formulated as

$$\mathrm{AIC}(\boldsymbol{\lambda}) = -2\ell_{\mathrm{p}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + 2(1 + p + \mathrm{df}(\boldsymbol{\lambda})),$$

where $\ell_{\mathrm{p}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda})$ denotes the penalized log-likelihood function evaluated at $\widehat{\boldsymbol{\theta}}$ for a fixed $\boldsymbol{\lambda}$. A grid (surface) for different values of $\boldsymbol{\lambda}$ and its corresponding $\mathrm{AIC}(\boldsymbol{\lambda})$ are helpful to choose the suitable smoothing parameters.

Other approach to selecting the smoothing parameters is when the degrees of freedom given in (3.8) depends only on λ_k and, therefore, the corresponding smoothing parameter can be specified. In other words, we specify an objective $\mathrm{df}(\lambda_k)$ for a function and then find the value λ_k that achieves this objective. Such an approach has been used by a number of authors [6, 18, 22, 44].

4. DIAGNOSTIC ANALYTICS

In this section, we derive the local influence method for different perturbation schemes to assess the potential influence of some observations on the RBS-SAM. These schemes are the case-weight, response and precision parameter perturbations.

4.1. The local influence method

Let $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^\top$ be an $n \times 1$ vector of perturbations restricted to some open subset $\Omega \in \mathbb{R}^n$ and $\ell_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})$ be the logarithm of the perturbed penalized likelihood function. It is assumed that exists $\boldsymbol{\omega}_0 \in \Omega$, a vector of no perturbation, such that $\ell_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega}_0) = \ell_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\lambda})$. To assess the influence of small perturbations on the MPL estimate $\widehat{\boldsymbol{\theta}}$, we can consider the likelihood displacement stated as

$$\mathrm{LD}(\boldsymbol{\omega}) = 2\Big(\ell_{\rm p}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) - \ell_{\rm p}(\widehat{\boldsymbol{\theta}}_\omega, \boldsymbol{\lambda})\Big) \geq 0,$$

where $\widehat{\boldsymbol{\theta}}_\omega$ is the MPL estimate under $\ell_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})$. The measure $\mathrm{LD}(\boldsymbol{\omega})$ is helpful for assessing the distance between $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_\omega$. In [9], it was suggested to study the local behavior of $\mathrm{LD}(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$. The procedure consists of selecting a unit direction $\boldsymbol{d} \in \Omega$, with $\|\boldsymbol{d}\| = 1$, and then to consider the plot of $\mathrm{LD}(\boldsymbol{\omega}_0 + a\boldsymbol{d})$ against a, where $a \in \mathbb{R}$. This plot is called lifted line. Each lifted line may be characterized by considering the normal curvature $C_d(\boldsymbol{\theta})$ around $a = 0$. The suggestion is to assume the direction $\boldsymbol{d} = \boldsymbol{d}_{\max}$ corresponding to the largest curvature $C_{d_{\max}}(\boldsymbol{\theta})$. The index plot of $\boldsymbol{d}_{\max}$ can identify those cases that, under small perturbations, provoke an importante potential influence on $\mathrm{LD}(\boldsymbol{\omega})$. According to [9], the normal curvature at the unitary direction $\boldsymbol{d}$ is expressed as

$$C_d(\boldsymbol{\theta}) = -2\big(\boldsymbol{d}^\top \boldsymbol{\Delta}_{\rm p}^\top \ddot{\boldsymbol{\ell}}_{\rm p}^{-1} \boldsymbol{\Delta}_{\rm p} \boldsymbol{d}\big),$$

with

$$\ddot{\boldsymbol{\ell}}_{\rm p} = \frac{\partial^2 \ell_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \quad \boldsymbol{\Delta}_{\rm p} = \frac{\partial^2 \ell_{\rm p}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\, \boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

Note that $-\ddot{\boldsymbol{\ell}}_{\rm p}$ is the penalized observed information matrix evaluated at $\widehat{\boldsymbol{\theta}}$ (see Subsection 3.2) and $\boldsymbol{\Delta}_{\rm p}$ is the penalized perturbation matrix evaluated at $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$. Observe that $C_d(\boldsymbol{\theta})$ denotes the local influence on the estimate $\widehat{\boldsymbol{\theta}}$ after perturbing the model or data. In [11], it was proposed to study the normal curvature at the direction $\boldsymbol{d} = \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is an $n \times 1$ vector with a one at the i-th position and zeros at the remaining positions. Thus, the normal curvature, called total local influence of the i-th case, assumes the form $C_{\boldsymbol{e}_i}(\boldsymbol{\theta}) = 2|c_{ii}|$, for $i \in \{1, ..., n\}$, where c_{ii} is the i-th principal diagonal element of the matrix $\boldsymbol{C} = \boldsymbol{\Delta}_{\rm p}^\top \ddot{\boldsymbol{\ell}}_{\rm p}^{-1} \boldsymbol{\Delta}_{\rm p}$.

The local influence results presented in the following sections are extensions of the work presented in [27] but for the case where the linear predictor includes a parametric term and also an additive term of smooth functions.

4.2. Case-weight perturbation

The perturbation of case-weight is considered to identify observations with high contribution to the likelihood function and that can exercise strong influence on the MPL estimates. Let us to consider the attributed weights for the cases in the penalized log-likelihood function as

$$\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \ell_i(\boldsymbol{\theta}) - \sum_{k=1}^{s} \frac{\lambda_k}{2} \boldsymbol{f}_k^{\top} \boldsymbol{K}_k \boldsymbol{f}_k,$$

where $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^{\top}$ is the vector of weights, with $0 \leq \omega_i \leq 1$, for $i \in \{1, ..., n\}$, and $\boldsymbol{\omega}_0 = (1, ..., 1)^{\top}$ denotes the vector of no perturbation. Differentiating $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})$ with respect to the elements of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, we obtain

$$\begin{aligned}
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}}\Big|_{\theta=\widehat{\theta},\ \omega=\omega_0} &= \boldsymbol{X}^{\top} \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_{\boldsymbol{z}}, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{f}_k \partial \boldsymbol{\omega}}\Big|_{\theta=\widehat{\theta},\ \omega=\omega_0} &= \boldsymbol{N}_k^{\top} \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_{\boldsymbol{z}}, \quad k \in \{1, ..., s\}, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \delta \partial \boldsymbol{\omega}}\Big|_{\theta=\widehat{\theta},\omega=\omega_0} &= \widehat{\boldsymbol{b}},
\end{aligned}$$

for $i \in \{1, ..., n\}$, with $\boldsymbol{D}_a$ and $\boldsymbol{b}$ being defined in previous sections, whereas that $\boldsymbol{D}_{\boldsymbol{z}} = \mathrm{diag}\{z_1, ..., z_n\}$.

4.3. Response perturbation

According to [27], the additive perturbation on the i-th response is given by $y_{i\omega_i} = y_i + \omega_i s(y_i)$, where $s(y_i) = (\widehat{\mu}_i^2/\widehat{\phi})^{1/2}$ and $\omega_i \in \mathbb{R}$, for $i \in \{1, ..., n\}$. Then, the perturbed penalized log-likelihood function is constructed from the expression defined in (2.10) with y_i being replaced by $y_{i\omega}$, that is,

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

where ℓ is given in (2.7) with $y_{i\omega_i}$ in the place of y_i, for $i \in \{1, ..., n\}$,. Here, the vector of no perturbation is stated as $\boldsymbol{\omega}_0 = (0, ..., 0)^{\top}$. Differentiating $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})$ with respect to the elements of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, we obtain, after some algebraic manipulation, that

$$\begin{aligned}
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}}\Big|_{\theta=\widehat{\theta},\ \omega=\omega_0} &= \boldsymbol{X}^{\top} \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_{\psi} \widehat{\boldsymbol{D}}_{\vartheta}, \quad \frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \delta \partial \boldsymbol{\omega}}\Big|_{\theta=\widehat{\theta},\ \omega=\omega_0} = \widehat{\boldsymbol{\tau}}^{\top} \boldsymbol{D}_{\vartheta}, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{f}_k \partial \boldsymbol{\omega}}\Big|_{\theta=\widehat{\theta},\ \omega=\omega_0} &= \boldsymbol{N}_k^{\top} \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_{\psi} \widehat{\boldsymbol{D}}_{\vartheta}, \quad k \in \{1, ..., s\},
\end{aligned}$$

where $\widehat{\boldsymbol{D}}_{\vartheta} = \mathrm{diag}\{\widehat{\vartheta}_1, ..., \widehat{\vartheta}_n\}$, $\widehat{\boldsymbol{D}}_{\psi} = \mathrm{diag}\{\widehat{\psi}_1, ..., \widehat{\psi}_n\}$, and $\widehat{\boldsymbol{D}}_{\tau} = \mathrm{diag}\{\widehat{\tau}_1, ..., \widehat{\tau}_n\}$, with $\widehat{\vartheta}_i = s(y_i)$,

$$\begin{aligned}
\widehat{\psi}_i &= -\frac{\widehat{\delta}(\widehat{\delta}+1)}{(\widehat{\delta} y_i + y_i + \widehat{\delta}\widehat{\mu}_i)^2} + \frac{(\widehat{\delta}+1)}{4\widehat{\mu}_i^2} + \frac{\widehat{\delta}^2}{4(\widehat{\delta}+1) y_i^2}, \\
\widehat{\tau}_i &= -\frac{\widehat{\mu}_i}{(\widehat{\delta} y_i + y_i + \widehat{\delta}\widehat{\mu}_i)^2} - \frac{1}{4\widehat{\mu}_i} + \frac{\widehat{\delta}(\widehat{\delta}+2)\widehat{\mu}_i}{4 y_i^2 (\widehat{\delta}+1)^2}, \quad i \in \{1, ..., n\}.
\end{aligned}$$

4.4. Perturbation on the precision parameter

Perturbation of the precision parameter is used for evaluating the sensitivity of the MPL estimate to small modifications of δ. Initially, the RBS-SAM assumes that the precision parameter is constant across data. However, under the perturbed model, the precision parameter is non-constant across cases, this is, $Y_i \sim \mathrm{RBS}\big(\mu_i, \delta_i\big)$, where $\delta_i = \delta/\omega_i$, with $\omega_i > 0$, for $i \in \{1, ..., n\}$. Under this perturbation, the vector of no perturbation is given by $\boldsymbol{\omega}_0 = (1, ..., 1)^\top$. Then, the perturbed penalized log-likelihood function is constructed from the expression given in (2.10) with δ being replaced by δ_i. Taking derivatives of $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})$ with respect to the elements of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, we obtain, after some algebraic manipulation, that

$$\begin{aligned}
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}}\Big|_{\theta=\widehat{\theta},\ \omega=\omega_0} &= \boldsymbol{X}^\top \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_\varpi, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{f}_k \partial \boldsymbol{\omega}}\Big|_{\theta=\widehat{\theta},\ \omega=\omega_0} &= \boldsymbol{N}_k^\top \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_\varpi, \quad k \in \{1, ..., s\}, \\
\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \delta \partial \boldsymbol{\omega}}\Big|_{\theta=\widehat{\theta},\quad \omega=\omega_0} &= \widehat{\boldsymbol{\varphi}}^\top,
\end{aligned}$$

where $\widehat{\boldsymbol{D}}_\varpi = \mathrm{diag}\{\widehat{\varpi}_1, ..., \widehat{\varpi}_n\}$ and $\boldsymbol{\varphi} = (\widehat{\varphi}_1, ..., \widehat{\varphi}_n)^\top$, with

$$\begin{aligned}
\varpi_i &= -\frac{\widehat{\delta} y_i}{(\widehat{\delta} y_i + y_i + \widehat{\delta}\widehat{\mu}_i)^2} - \frac{\widehat{\delta} y_i}{4\widehat{\mu}_i^2} + \frac{\widehat{\delta}^2(\widehat{\delta}+2)}{4y_i(\widehat{\delta}+1)^2}, \quad i \in \{1, ..., n\}, \\
\widehat{\varphi}_i &= -\frac{1}{2} + \frac{1}{2(\widehat{\delta}+1)^2} - \frac{y_i(y_i+\widehat{\mu}_i)}{(\widehat{\delta} y_i + y_i + \widehat{\delta}\widehat{\mu}_i)^2} + \frac{y_i}{4\widehat{\mu}_i} + \frac{\widehat{\delta}^2\widehat{\mu}_i(\widehat{\delta}+3)}{4y_i(\widehat{\delta}+1)^3} + \frac{\widehat{\delta}\widehat{\mu}_i}{y_i(\widehat{\delta}+1)^3}.
\end{aligned}$$

5. APPLICATION TO REAL POLLUTION DATA

In this section, we provide the empirical application of the proposed model to environmental data obtained from the website of the Chilean Ministry of Environment using the `R` software. This application is motivated, as mentioned, by the fact that inclusion of nonparametric functions greatly enhances the modeling when accommodating non-linear effects of covariates. These covariates correspond in our case to contents of pollutants and meteorological variables as atmospheric pressure, precipitation, relative humidity, temperature, and wind speed, with the response variable being the particulate matter (PM) content.

5.1. Data and definition of the problem

PM pollution is one of the main global urban environmental problems, affecting human health and life quality. According to the World Health Organization (WHO), nine out of every ten people on the planet breathe air that contains high levels of pollutants and seven million people die every year due to this cause (`www.who.int`). PM is classified according to its diameter, because particle size determines sites of deposition within the respiratory tract.

Coarser particles (those with a diameter over 10 micrometers–μm–) do not penetrate into airways. Instead, these particles are deposited in the upper respiratory tract and are cleared by cilia action. Inhalable particles measuring less than 10 μm are called PM10, whereas those smaller than 2.5 μm are called PM2.5. As size decreases, there is a higher possibility for PM to penetrate deeper into smaller alveoli and airways. Specifically, various effects are produced from exposure to PM, but the nature of those induced effects vary according to the PM composition. Indeed, there is evidence of an increase in the risk of cardiovascular diseases and mortality from exposure to PM2.5, which occurs even after short time periods [8].

In Chile, according to the Ministry of the Environment, 3,494 people died prematurely due to critical air levels during 2017 mainly due to extreme contents of PM2.5, and nine million inhabitants are exposed to levels of pollution that exceeds the air quality standards (`https://bit.ly/2u40gDq`). Santiago of Chile is one of the most polluted cities in the world in terms of PM2.5 and PM10, because of a combination of anthropogenic, meteorological and topographic factors [31]. Several studies indicate that inhabitants of Santiago are under risk, because of the city's poor air quality. As urban air quality declines, the risk of stroke, heart disease, lung cancer, and chronic and acute respiratory diseases, including asthma, increases for the people who live in cities with high air pollution levels. Specifically for Santiago, several investigations provide evidence that exposure to air pollutants produces a risk to the inhabitants of this city; see [34] and references therein.

Periodic episodes of extreme levels of air pollution sometimes occur for certain atmospheric contaminants. Such episodes and their associated high contents vary with geographical and meteorological fluctuations, depending on changes in both source and type of emissions. Because of these variations, PM contents are treated as non-negative random variables that can be modeled by statistical distributions. Frequently, these distributions are asymmetrical and present positive skewness [31]. The current official methodology employed by the Chilean authority in Santiago to predict PM10 contents is based on a multiple regression model using contents of atmospheric pollutants and meteorological variables as covariates [39]. It helps to forecast the maximum value of the 24-hour average content of PM10 in μg/normalized cubic meters (Nm^3) for the period from 00:00 to 24:00 hours of the next day. Furthermore, in 2015, through Supreme Decree number 15/2015 and resolution number 9664/2015, it was instructed by the Chilean Ministry of Health to declare sanitary alert employing also PM2.5 contents.

5.2. Descriptive data analysis

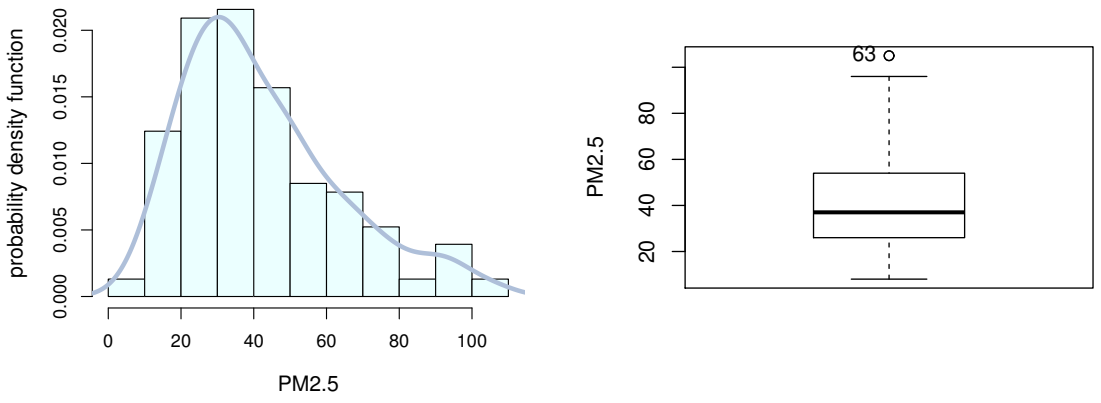
We consider the environmental data set related to air pollution. In particular, the data provided by the National Air Quality Information System (`https://sinca.mma.gob.cl`) corresponding to air pollution in the commune of Pudahuel in Santiago during the critical episodes management (CEM) period (01 April 2019 to 31 August 2019). In this application, we are interested on detecting the association of polluting contents with meteorological variables by utilizing the RBS-SAM. For motivating the semi-parametric models, we consider PM2.5 as response variable and two covariates, PM10 contents and WIND (speed wind in meters/second). In the CEM period, we work with a total of 153 observations.

Table 1 provides a descriptive summary of PM2.5, which includes mean ($\overline{y}$), median (MD), standard deviation (SD), coefficient of variation (CV), coefficient of skewness (CS), coefficient of kurtosis (CK), minimum ($y_{(1)}$), maximum ($y_{(n)}$), and the total of observations (n). Figure 1 contains the histogram and boxplot of PM2.5 contents. The primary air quality regulation for PM2.5 is 50μg/Nm^3, as 24-hour level.

Table 1: Descriptive statistics for the response variable PM2.5 in the CEM period with data from Santiago, Chile.

$\overline{y}$	MD	SD	CV	CS	CK	$y_{(1)}$	$y_{(n)}$	n
42.27	37	21.51	0.51	0.89	3.26	8	105	153

According to Table 1, the primary regulation is exceeded for the response; see, for example, the maximum contents registered in the CEM period. Note that CS = 0.89, indicating a slight asymmetry in the data of the response variable, and its CK = 3.26, indicating a probability density function with heavy tails in relation to the normal distribution. In addition, from the histogram displayed in Figure 1, we note that the values of PM2.5 have an empirical distribution that is positively skewed, while from Figure 1(b), we identify case #63 as an atypical observation in the boxplot. Consequently, from Table 1 and Figure 1, we propose that the RBS-SAM may be suitable for describing the mean of the data, the non-constant variance and asymmetry detected in the distribution of these data.



(a) **(b)**

Figure 1: Histogram (a) and boxplot (b) for the response variable PM2.5 in CEM period with data from Santiago, Chile.

Figure 2 contains the scatter plots between the response variable and each covariate. In Figure 2(a), we see that the linear correlation between PM2.5 and the covariate PM10 seems to be positive, which is supported by the correlation coefficient between both variables (0.9261). In addition, we observe that the variability of PM2.5 tends to increase as the PM10 values increase, which could be an indication of a non-constant variance in the data.

Figure 2(a) also shows evidence that the straight line do not go through the origin, so that considering the intercept in the parametric component of the selected model could be helpful. Figure 2(b) indicates that the relationship between PM2.5 and WIND seems to be non-linear.

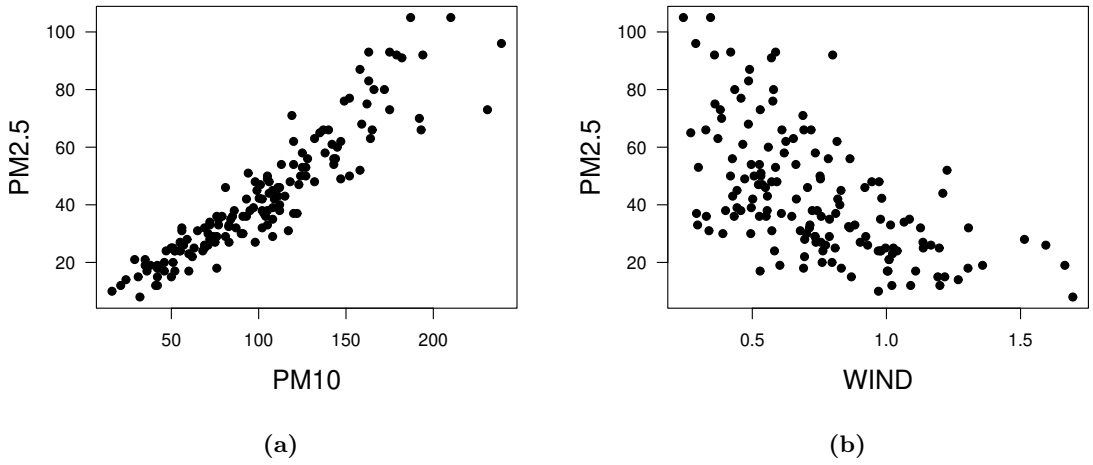


Figure 2: Scatter plots of PM2.5 versus PM10 (a) and PM2.5 versus WIND (b) with data from Santiago, Chile.

After this descriptive data analysis, we can see that the response variable presents characteristics of the RBS distribution, such as positive asymmetry. In addition, as mentioned, the BS distribution has mathematical arguments that allows us to justify its use when describing environmental data. Furthermore, from the scatter plots, note that the covariates contribute linearly and non-linearly to the response variable. Therefore, a semi-parametric structure may be proposed to model PM2.5 contents as a function of PM10 and WIND, contributing parametrically and nonparametrically to the model, respectively.

5.3. Model fitting

The trends described in the exploratory data analysis suggest an RBS-SAM among PM2.5 and the covariates, assuming the identity link function, that is,

$$h(\mu_i) = \mu_i = \beta_1 + x_i\beta_2 + f(t_i), \quad i \in \{1, ..., 153\},$$

where x_i and t_i denotes the values of PM10 and WIND from the i-th case, respectively, $(\beta_1, \beta_2)^\top$ is a parameter vector, and f is a smooth function. We apply the procedure described in Subsection 3.6 to select the smoothing parameter, which results to be $\lambda = 0.0001$.

To estimate the parameters of the parametric component of the model, we maximize the penalized log-likelihood function as described in Subsection 3.3, obtaining $\widehat{\beta}_1 = -30.013$, $\widehat{\beta}_2 = 0.387$ and $\widehat{\delta} = 60.991$, whose corresponding SEs are 0.000095, 0.0068 and 0.0000011, respectively. To assess that the RBS-SAM is adequate to describe the mean of the response variable, we verify the assumptions established for the model. First, note that β_1 and β_2 are highly significant at 5%, since both empirical p-values (omitted here) are close to zero, as expected by the exploratory data analysis that we performed. Thus, the parametric component of the selected model seems to be adequate.

5.4. Residual analysis

Figure 3 shows the plot of the partial residuals defined as

$$r_i^{(\mathrm{p})} = \widehat{\mu}_i - (\widehat{\beta}_1 + x_i\widehat{\beta}_2), \quad i \in \{1, ..., n\}.$$

We note that the effect of the WIND covariate seems to be non-linear on the mean of the PM2.5 response variable, which indicates that it is reasonable to quantify such an effect through a smooth function.

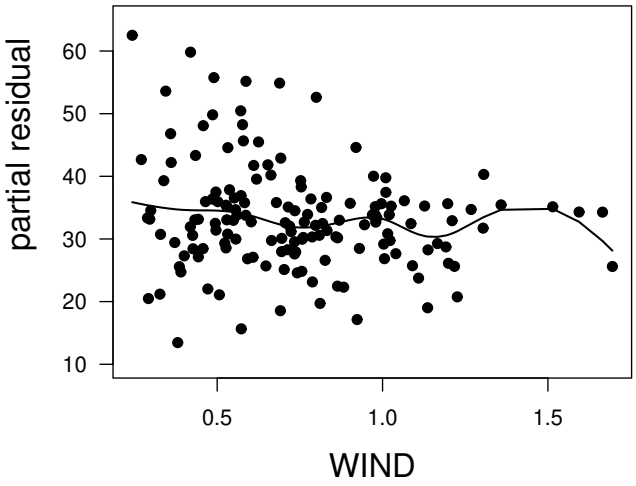


Figure 3: Plot of partial residuals versus WIND covariate with the estimated smooth function overlapped with data from Santiago, Chile.

Figure 4 contains the theoretical quantile versus empirical quantile (QQ) plot (a), histogram (b) and index plot (c) for the standardized residuals defined as

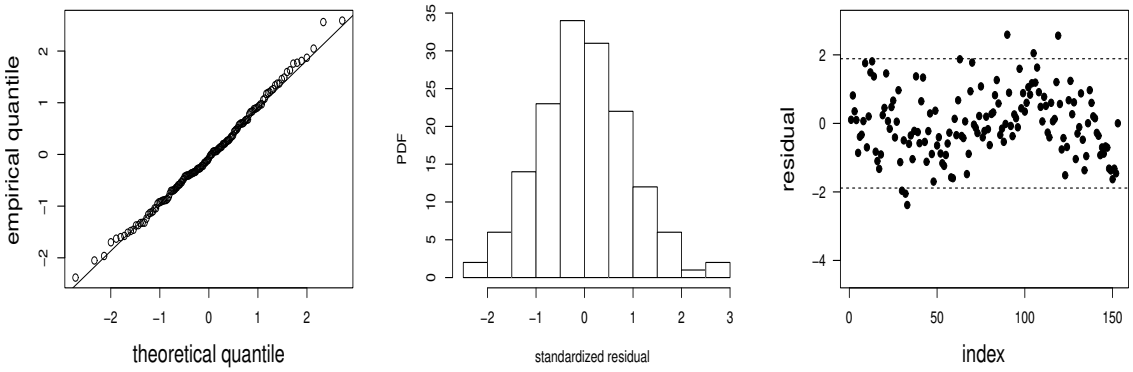
$$r_i^{(\mathrm{s})} = \frac{y_i - \mu_i}{\sqrt{\widehat{\mathrm{Var}}(Y_i)}} = \frac{\widehat{\phi}^{1/2}(y_i - \widehat{\mu}_i)}{\sqrt{\widehat{\mu}_i^2}}, \quad i \in \{1, ..., n\},$$

with ϕ and μ_i being stated in Subsections 2.2, for the model formulated in (2.5).

To verify the distributional assumption established in the model, we consider Figure 4(a). This figure does not show inusual features, so that the response variable seem to be well described by the RBS-SAM. In addition, the independence assumption is also verified. In Figure 4(b), we see a considerable symmetry of the standardized residuals. From Figure 4(c), we note that the standardized residuals take values in the interval $[-2.0, 2.0]$ mostly, except for some values that are outside the bands defined as two times the standard deviation for each residual.

For comparative purposes, we also consider a generalized additive model (GAM) [15, 16] assuming a Gaussian family with link function equal to identity to describe the data set of

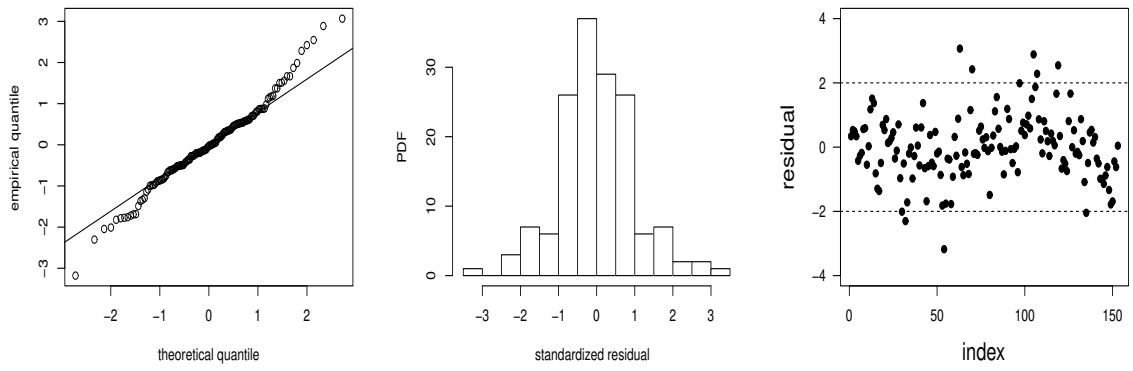
pollution. The estimates of the parametric part of the model are $\widehat{\beta}_1 = 0.668$ and $\widehat{\beta}_2 = 0.428$, whose estimated SEs are 1.956 and 0.018, respectively, with a total of three degrees of freedom associated with this fit.



(a) **(b)** **(c)**

Figure 4: QQ plot (a), histogram (b), and index plot (c) for the standardized residuals of RBS-SAM with data from Santiago, Chile.

Figure 5 shows a residual analysis of the model. Observe the behavior of the tails of the distribution in Figure 5(a) and 5b. In addition, in Figure 5(c), note that there are several points outside the bands defined as twice the standard deviation of the residuals. The corresponding AIC of the fitted model is 1081.612. According to the AIC, the RBS-SAM presents a better fit, since the value of the AIC is 1032.003, with this value being less than the value obtained for the adjusted GAM.



(a) **(b)** **(c)**

Figure 5: QQ-plot (a), histogram (b), and index plot (c) for the standardized residuals of GAM with data from Santiago, Chile.

5.5. Local influence analysis

Figures 6 (a)–(c) present the index plots of C_i for the case-weight scheme. Considering Figure 6 (a)–(b), we notice that cases #9 and #151 have the highest potential influence on $\widehat{\boldsymbol{\beta}}$ and $\widehat{\delta}$, whereas case #63 is more influential on $\widehat{\boldsymbol{f}}$; see Figure 6(c).

Figures 6 (d)–(f) display the index plot of C_i for the response perturbation scheme. We observe that case #151 has a small influence on $\widehat{\boldsymbol{\beta}}$ and $\widehat{\delta}$, whereas that cases #74 and #151 have the highest potential influence on $\widehat{\boldsymbol{f}}$.

Figures 6 (g)–(i) show the index plot of C_i for the precision perturbation scheme. We see that cases #9 and #151 have the highest potential influence on $\widehat{\boldsymbol{\beta}}$ and $\widehat{\delta}$, whereas case #63 is the more influential on $\widehat{\boldsymbol{f}}$; see Figure 6(i). Note that cases #9, #63, #74 and #151 correspond to 09 April, 02 June, 13 June, and 29 August 2019, respectively.

We now analyze how the MPL estimates change when the cases detected as potentially influential are removed. To perform this analysis, the set of case(s) $\{9\}$, $\{63\}$, $\{151\}$, $\{9, 63\}$, $\{9, 151\}$, $\{63, 151\}$, and $\{9, 63, 151\}$ are removed and the estimates of the parameters are calculated again. Table 2 provide the relative changes (RC) in the parameter estimates, in their corresponding estimated SEs, and the associated p-value. These changes are calculated from $\mathrm{RC}_{\widehat{\beta}_{j(i)}} = |(\widehat{\beta}_j - \widehat{\beta}_{j(i)})/\widehat{\beta}_j| \times 100\%$ and $\mathrm{RC}_{\mathrm{SE}(\widehat{\beta}_j)_{(i)}} = |(\mathrm{SE}(\widehat{\beta}_j) - \mathrm{SE}(\widehat{\beta}_j)_{(i)})/\mathrm{SE}(\widehat{\beta}_j)| \times 100\%$, where $\widehat{\beta}_{j(i)}$ and $\mathrm{SE}(\widehat{\beta}_j)_{(i)}$ denote the MPL estimates of β_j and their corresponding SEs, obtained after extracting case i, for $j \in \{1, 2\}$ and $i \in \{1, ..., 153\}$. Table 2 reports that the highest values of RCs are associated with β_1, in particular, for cases $\{9\}$ and $\{9, 63\}$, corresponding to the most influential observations on the parametric, nonparametric and precision components. Note that the influential cases on the parametric component are not necessarily the same on the nonparametric component, as reported in [24]. In addition, note that the significance of the parameters at 5% does not change since the p-values remain below 0.01. Table 2 reports that the diagnostic measures derived here identify potentially influential cases, being them cases $\{9, 63\}$, which affect the inference of the model, but not their significance. These cases correspond to the dates: 09 April 2019 and 02 June 2019 of the CEM period. Note that, for case #9, a low PM2.5 content was recorded, close to the minimum, as well as for PM10, while the wind speed was high and close to the maximum recorded. Regarding case #63, this was the day with maximum PM2.5 content recorded throughout the CEC period, while on this day the wind speed recorded the minimum measurement. In addition, for case #151, the PM2.5 content registered its minimum value in the CEC period and the PM10 content was very close to the minimum value recorded, while the wind speed reached its maximum value. Considering the above results, the maximum or minimum PM2.5 content for cases #9, #63 and #151 are strongly related to wind speed, since on those days this speed reached extreme measurements. In summary, the diagnostic analytics based on the local influence method and residuals confirm that the RBS-SAM presented in Section 2 is suitable for modeling environmental data, even if there are outliers and potentially influential observations.

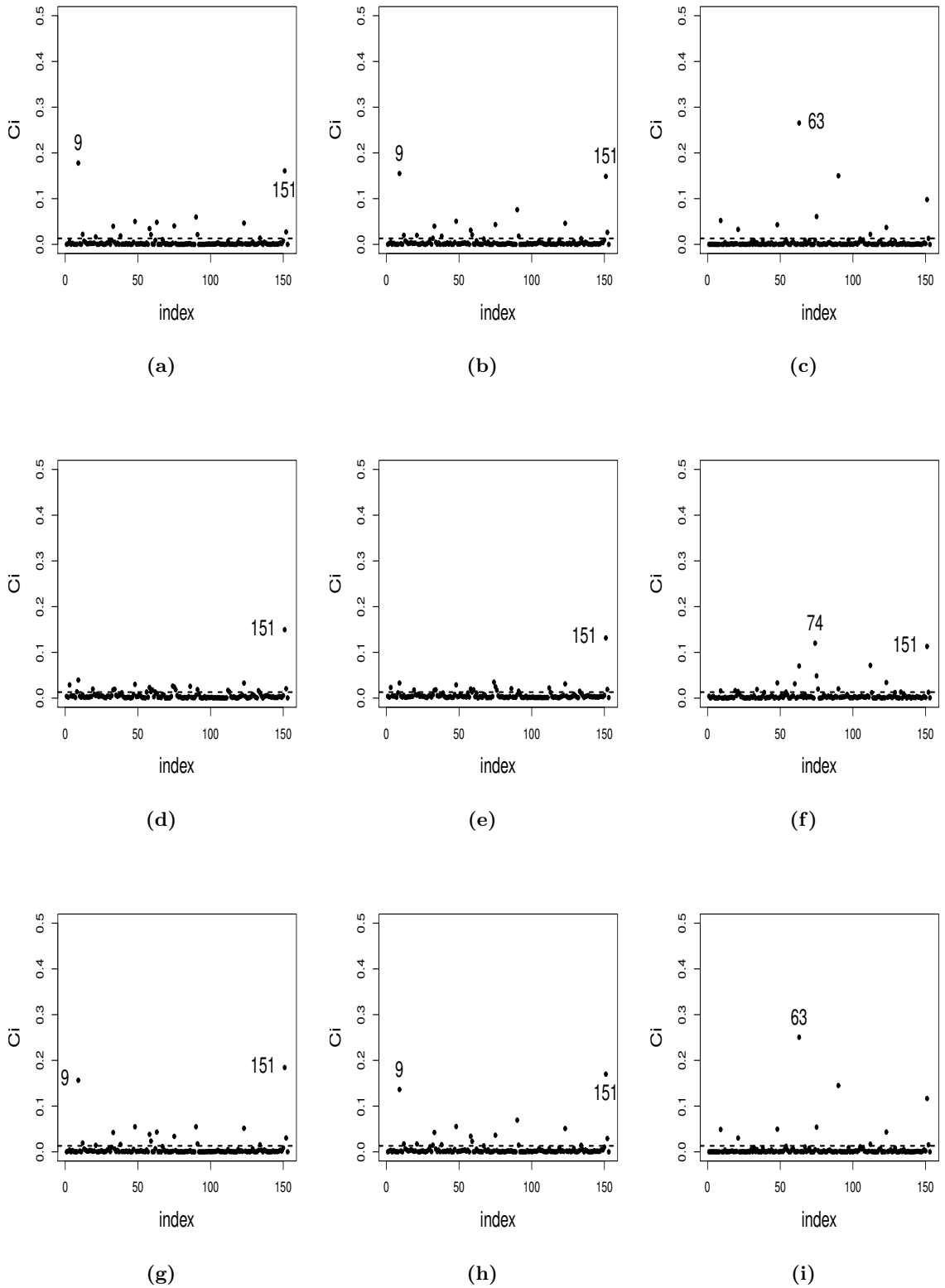


Figure 6: Index plots of C_i for $\boldsymbol{\beta}$ (a,d,g), δ (b,e,h) and $\boldsymbol{f}$ (c,f,i) under case-weight (a,b,c); response (d,e,f); and precision (g,h,i) perturbations with data from Santiago, Chile.

Table 2: RCs (in %) in MPL estimates and in the corresponding estimated SEs for the indicated removed case(s), and respective p-values using environmental data and the RBS-SAM with data from Santiago, Chile.

Removed case(s)	RCs in the estimate of	β_1	β_2	δ
None	Parameter	—	—	—
	SE	—	—	—
	(p-value)	< 0.01	< 0.01	—
{9}	Parameter	8.52	0.14	1.18
	SE	0.15	1.55	7.85
	(p-value)	< 0.01	< 0.01	—
{63}	Parameter	0.47	0.37	0.94
	SE	0.26	1.57	7.94
	(p-value)	< 0.01	< 0.01	—
{151}	Parameter	1.28	0.27	1.52
	SE	1.09	2.17	9.23
	(p-value)	< 0.01	< 0.01	—
{9, 63}	Parameter	7.29	0.41	2.14
	SE	0.21	1.67	7.75
	(p-value)	< 0.01	< 0.01	—
{9, 151}	Parameter	2.84	0.33	1.12
	SE	1.33	2.40	9.37
	(p-value)	< 0.01	< 0.01	—
{63, 151}	Parameter	0.92	0.84	2.49
	SE	1.03	2.29	9.13
	(p-value)	< 0.01	< 0.01	—
{9, 63, 151}	Parameter	2.13	0.90	2.10
	SE	1.25	2.52	9.27
	(p-value)	< 0.01	< 0.01	—

6. CONCLUDING REMARKS

In this study, we proposed and derived a reparametrized Birnbaum–Saunders semi-parametric additive model. This model allowed us to describe the mean of a random variable whose dispersion is non-constant through covariates. These covariates can contribute parametrically or nonparametrically in the model. Furthermore, it was possible to maintain the original scale of the data, since performing transformations on the modeled variable can reduce interpretability. We used a back-fitting algorithm to obtain the maximum penalized likelihood estimates by using the cubic smoothing splines.

To find potentially influential observations, we derived local influence techniques for the proposed model under case-weight, response variable and precision parameter perturbations. We apply the proposed model to real pollution data, verifying that the reparametrized Birnbaum–Saunders semi-parametric additive model is adequate to work with this type of data.

As future research, a package in R will be implemented so that various users have this model computationally to their disposal to be applied in practice. Currently, the R codes are available from the authors upon request. Also, the reparametrized Birnbaum–Saunders semi-parametric model can be extended to the reparametrized Birnbaum–Saunders semi-parametric case with varying precision, thin-plate spline, and partially varying-coefficient models. In addition, the use of functional, latent, spatial, and temporal structures, as well as reduction of dimensionality employing partial least squares regression, and mixture models that can facilitate the parameters estimation, are aspects to be considered in further studies with optimism because they have been explored in the Birnbaum–Saunders parametric case [12, 19, 25, 26, 35].

ACKNOWLEDGMENTS

The authors would like to thank the Editors and Reviewers very much for their constructive comments on an earlier version of this article which resulted in this improved version. The research of Carolina Marchant was partially funded by the National Agency for Research and Development (ANID) of the Chilean government under the Ministry of Science, Technology, Knowledge and Innovation, grants FONDECYT 11190636 and ANID-Millennium Science Initiative Program NCN17_059. The research of Germán Ibacache-Pulgar was partially funded by ANID, grant FONDECYT 11130704. The research of Víctor Leiva was partially funded by ANID, grant FONDECYT 1200525.

REFERENCES

[1] Arrue, J.; Arellano, R.; Gomez, H.W. and Leiva, V. (2020). On a new type of Birnbaum–Saunders models and its inference and application to fatigue data, *Journal of Applied Statistics*, **47**, 2690–2710.

[2] Balakrishnan, N.; Gupta, R.; Kundu, D.; Leiva, V. and Sanhueza, A. (2011). On some mixture models based on the Birnbaum–Saunders distribution and associated inference, *Journal of Statistical Planning and Inference*, **141**, 2175–2190.

[3] Balakrishnan, N. and Kundu, D. (2019). Birnbaum–Saunders distribution: a review of models, analysis, and applications, *Applied Stochastic Models in Business and Industry*, **35**, 4–49.

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

[5] Bourguignon, M.; Leao, J.; Leiva, V. and Santos-Neto, M. (2017). The transmuted Birnbaum–Saunders distribution, *REVSTAT – Statistical Journal*, **15**, 601–628.

[6] Buja, A.; Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models, *Annals of Statistics*, **17**, 453–555.

[7] Carrasco, J.M.F.; Figueroa-Zúñiga, J.; Leiva, V.; Riquelme, M. and Aykroyd, R.G. (2020). An errors-in-variables model based on the Birnbaum–Saunders and its diagnostics with an application to earthquake data. *Stochastic Environmental Research and Risk Assessment*, **34**, 369–380.

[8] Cavieres, M.F.; Leiva, V.; Marchant, C. and Rojas, F. (2021). A methodology for data-driven decision making in the monitoring of particulate matter environmental contamination in Santiago of Chile, *Reviews of Environmental Contamination and Toxicology*, **250**, 45–67.

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

[10] Dasilva, A.; Dias, R.; Leiva, V.; Marchant, C. and Saulo, H. (2020). Birnbaum–Saunders regression models: a comparative evaluation of three approaches, *Journal of Statistical Computation and Simulation*, **90**, 2552–2570.

[11] Escobar, L. and Meeker, W. (1992). Assessing influence in regression analysis with censored data, *Biometrics*, **48**, 507–28.

[12] 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*, **17**, 114–128.

[13] Gomez-Deniz, E.; Leiva, V.; Calderin-Ojeda, E. and Chesneau, C. (2022). A novel claim size distribution based on a Birnbaum–Saunders and gamma mixture capturing extreme values in insurance: estimation, regression, and applications, *Computational and Applied Mathematics*, **41**, 171.

[14] Green, P. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty*, Chapman and Hall, New York, USA.

[15] Hastie, T. and Tibshirani, R. (1986). Generalized additive models, *Statistical Science*, **1**, 293–310.

[16] Hastie, T. and Tibshirani, R. (1987). Generalized additive models: some applications, *Journal of the American Statistical Association*, **82**, 371–386.

[17] Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall, New York, USA.

[18] Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models, *Journal of the Royal Statistical Society B*, **55**, 757–796.

[19] Huerta, M.; Leiva, V.; Liu, S.; Rodriguez, M. and Villegas, D. (2019). On a partial least squares regression model for asymmetric data with a chemical application in mining, *Chemometrics and Intelligent Laboratory Systems*, **190**, 55–68.

[20] Ibacache-Pulgar, G.; Paula, G.A. and Cysneiros, F.J.A. (2013). Semi-parametric additive models under symmetric distributions, *TEST*, **22**, 103–121.

[21] Ibacache-Pulgar, G.; Paula, G.A. and Galea, M. (2012). Influence diagnostics for elliptical semi-parametric mixed models, *Statistical Modeling*, **12**, 165–193.

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

[23] Ibacache-pulgar, G.; Figueroa-Zúñiga, J. and Marchant, C. (2021). Semi-parametric additive beta regression models: inference and local influence diagnostics, *REVSTAT – Statistical Journal*, **19**, 255–274.

[24] Kim, C.; Park, B.U. and Kim, W. (2002). Influence diagnostics in semi-parametric regression models, *Statistics and Probability Letters*, **60**, 49–58.

[25] Kotz, S.; Leiva, V. and Sanhueza, A. (2010). Two new mixture models related to the inverse Gaussian distribution, *Methodology and Computing in Applied Probability*, **12**, 199–212.

[26] 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.

[27] Leiva, V.; Santos-Neto, M.; Cysneiros, F.J.A. and Barros, M. (2014). Birnbaum–Saunders statistical modeling: a new approach, *Statistical Modeling*, **14**, 21–48.

[28] Leiva, V.; Marchant, C.; Ruggeri, F. and Saulo, H. (2015). A criterion for environmental assessment using Birnbaum–Saunders attribute control, *Environmetrics*, **26**, 463–476.

[29] Leiva, V.; Ferreira, M.; Gomes, M.I. and Lillo, C. (2016). Extreme value Birnbaum–Saunders regression models applied to environmental data, *Stochastic Environmental Research and Risk Assessment*, **30**, 1045–1058.

[30] Leiva, V.; Castro, C.; Vila, R. and Saulo, H. (2024). Unveiling patterns and trends in research on cumulative damage models for statistical and reliability analyses: Bibliometric and thematic explorations with data analytics, *Chilean Journal of Statistics*, **15**, 81–109.

[31] Marchant, C.; Leiva, V.; Cavieres, M. and Sanhueza, A. (2013). Air contaminant statistical distributions with application to PM10 in Santiago, Chile, *Reviews of Environmental Contamination and Toxicology*, **223**, 1–31.

[32] Marchant, C.; Leiva, V. and Cysneiros, F.J.A. (2016). A multivariate log-linear model for Birnbaum–Saunders distributions, *IEEE Transactions on Reliability*, **65**, 816–827.

[33] Marchant, C.; Leiva, V.; Cysneiros, F.J.A. and Liu, S. (2018). Robust multivariate control charts based on Birnbaum–Saunders distributions, *Journal of Statistical Computation and Simulation*, **88**, 182–202.

[34] Marchant, C.; Leiva, V.; Christakos, G. and Cavieres, M.A. (2019). Monitoring urban environmental pollution by bivariate control charts: new methodology and case study in Santiago, Chile, *Environmetrics*, **30**, e2551.

[35] Martinez, S.; Giraldo, R. and Leiva, V. (2019). Birnbaum–Saunders functional regression models for spatial data, *Stochastic Environmental Research and Risk Assessment*, **33**, 1765–1780.

[36] Mazucheli, J.; Menezes, A.F.B. and Dey, S. (2018). The unit-Birnbaum–Saunders distribution with applications, *Chilean Journal of Statistics*, **9**(1), 47–57.

[37] Mazucheli, M.; Leiva, V.; Alves, B. and Menezes, A.F.B. (2021). A new quantile regression for modeling bounded data under a unit Birnbaum–Saunders distribution with applications in medicine and politics, *Symmetry*, **13**, 682.

[38] Mazucheli, M.; Alves, B.; Menezes, A.F.B. and Leiva, V. (2022). An overview on parametric quantile regression models and their computational implementation with applications to biomedical problems including COVID-19 data, *Computer Methods and Programs in Biomedicine*, **221**, 106816.

[39] Morales, R.G.E.; Llanos, A.; Merino, M. and Gonzalez-Rojas, C.H. (2012). A semi-empirical method of PM-10 atmospheric pollution forecast at Santiago de Chile city, *Nature, Environment and Pollution Technology*, **11**, 181–186.

[40] Puentes, R.; Marchant, C.; Leiva, V.; Figueroa-Zúñiga, J. and Ruggeri, F. (2021). Predicting PM2.5 and PM10 levels during critical episodes management in Santiago, Chile, with a bivariate Birnbaum–Saunders log-linear model, *Mathematics*, **9**, 645.

[41] R Development Core Team (2021). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[42] Rieck, J.R. and Nedelman, J.R. (1991). A log–linear model for the Birnbaum–Saunders distribution, *Technometrics*, **33**, 51–60.

[43] Reyes, J.; Arrue, J.; Leiva, V. and Martin-Barreiro, C. (2021). A new Birnbaum–Saunders distribution and its mathematical features applied to bimodal real-world data from environment and medicine, *Mathematics*, **9**, 1891.

[44] Rigby, R.A. and Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape, *Applied Statistics*, **54**, 507–554.

[45] Sanchez, L.; Leiva, V.; Galea, M. and Saulo, H. (2021). Birnbaum–Saunders quantile regression and its diagnostics with application to economic data, *Applied Stochastic Models in Business and Industry*, **3**, 53–73.

[46] Santos–Neto, M.; Cysneiros, F.J.A.; Leiva, V. and Ahmed, S.E. (2012). On new parametrizations of the Birnbaum–Saunders distribution, *Pakistan Journal of Statistics*, **1**, 1–26.

[47] Santos-Neto, M.; Cysneiros, F.J.A.; Leiva, V. and Barros, M. (2014). On a reparameterized Birnbaum–Saunders distribution and its moments, estimation and applications, *REVSTAT – Statistical Journal*, **12**, 247–272.

[48] Saulo, H.; Souza, R.; Vila, R.; Leiva, V. and Aykroyd, R.G. (2021). Modeling mortality based on pollution and temperature using a new Birnbaum–Saunders autoregressive moving average structure with regressors and related-sensors data, *Sensors*, **21**, 6518.

[49] Saulo, H.; Leao, J.; Leiva, V. and Aykroyd, R.G. (2019). Birnbaum–Saunders autoregressive conditional duration models applied to high-frequency financial data, *Statistical Papers*, **60**, 1605–1629.

[50] Segal, M.R. and Bacchetti, P. (1994). Variances for maximum penalized likelihood estimates obtained via the EM algorithm, *Journal of the Royal Statistical Society B*, **56**, 345–352.

[51] Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline, *Journal of the Royal Statistical Society B*, **45**, 133–150.