Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Generalized symmetrical partial linear model

Julio Cezar Souza Vasconcelos & Cristian Villegas

JOURNAL OF APPLIED STATISTICS
https://doi.org/10.1080/02664763.2020.1726301

Taylor & Francis
Taylor & Francis Group

APPLICATION NOTE

Check for updates

Generalized symmetrical partial linear model

Julio Cezar Souza Vasconcelos and Cristian Villegas

University of São Paulo, Piracicaba, Brazil

ABSTRACT
In this work, we propose a new model called generalized symmetrical partial linear model, based on the theory of generalized linear models and symmetrical distributions. In our model the response variable follows a symmetrical distribution such a normal, Student-t, power exponential, among others. Following the context of generalized linear models we consider replacing the traditional linear predictors by the more general predictors in whose case one covariate is related with the response variable in a non-parametric fashion, that we do not specified the parametric function. As an example, we could imagine a regression model in which the intercept term is believed to vary in time or geographical location. The backfitting algorithm is used for estimating the parameters of the proposed model. We perform a simulation study for assessing the behavior of the penalized maximum likelihood estimators. We use the quantile residuals for checking the assumption of the model. Finally, we analyzed real data set related with pH rivers in Ireland.

ARTICLE HISTORY
Received 5 May 2019
Accepted 2 February 2020

KEYWORDS
Backfitting algorithm; cubic splines; generalized linear models; partial linear model; symmetrical distributions

1. Introduction

The generalized linear models proposed by Nelder and Wedderburn [16] have been applied in several areas of knowledge, such as agronomy, forestry, biological sciences, toxicology, among others. Green and Yandell [11] proposed the semiparametric generalized linear models in which the authors added in the linear predictor a non-parametric component. In particular, they considered the logistic regression model from a biological assay of the flame extinguishing agent conducted in the United States in the National Toxicology Program [5]. Green and Yandell [11] fitted the logistic regression model, in which the first three covariates were considered parametrically and the fourth variable (age-at-death) non-parametrically in the linear predictor. The symmetric distributions studied, for instance, by Kelker [14], Cambanis et al. [2], Chmielewski [3] and Lange and Sinsheimer [15] have been used as an interesting alternative to the normal distribution when there are atypical observations, since they have heavier tails than normal distribution. An extension of the symmetric regression models [9] was done by Villegas et al. [19], in which they use a link function in the same sense of generalized linear models [16]. The goal of this work is to develop a model based on semiparametric generalized linear models [11] and symmetrical distributions considering in the linear predictor a parametric part plus a non-parametric

CONTACT Julio Cezar Souza Vasconcelos juliocezarvasconcelos@hotmail.com

function whose form is not specified. For the selection of the model, we used the generalized Akaike information criterion (*GAIC*). The estimates will be obtained by the backfitting algorithm. In addition, quantile residual with simulated envelope will be used to verify the goodness of fit. The data used are a subset of the data analyzed in Cruikshanks et al. [4], a technical report by the Environmental Protection Agency, Wexford (Ireland). The research sampled 257 rivers in Ireland. One of the aims was to find a different tool for identifying acid-sensitive waters, which currently uses measures of pH. It will be considered in this work the following variables: pH, days, forested (factor variable with two levels: non-forested and forested sites), field pH and temperature. There are some missing data. We will use pH as a response variable and the others as covariates. The paper is organized as follows. In Section 2 we present a brief review of symmetric distributions. Section 3 gives the generalized symmetrical partial linear model and the backfitting algorithm for the parameter estimation. Section 4 presents the quantile residuals for checking the assumptions of the model. In Section 5 we develop a small simulation study. In Section 6, the methodology is applied to self-reports of height and weight. Finally, conclusions are given in Section 7.

2. The symmetrical distributions

This section presents some distributions belonging to the univariate symmetrical distribution family. A random variable Y has a symmetrical distribution, with location parameter μ, scale parameter ϕ and density generating function $h(\cdot)$, if its probability density function can be expressed as

$$f(y;\mu,\phi) = \frac{1}{\sqrt{\phi}}h(u), \qquad (1)$$

where $u = (y-\mu)^2/\phi > 0$, $\mu \in \mathbb{R}, \phi > 0$ and $h(\cdot)$ is a positive definite function in $\mathbb{R}^+$, in such a way that $\int_0^\infty u^{-1/2}h(u)\,\mathrm{d}u = 1$ [14]. We will denote: $Y \sim S(\mu,\phi,h)$. The characteristic function takes the form $\exp(it\mu)\psi(t^2\phi)$ for some function $\psi(\cdot)$, where $i = \sqrt{-1}$. Kelker [14] showed that $u^{(1/2)(1-k)-1}h(u)$ is integrable, so there are moments of order k of the random variable Y. Moreover, if the expectation and variance of Y exist, then they are given by $E(Y) = \mu$ and $\mathrm{Var}(Y) = \xi_h\phi$, respectively, where $\xi_h = -2\psi'(0)$ is a positive constant. In most symmetric distributions, the $h(\cdot)$ function depends on an additional parameter, called the shape parameter, which will be indicated as ν, controlling the kurtosis of the distribution. This parameter can be fixed or estimated from the data. Based on the distributions with these characteristics, attention will be focused on the following distributions: normal, power exponential and Student-t. Next, the probability density functions of some distributions will be defined.

(1) normal: $h(u) = (2\pi)^{-0.5}\exp\{-0.5u\}$.
(2) power exponential: $h(u) = (\nu\exp\{-0.5|u/c|^\nu\})/(c2^{1+1/\nu}\Gamma(1/\nu))$, $\nu > 0$ where $c = (2^{-2/\nu}\Gamma(1/\nu)/\Gamma(3/\nu))^{0.5}$
(3) Student-t: $h(u) = (\nu^{0.5\nu}/B(0.5, 0.5\nu))(\nu+u)^{-0.5(\nu+1)}$, $\nu > 0$.
(4) hyperbolic: $h(u) = (2K_1(\nu))^{-1}\exp\{-\nu\sqrt{1+u}\}$, $\nu > 0$ and $K_\lambda(\nu) = 0.5\int_0^\infty x^{\lambda-1}\exp\{-0.5\nu(x+x^{-1})\}\,\mathrm{d}x$ is the modified Bessel function of the third kind λ.

(5) slash: $h(u) = (\nu/\sqrt{2\pi})G(\nu + 0.5, 0.5u)$, $\nu > 0$ and $G(a, b) = \int_0^1 x^{a-1} \exp\{-bx\}\,dx$ is the incomplete gamma function.
(6) Cauchy: $h(u) = 1/(\pi(1 + u))$.
(7) generalized-t: $h(u) = (s^{0.5r}/B(0.5, 0.5r))(s + u)^{-0.5(r+1)}$, $s, r > 0$. It includes Student-t ($s = r = \nu$) and Cauchy distribution ($s = r = 1$).
(8) type I logistic: $h(u) = c(\exp\{-u\}/(1 + \exp\{-u\})^2)$, $c \approx 1.484300029$.
(9) type II logistic: $h(u) = \exp\{-\sqrt{u}\}/(1 + \exp\{-\sqrt{u}\})^2$.
(10) generalized logistic: $h(u) = (\alpha/B(m, m))[\exp\{-\alpha\sqrt{u}\}/(1 + \exp\{-\alpha\sqrt{u}\})]^m$, $m, \alpha > 0$. The type II logistic density corresponds to the case $m = 1, \alpha = 1$;
(11) contaminated normal

$$\begin{aligned} h(u) = (1 - \epsilon)\frac{1}{\sqrt{2\pi}} \exp\{-0.5u\} \\ + \epsilon\frac{1}{\sigma\sqrt{2\pi}} \exp\{-0.5u/(\sigma^2)\}, \quad \sigma > 0, \quad 0 \le \epsilon \le .1 \end{aligned}$$

For further details about symmetrical distributions see, for instance, Fang et al. [7] and Ferrari and Uribe-Opazo [8].

3. Generalized symmetrical partial linear model

The aim of this section is to present the new model called the generalized symmetrical partial linear model, that is an extension of a partial linear model in the same sense of generalized linear models, by incorporating a link function between the mean response that follows a symmetrical distribution.

3.1. The model

Let $Y_1, \ldots, Y_n$ be n independent random variables, each with probability density function expressed by

$$f(y_i; \mu_i, \phi) = \frac{1}{\sqrt{\phi}} h(u_i), \tag{2}$$

where $u_i = (y_i - \mu_i)^2/\phi$, $\mu_i \in \mathbb{R}$ is the location parameter, $\phi > 0$ is the scale parameter and $h(\cdot) > 0$ is the density generating function. We will assume that the parameter μ_i satisfies $g(\mu_i) = \eta_i = \mathbf{x}_i^T\boldsymbol{\beta} + f(t_i)$, with $g(\cdot)$ being the link function, a monotone and at least twice differentiable function as in GLMs. The model defined above is called Generalized Symmetric Partial Linear Model, where η_i is the linear predictor, $f(t_i)$ is an arbitrary function that depends on the covariate t_i which is non-parametrically controlled and $g(\mu_i)$ is the link function. Furthermore, the function $f(t_i)$, could be considered as an intercept term believed to vary in time or with geographical location. On the other hand, since μ_i may assume any real value, possible link functions could be $g(\mu_i) = \mu_i = \eta_i$ and $g(\mu_i) = \mu_i^{-1} = \eta_i$. However, if for a particular symmetric data set μ_i is positive, other possible link functions could be $g(\mu_i) = \log(\mu_i) = \eta_i$ and $g(\mu_i) = \mu_i = \sqrt{\eta_i}$. It is important to observe that the definition of canonical link is not available in generalized symmetrical partial linear models.

3.2. Penalized likelihood function

The log-likelihood function of the model given in (2) yields

$$L(\boldsymbol{\theta}) = -\frac{n}{2}\log\phi + \sum_{i=1}^{n}\log h(u_i). \tag{3}$$

For estimating the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}^T, \phi)^T$ of model given in (3) we need to impose some restriction on the non-parametric function $f(t)$ by assuming that it is smooth. Therefore, we introduce a penalty on the second derivative in the logarithm of the likelihood function [17]. Thus, the logarithm of the penalized likelihood function for the generalized symmetrical partial linear model is given by:

$$L_p(\boldsymbol{\theta}, \lambda) = -\frac{n}{2}\log\phi + \sum_{i=1}^{n}\log h(u_i) - \frac{\lambda}{2}\int_a^b [f''(t)]^2\,\mathrm{d}t, \tag{4}$$

where $f(t)$ is a twice continuously differentiable smoothing function, $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}^T, \phi)^T$. The smoothing parameter $\lambda > 0$ controls the tradeoff between the bias and the variance of $f(t)$. The integral in equation (4) can be written of the form [10]

$$\int_a^b [f''(t)]^2\,\mathrm{d}t = \mathbf{f}^T\mathbf{K}\mathbf{f}, \quad \text{for } a \le t \le b, \tag{5}$$

in which the positive definite matrix $\mathbf{K}$ $(q \times q)$ is defined as $\mathbf{K} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T$, where $\mathbf{Q}$ is a tridiagonal matrix $q \times (q-2)$ and d_i the distance between two subsequent and different control points called nodes i and $i+1$, this is, $d_i = t_{i+1}^0 - t_i^0$, $i = 1, \ldots, q-1$. Let $\mathbf{Q}$ be the $q \times (q-2)$ matrix, in which its elements q_{ij} for $i = 1, \ldots, q$ and $j = 2, \ldots, q-1$ are given by:

$$q_{j-1,j} = d_{j-1}^{-1}, \quad q_{jj} = -d_{j-1}^{-1} - d_j^{-1}, \quad \text{and} \quad q_{j+1,j} = d_j^{-1}$$
$$q_{ij} = 0 \quad \text{for } |i-j| \ge 2.$$

The symmetric matrix $\mathbf{R}$ is $(q-2) \times (q-2)$, in which its elements r_{ij} with $i = 2, \ldots, q-1$ and $j = 2, \ldots, q-1$, are given by:

$$r_{ii} = \tfrac{1}{3}(d_{i-1} + d_i) \quad \text{for } i = 2, \ldots, q-1,$$
$$r_{i,i+1} = r_{i+1,i} = \tfrac{1}{6}d_i \quad \text{for } i = 2, \ldots, q-2 \quad \text{and}$$
$$r_{ij} = 0 \quad \text{for } |i-j| \ge 2.$$

It is important to note that cubic splines are being used in (4) and the solution of (5) exists if, $f(t)$ is a cubic spline in which the nodes are the distinct values of $t_1, \ldots, t_n$, ordered denoted by $t_1^0 < t_2^0 < \cdots < t_q^0$. Finally, based on (4) and (5) the logarithm of the penalized likelihood function can be expressed as:

$$L_p(\boldsymbol{\theta}, \lambda) = -\frac{n}{2}\log\phi + \sum_{i=1}^{n}\log h(u_i) - \frac{\lambda}{2}\mathbf{f}^T\mathbf{K}\mathbf{f}, \tag{6}$$

with $\mathbf{f} = (f(t_1^0), f(t_2^0), \ldots, f(t_q^0))^T$ a $q \times 1$ vector.

3.3. Penalized score function and penalized hessian matrix

The penalized score function is given by:

$$\mathbf{U}^p(\boldsymbol{\theta}) = \frac{L_p(\boldsymbol{\theta},\lambda)}{\partial\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{U}^p_\beta(\boldsymbol{\theta}) \\ \mathbf{U}^p_f(\boldsymbol{\theta}) \\ U^p_\phi(\boldsymbol{\theta}) \end{pmatrix}.$$

Matricially $\mathbf{U}^p(\boldsymbol{\theta})$ could be writen as, see Appendix 1 for details:

$$\begin{aligned}
\mathbf{U}^p_\beta(\boldsymbol{\theta}) &= \frac{1}{\phi}\mathbf{X}^T\mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\xi})(\mathbf{y}-\boldsymbol{\mu}),\\
\mathbf{U}^p_f(\boldsymbol{\theta}) &= \frac{1}{\phi}\mathbf{N}^T\mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\xi})(\mathbf{y}-\boldsymbol{\mu}) - \lambda\mathbf{K}\mathbf{f}, \quad \text{and}\\
U^p_\phi(\boldsymbol{\theta}) &= -\frac{n}{2\phi} + \frac{1}{2\phi^2}(\mathbf{y}-\boldsymbol{\mu})^T\mathbf{D}(\mathbf{v})(\mathbf{y}-\boldsymbol{\mu}),
\end{aligned}$$

where $\mathbf{D}(\mathbf{v}) = \text{diag}\{v_1,\ldots,v_n\}$, $\mathbf{D}(\boldsymbol{\xi}) = \text{diag}\{\dot{a}_1,\ldots,\dot{a}_n\}$, $\mathbf{y} = (y_1,\ldots,y_n)^T$, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n)^T$, $\mathbf{N}$ is an $(n\times q)$ incidence matrix (see [13]), $v_i = -2W_h(u_i)$, $W_h(u_i) = \mathrm{d}\log(h(u_i))/\mathrm{d}u_i = h'(u_i)/h(u_i)$ and $\dot{a}_i = \mathrm{d}\mu_i/\mathrm{d}\eta_i$. The elements of the penalized hessian matrix are given by:

$$\ddot{\mathbf{L}}^p(\boldsymbol{\theta}) = \frac{\partial^2 L_p(\boldsymbol{\theta},\lambda)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = \begin{pmatrix} \ddot{\mathbf{L}}^p_{\beta\beta}(\boldsymbol{\theta}) & \ddot{\mathbf{L}}^p_{\beta f}(\boldsymbol{\theta}) & \ddot{\mathbf{L}}^p_{\beta\phi}(\boldsymbol{\theta}) \\ \ddot{\mathbf{L}}^p_{f\beta}(\boldsymbol{\theta}) & \ddot{\mathbf{L}}^p_{ff}(\boldsymbol{\theta}) & \ddot{\mathbf{L}}^p_{f\phi}(\boldsymbol{\theta}) \\ \ddot{\mathbf{L}}^p_{\phi\beta}(\boldsymbol{\theta}) & \ddot{\mathbf{L}}^p_{\phi f}(\boldsymbol{\theta}) & \ddot{L}^p_{\phi\phi}(\boldsymbol{\theta}) \end{pmatrix}.$$

Matricially, the elements of $\ddot{\mathbf{L}}^p(\boldsymbol{\theta})$, could be written as, see Appendix 2 for details:

$$\begin{aligned}
\ddot{\mathbf{L}}^p_{\beta\beta}(\boldsymbol{\theta}) &= -\frac{1}{\phi}\mathbf{X}^T[\mathbf{D}(\mathbf{a})\mathbf{D}(\boldsymbol{\xi})^2 - \mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\rho})\mathbf{D}(\mathbf{y}-\boldsymbol{\mu})]\mathbf{X},\\
\ddot{\mathbf{L}}^p_{ff}(\boldsymbol{\theta}) &= -\frac{1}{\phi}\mathbf{N}^T[\mathbf{D}(\mathbf{a})\mathbf{D}(\boldsymbol{\xi})^2 - \mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\rho})\mathbf{D}(\mathbf{y}-\boldsymbol{\mu})]\mathbf{N} - \lambda\mathbf{K},\\
\ddot{L}^p_{\phi\phi}(\boldsymbol{\theta}) &= \frac{n}{2\phi^2} + \frac{1}{\phi^2}\left[\mathbf{u}^T\mathbf{D}(\mathbf{c})\mathbf{u} - \frac{1}{\phi}(\mathbf{y}-\boldsymbol{\mu})^T\mathbf{D}(\mathbf{v})(\mathbf{y}-\boldsymbol{\mu})\right],\\
\ddot{\mathbf{L}}^p_{\beta\phi}(\boldsymbol{\theta}) &= \frac{2}{\phi^2}\mathbf{X}^T\mathbf{b}(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\xi},\\
\ddot{\mathbf{L}}^p_{f\phi}(\boldsymbol{\theta}) &= \frac{2}{\phi^2}\mathbf{N}^T\mathbf{b}(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\xi},\\
\ddot{\mathbf{L}}^p_{\beta f}(\boldsymbol{\theta}) &= -\frac{1}{\phi}\mathbf{X}^T[\mathbf{D}(\mathbf{a})\mathbf{D}(\boldsymbol{\xi})^2 - \mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\rho})\mathbf{D}(\mathbf{y}-\boldsymbol{\mu})]\mathbf{N},
\end{aligned}$$

where $a_i = [v_i - 4W'_h(u_i)u_i]$, $u_i = (y_i-\mu_i)^2/\phi$, $c_i = W'_h(u_i) = \mathrm{d}W_h(u_i)/\mathrm{d}u_i$, $\ddot{a}_i = \mathrm{d}^2\mu_i/\mathrm{d}\eta_i^2$ and $b_i = [W_h(u_i) + W'_h(u_i)u_i]$, $\mathbf{b} = (b_1,\ldots,b_n)^T$, $\boldsymbol{\xi} = (\dot{a}_1,\ldots,\dot{a}_n)^T$, $\mathbf{D}(\mathbf{a}) = \text{diag}\{a_1,\ldots,a_n\}$, $\mathbf{D}(\mathbf{c}) = \text{diag}\{c_1,\ldots,c_n\}$, $\mathbf{D}(\boldsymbol{\rho}) = \text{diag}\{\ddot{a}_1,\ldots,\ddot{a}_n\}$ and $\mathbf{D}(\mathbf{y}-\boldsymbol{\mu}) = \text{diag}\{y_1-\mu_1,\ldots,y_n-\mu_n\}$. For $g(\mu_i) = \mu_i = \eta_i$, $\boldsymbol{\xi} = (1,\ldots,1)^T$, $\mathbf{D}(\boldsymbol{\xi}) = \mathbf{I}_n$, $\boldsymbol{\rho} = (0,\ldots,0)^T$ and therefore $\mathbf{D}(\boldsymbol{\rho}) = \mathbf{0}_n$. We could use several methods for estimating the parameters of (3), see for instance, Hastie and Tibshirani [12], Rigby and Stasinopoulos [18] and Wood [20]. We use in particular the backfitting algorithm following the recommendation given in Rigby and Stasinopoulos [18].

3.4. Backfitting algorithm

The aim of this section is to determine the estimates of the parameters that maximize the logarithm of the penalized likelihood function. The parameter $\boldsymbol{\theta}$ obtained through the penalized likelihood must satisfy the inequality

$$\mathbf{L}_p(\widehat{\boldsymbol{\theta}}, \lambda) \geq \sup_{\theta \in \Theta} \mathbf{L}_p(\boldsymbol{\theta}, \lambda).$$

Considering the penalized score vector $\mathbf{U}^p(\boldsymbol{\theta})$ the backfitting algorithm following Ibacache-Pulgar and Paula [13] for generalized symmetrical partial linear model is defined as follows. Solving the score equations $\mathbf{U}^p_{\beta}(\boldsymbol{\theta}) = \mathbf{0}$, $\mathbf{U}^p_f(\boldsymbol{\theta}) = \mathbf{0}$ and $U^p_{\phi}(\boldsymbol{\theta}) = 0$. Considering $g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f}$, we obtain the following equations:

$$\mathbf{X}^T\mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\xi})\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \mathbf{N}\mathbf{f}),$$
$$(\mathbf{N}^T\mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\xi})\mathbf{N} + \lambda\phi\mathbf{K})\mathbf{f} = \mathbf{N}^T\mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$
$$(\mathbf{y} - \boldsymbol{\mu})^T\mathbf{D}(\mathbf{v})(\mathbf{y} - \boldsymbol{\mu}) = n\phi.$$

Therefore, the backfitting algorithm is given by:

$$\boldsymbol{\beta}^{(r+1)} = (\mathbf{X}^T\mathbf{D}(\mathbf{v}^{(r)})\mathbf{D}(\boldsymbol{\xi}^{(r)})\mathbf{X})^{-1}\mathbf{X}^T\mathbf{D}(\mathbf{v}^{(r)})\mathbf{D}(\boldsymbol{\xi}^{(r)})(\mathbf{y} - \mathbf{N}\mathbf{f}^{(r)}), \quad (7)$$
$$\mathbf{f}^{(r+1)} = (\mathbf{N}^T\mathbf{D}(\mathbf{v}^{(r)})\mathbf{D}(\boldsymbol{\xi}^{(r)})\mathbf{N} + \lambda\phi^{(r)}\mathbf{K})^{-1}\mathbf{N}^T\mathbf{D}(\mathbf{v}^{(r)})\mathbf{D}(\boldsymbol{\xi}^{(r)})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(r+1)}), \quad (8)$$
$$\phi^{(r+1)} = \frac{1}{n}(\mathbf{y} - \boldsymbol{\mu}^{(r)})^T\mathbf{D}(\mathbf{v}^{(r)})(\mathbf{y} - \boldsymbol{\mu}^{(r)}), \quad \text{for } r = 1, 2, 3, \ldots \quad (9)$$

For starting the iterative process (7)–(9) we need some initial values $\boldsymbol{\beta}^{(0)}$, $\mathbf{f}^{(0)}$ and $\phi^{(0)}$. We could consider as starting values the estimates from the normal model with identity link function, that is, $Y_i \sim N(\mu_i, \phi)$ with $g(\mu_i) = \mu_i = \mathbf{x}_i^T\boldsymbol{\beta} + f(t_i)$, therefore

$$\boldsymbol{\beta}^{(0)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{y} - \mathbf{N}\mathbf{f}^{(0)}), \quad (10)$$
$$\mathbf{f}^{(0)} = (\mathbf{N}^T\mathbf{N} + \lambda\phi^{(0)}\mathbf{K})^{-1}\mathbf{N}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(0)}), \quad (11)$$
$$\phi^{(0)} = \frac{1}{n}(\mathbf{y} - \boldsymbol{\mu}^{(0)})^T(\mathbf{y} - \boldsymbol{\mu}^{(0)}), \quad (12)$$

with $\boldsymbol{\mu}^{(0)} = \mathbf{X}\boldsymbol{\beta}^{(0)} + \mathbf{N}\mathbf{f}^{(0)}$. On the other hand, for the model selection we will use the generalized Akaike information criterion (*GAIC*), based on Rigby and Stasinopoulos [18], given by

$$GAIC = -2\mathbf{L}_p(\widehat{\boldsymbol{\theta}}, \lambda) + 2(p + 1 + edf(\lambda)), \quad (13)$$

with $\mathbf{L}_p(\widehat{\boldsymbol{\theta}}, \lambda)$ denotes the penalized log-likelihood function assessed at $\widehat{\boldsymbol{\theta}}$ for a fixed λ, p is the number of parameters of $\boldsymbol{\beta}$ and $edf(\lambda)$ is the effective degress of freedom for a fixed λ. For further details about $edf(\lambda)$ see, for instance, [10]. The *GAIC* criterion consists in minimizing the function given in Equation (13).

4. Quantile residual

After fitting the generalized symmetrical partial linear model we need to study the goodness of fit for concluding if the model assumptions were satisfied, we do that with the quantile residual proposed by Dunn and Smyth [6]. The quantile residual for our model is defined as

$$r_{q,i} = \Phi^{-1}\{F(y_i; \hat{\mu}_i, \hat{\phi})\}, \tag{14}$$

where $F(y_i; \hat{\mu}_i, \hat{\phi})$ is the cumulative distribution function of $Y_i \sim S(\mu_i, \phi, h)$ with μ_i and ϕ consistently estimated and $\Phi(\cdot)^{-1}$ is the inverse cumulative standard normal distribution. For a better understanding of the normal quantile plot, Atkinson [1] add a simulated envelope to the plot. The envelope is made based on the minimum and maximum values of quantile residuals. For the fitted model to be plausible, at least 95% of the points must be contained within the envelope, otherwise the model did not fit the data well. We used the quantile residual because they produce residuals that are exactly normal by inverting the fitted distribution function for each response value and finding the equivalent standard normal quantile. More details about the quantile residual could be found in [6]. We chose to use the quantile residuals due to the computational implementation in *gamlss* package from software R, proposed by Rigby and Stasinopoulos [18].

5. Simulation study

For assessing the quality of penalized maximum likelihood estimation (PMLE) related to generalized symmetrical partial linear model considering the systematic part in a partially way, the estimates of the parameters will be calculated using the *gamlss* package from software R [18]. The Monte Carlo study via penalized maximum likelihood estimation (PMLE) it is important for checking bias of an estimator. In this case the simulation study was done for checking the precision of parameters associated with the explanatory variables x_1, x_2 and t_3. It will be done an analysis of the coefficient related to the variables x_1 and x_2, due to the coefficient related to the non-parametric part do not have a direct explanation, so it will be analyzed via scatterplots, in which each of this plot presents a true smooth curve given by $f(t_{i3}) = \sin^2(t_{i3}) + 5$ and the fitted curves considering cubic splines. Therefore, we expect for increasing sample size, fitted curves close to the real curve. We considered sample sizes of $n = 100$, 250 and 1000. It was simulated 2000 times considering 3 scenarios with the following distributions: normal (scenario 1), Student-t (scenario 2) and power exponential (scenario 3). Furthermore, for each scenario it was considered the identity and logarithm link function with the linear predictor η_i defined as $\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + f(t_{i3})$, with $x_{i1} \sim N(0, 3)$ and $x_{i2} \sim Ber(0.6)$ and $t_{i3} \sim U(0, 5)$. Other two scenarios were tested (omitted here), that is, $n = 30$ and $n = 500$. For a small sample size ($n = 30$), we obtained bad results, and for a sample size greater than 30 ($n > 30$) we obtained good results. We considered $\beta_1 = 0.2$, $\beta_2 = 0.05$ and $\phi = 1.3$. For Student-t and power exponential distributions we took $\nu = 3$. For each fitted model, we calculated the average estimator (AE), bias and mean squared error (MSE). We consider in the linear predictor the additive term as a cubic spline available in the package *gamlss* from software R [18], evaluated in t_{i3}, that is, $cs(t_{i3})$. In Table 1 we present the results of the simulation study. We notice that the penalized maximum likelihood estimation β_1 and β_2 goes to zero when the sample size n

Table 1. AEs, biases and MSEs for the normal, Student-t and power exponential model considering the identity (a) and logarithm (b) link function.

	$n = 100$			$n = 250$			$n = 1000$		
Parameters	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
Normal distribution									
(a) β_1	0.20001	0.00001	0.00203	0.20075	0.00075	0.00076	0.19998	−0.00002	0.00019
(a) β_2	0.04719	−0.00281	0.07407	0.05310	0.00310	0.02871	0.04824	−0.00176	0.00751
(b) β_1	0.19932	−0.00068	0.00006	0.19968	−0.00032	0.00003	0.19998	−0.00002	0.00001
(b) β_2	0.05139	0.00139	0.00133	0.04959	−0.00041	0.00056	0.04987	−0.00013	0.00017
Student-t distribution									
(a) β_1	0.20053	0.00053	0.00322	0.19974	−0.00026	0.00120	0.19973	−0.00027	0.00028
(a) β_2	0.06543	0.01543	0.12426	0.04811	−0.00189	0.04482	0.04916	−0.00084	0.01118
(b) β_1	0.19948	−0.00052	0.00009	0.19937	−0.00063	0.00005	0.19952	−0.00048	0.00002
(b) β_2	0.05072	0.00072	0.00206	0.04997	−0.00003	0.00085	0.05028	0.00028	0.00021
Power exponential distribution									
(a) β_1	0.20095	0.00095	0.00213	0.19974	−0.00026	0.00076	0.20068	0.00068	0.00018
(a) β_2	0.04368	−0.00632	0.07786	0.04913	−0.00087	0.02627	0.05189	0.00189	0.00652
(b) β_1	0.19958	−0.00042	0.00012	0.19974	−0.00026	0.00004	0.19995	−0.00005	0.00001
(b) β_2	0.05269	0.00269	0.00487	0.04956	−0.00044	0.00127	0.04958	−0.00042	0.00027

increased. Furthermore, the AEs of the parameters are close to the real values when the sample size increased.

Figures 1–3 show the fitted curves for the normal, Student-t and power exponential for the different sample sizes and link functions. We could see, as expected, that the variability among the estimates of the non-parametric function reduces as the sample size increases.

6. Application: pH rivers in Ireland in 2003

In this section, generalized symmetrical partial linear model is applied to the data analyzed in [4,21]. We modeled the pH data as a function of the explanatory variables days, forested

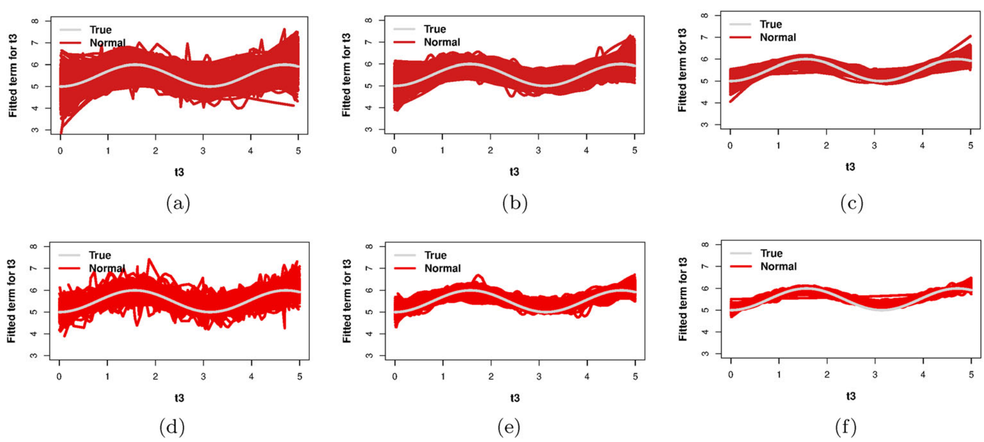


Figure 1. Normal model with the identity link function for (a) $n = 100$, (b) $n = 250$, (c) $n = 1000$ and logarithm link function for (d) $n = 100$, (e) $n = 250$, (f) $n = 1000$, with the true curve and fitted curves.

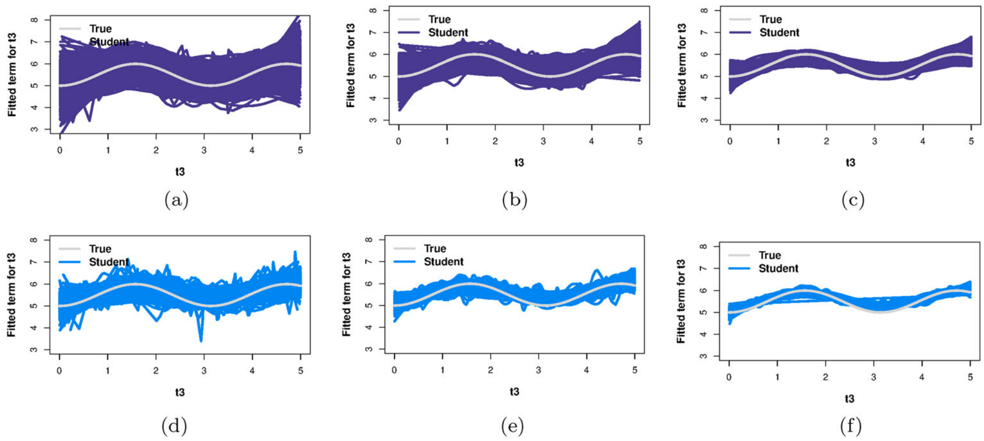


Figure 2. Student-t model with the identity link function for (a) $n = 100$, (b) $n = 250$, (c) $n = 1000$ and logarithm link function for (d) $n = 100$, (e) $n = 250$, (f) $n = 1000$, with the true curve and fitted curves.

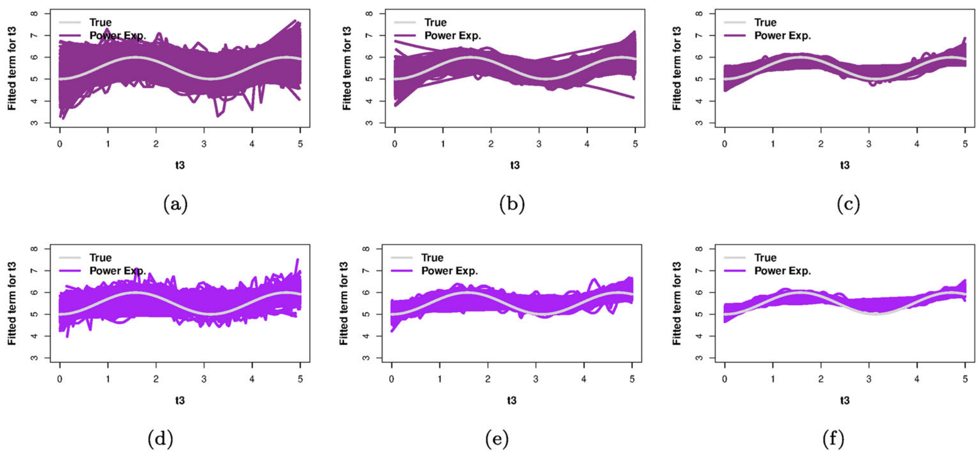


Figure 3. Power exponential model with the identity link function for (a) $n = 100$, (b) $n = 250$, (c) $n = 1000$ and logarithm link function for (d) $n = 100$, (e) $n = 250$, (f) $n = 1000$, with the true curve and fitted curves.

(as dummy variable 1 = non-forested sites and 2 = forested sites), field pH and the covariate temperature. Only 2003 data will be used, and several recordings were discarded as well as missing values were omitted, so the dataset has 204 complete observations for analyzing. Therefore, the results obtained in this work will be different to the results shown in the original report by Cruikshanks et al. [4]. In Figure 4 note that the histogram pH with the density curve using R software and ggplot2 package. In this figure, it is possible to observe a symmetrical behavior with atypical observations. Thus, we will use as an alternative to the normal distribution the exponential power and the Student-t distributions. Figure 5(a) shows the boxplot for pH by non-forested and forested sites. Figure 5(b,c) present the linear relationship between pH versus field pH and pH versus days, respectively. Finally, Figure 5(d) presents the nonlinear relationship between pH versus temperature.

6.1. Proposed models and model selection

We will propose to study the linear relationship between the explanatory variables (days, forested, field pH, temperature) and response variable (measured pH) the following model

$$\mathrm{pH}_i \sim S(\mu_i, \phi, h) \quad \text{with} \tag{15}$$

$$g(\mu_i) = \beta_0 + \beta_1 \,\mathrm{days}_i + \beta_2 \,\mathrm{forested}_i + \beta_3 \,\mathrm{fieldpH}_i + f(\mathrm{temperature}_i). \tag{16}$$

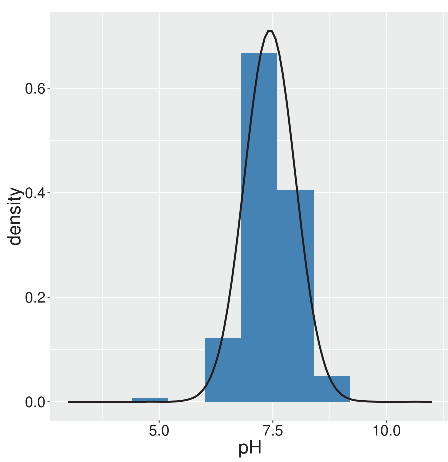
We will consider normal, power exponential and Student-t distributions for pH and as link functions: identity, square root, reciprocal and logarithmic. Therefore, the 12 models will be fitted. Based on Table 2, the model with the lowest *GAIC* measurement was the Student-t model with reciprocal link function. Table 3 shows the penalized likelihood maximum estimates, standard errors, t values and p values for the Student-t model considering the reciprocal link function. Note that only the forested covariate is not significant at the 5% significance level.

Figure 6(a) shows the estimated smooth term for the Student-t model considering reciprocal link function. Note a nonlinear effect temperature (as shown in Figure 5(d)), this explanatory variable can be considered in the linear predictor in a non-parametric manner. Figure 6(a) shows at about 6 degrees a slight decrease in pH, but from this value up to about 8 degrees the pH reduction is stronger.

To study possible divergence and check the quality of fit, we present the quantile probability plot for simulated envelope *quantile residuals*. From Figure 6(b), the Student-t model with the reciprocal link function can be considered a suitable model to explain this data set because do not present observations outside the envelope.

7. Concluding remarks

In this paper, we presented an extension of the generalized symmetric linear model by adding a non-parametric function to the linear predictor, called the generalized symmetric



(a)

Figure 4. Histogram of measured pH with the density curve.

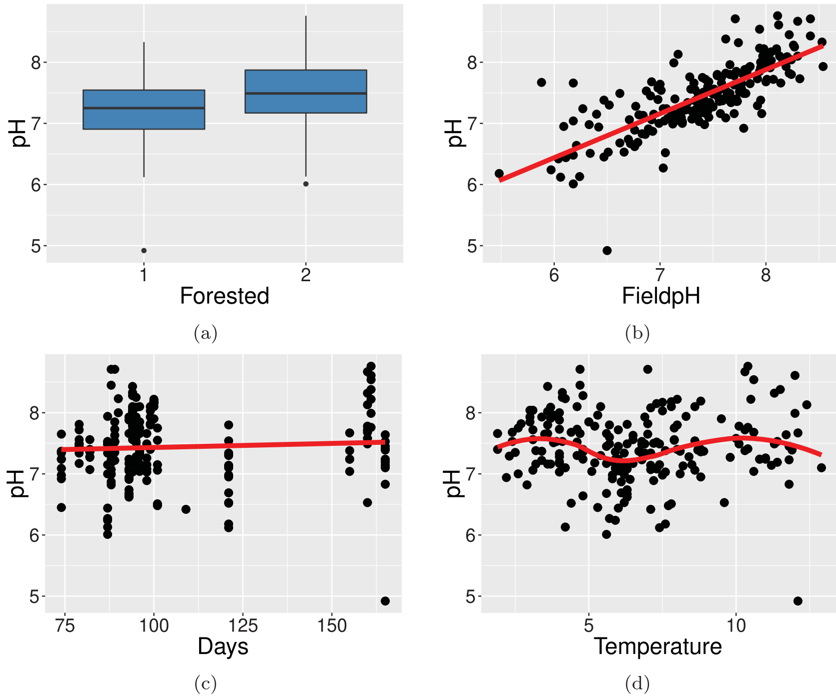


Figure 5. (a) Boxplot of the non-forested and forested sites by measured pH, (b) Scatterplot of measured pH versus Field pH, (c) Scatterplot of measured pH versus days and (d) Scatterplot of measured pH versus temperature for pH rivers in Ireland in the year of 2003.

Table 2. Results of the *GAIC* criterion for the normal, Student-t and power exponential models, with different link functions.

Distribution	Link function	*GAIC*
Student-t	**reciprocal**	**79.3725**
Student-t	log	81.8273
Student-t	sqrt	84.1003
Student-t	identity	87.0416
Power Exp.	reciprocal	82.3850
Power Exp.	log	83.5172
Power Exp.	sqrt	85.4131
Power Exp.	identity	88.0200
Normal	reciprocal	129.1632
Normal	log	131.9331
Normal	sqrt	133.7137
Normal	identity	135.7423

partial linear model. In this model the response variable follows a symmetrical distribution such a normal, Student-t, power exponential, among others. Following the context of generalized linear models we considered replacing the traditional linear predictors by the more general predictors in whose case one covariate is related with the response variable in a non-parametric fashion, that we do not specified the parametric function. For the non-parametric function, we considered a cubic spline and estimates the parameter by using the backfitting algorithm. Model selection was based on the generalized Akaike information

Table 3. Penalized maximum likelihood estimates (PMLEs), standard errors (SEs) and p-values for the Student-t model considering reciprocal link function.

Parameter	Estimate	SE	*t*-Value	*p*-Value
β_0	0.2597	0.0046	55.481	< 0.0001
β_1	-0.0001	0.00001	-6.0250	< 0.0001
β_2	-0.0005	0.0008	-0.6530	0.5150
β_3	-0.0151	0.0005	-25.4520	< 0.0001
ϕ	0.1991	0.0969	-16.6500	< 0.0001
ν	3.1174	0.2590	4.3890	< 0.0001
λ	0.0103			

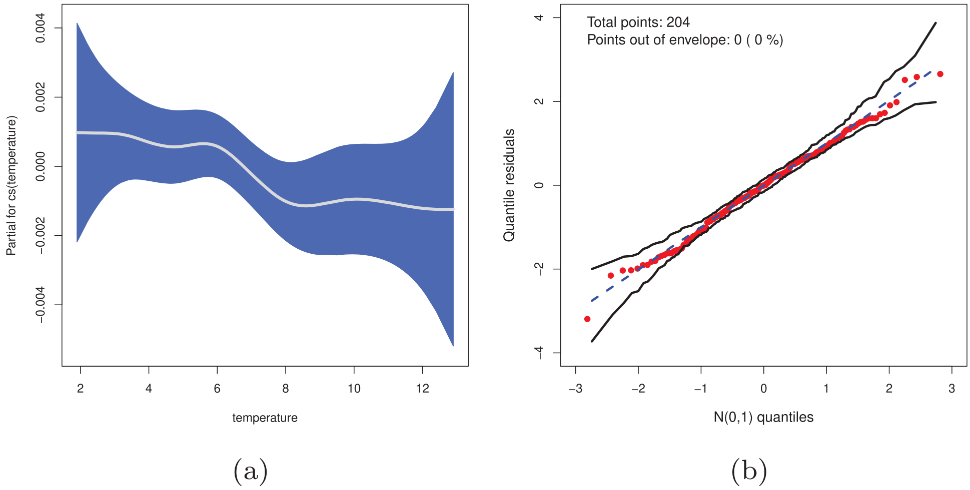


Figure 6. (a) 95% pointwise confidence bands for cs(temperature) from the Student-t model with reciprocal link function and (b) Normal probability plot with simulated envelope for the quantile residuals from the Student-t model with reciprocal link function.

criterion. The goodness of fit was performed by using the quantile residual. We developed a small simulation study for assessing the behavior of the penalized maximum likelihood estimators. A real data set was fitted for different response variables with different link functions. Finally, the code for the development of the application in the software R, is available from the authors upon request.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil.

References

[1] A.C. Atkinson, *Plots, Transformations and Regression; An Introduction to Graphical Methods of Diagnostic Regression Analysis*, No. 04, QA278. 2, A8, 1985.

[2] S. Cambanis, S. Huang, and G. Simons, *On the theory of elliptically contoured distributions*, J. Multivar. Anal. 11 (1981), pp. 368–385.
[3] M.A. Chmielewski, *Elliptically symmetric distributions: A review and bibliography*, Int. Stat. Rev. 49 (1981), pp. 67–74.
[4] R. Cruikshanks, R. Laursiden, A. Harrison, M.G.H. Hartl, M. Kelly-Quinn, P.S. Giller, and J. O'Halloran, *Evaluation of the use of the Sodium Dominance Index as a potential measure of acid sensitivity (2000-LS-3.2.1-M2)*, Synthesis Report, Environmental Protection Agency, Dublin, 2006, 26 pp.
[5] G. Dinse and S.W. Lagakos, *Regression analysis of tumour prevalence data*, J. R. Stat. Soc. C (Appl. Stat.) 32 (1983), pp. 236–248.
[6] P. Dunn and G.K. Smyth, *Randomized quantile residuals*, J. Comput. Graph. Stat. 5 (1996), pp. 236–244.
[7] K.T. Fang, S. Kotz, and K.W. Ng, *Symmetric Multivariate Related Distributions*, Chapman, London, 1990.
[8] S.L.P. Ferrari and M.A. Uribe-Opazo, *Corrected likelihood ratio tests in class of symmetric linear regression models*, Braz. J. Probab. Stat. 15 (2001), pp. 49–67.
[9] M. Galea, G.A. Paula, and M. Uribe-Opazo, *On influence diagnostic in univariate elliptical linear regression models*, Stat. Pap. 44 (2003), pp. 23–45.
[10] P. Green and B. Silverman, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, CRC Press, London, 1993.
[11] P. Green and B. Yandell, *Semi-Parametric Generalized Linear Models. Generalized Linear Models*, Springer, New York, NY, 1985, pp. 44–55.
[12] T.J. Hastie and R.J. Tibshirani, *Generalized Additive Models*, Chapman and Hall, London, 1990.
[13] G. Ibacache-Pulgar and G.A. Paula, *Local influence for Student-t partially linear models*, Comput. Stat. Data Anal. 55 (2011), pp. 1462–1478.
[14] D. Kelker, *Distribution theory of spherical distributions and a location-scale parameter generalization*, Sankhyā: Ind. J. Stat. A 32 (1970), pp. 419–430.
[15] K.L. Lange and J.S. Sinsheimer, *Normal/independent distributions and their applications in robust regression*, J. Comput. Graph. Stat. 2 (1993), pp. 175–198.
[16] J.A. Nelder and R.W.M. Wedderburn, *Generalized linear models*, J. R. Stat. Soc. A 135 (1972), pp. 370–384.
[17] F. O'Sullivan, *A statistical perspective on ill-posed inverse problems*, Stat. Sci. 1 (1986), pp. 502–518.
[18] R.A. Rigby and D.M. Stasinopoulos, *Generalized additive models for location, scale and shape*, J. R. Stat. Soc. C (Appl. Stat.) 54 (2005), pp. 507–554.
[19] C. Villegas, G.A. Paula, F.J.A. Cysneiros, and M. Galea, *Influence diagnostics in generalized symmetric linear models*, Comput. Stat. Data Anal. 59 (2013), pp. 161–170.
[20] S.N. Wood, *Generalized Additive Models. An Introduction with R*, Chapman and Hall, London, 2006.
[21] A.F. Zuur, E.N. Ieno, N.J. Walker, A.A. Saveliev, and G.M. Smith, *Mixed Effects Models and Extensions in Ecology with R*, Springer-Verlag, New York, 2009.

Appendices

Appendix 1. Penalized score function

In this section, we calculate the first derivative of the logarithm of the penalized likelihood function $L_p(\boldsymbol{\theta}, \lambda)$. The penalized score function, $\mathbf{U}^p(\boldsymbol{\theta})$, is equal to

$$\mathbf{U}^p(\boldsymbol{\theta}) = \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \mathbf{U}^p_\beta(\boldsymbol{\theta}) \\ \mathbf{U}^p_f(\boldsymbol{\theta}) \\ U^p_\phi(\boldsymbol{\theta}) \end{pmatrix}.$$

The first derivative of $L_p(\boldsymbol{\theta},\lambda)$, with respect to $\boldsymbol{\beta}$ can be computed using the product and chain rules:

$$\begin{aligned}
\mathbf{U}^p_\beta(\boldsymbol{\theta}) = \frac{\partial L_p(\boldsymbol{\theta},\lambda)}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n} \frac{\mathrm{d}\log(h(u_i))}{\mathrm{d}u_i}\frac{\partial u_i}{\partial \boldsymbol{\beta}} \\
&= \sum_{i=1}^{n} \frac{\mathrm{d}\log(h(u_i))}{\mathrm{d}u_i}\left[\frac{-2(y_i-\mu_i)}{\phi}\frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}\right] \\
&= \frac{1}{\phi}\sum_{i=1}^{n} v_i(y_i-\mu_i)\dot{a}_i\mathbf{x}_i,
\end{aligned}$$

with $v_i = -2W_h(u_i)$, $W_h(u_i) = \mathrm{d}\log(h(u_i))/\mathrm{d}u_i$ and $\dot{a}_i = \mathrm{d}\mu_i/\mathrm{d}\eta_i$. The first derivative of $L_p(\boldsymbol{\theta},\lambda)$, with respect to $\mathbf{f}$ can be computed using the product and chain rules:

$$\begin{aligned}
\mathbf{U}^p_f(\boldsymbol{\theta}) = \frac{\partial L_p(\boldsymbol{\theta},\lambda)}{\partial \mathbf{f}} &= \sum_{i=1}^{n} \frac{\mathrm{d}\log(h(u_i))}{\mathrm{d}u_i}\frac{\partial u_i}{\partial \mathbf{f}} - \frac{\lambda}{2}\left\{\frac{\partial \mathbf{f}^{\mathrm{T}}\mathbf{K}\mathbf{f}}{\partial \mathbf{f}}\right\} \\
&= \sum_{i=1}^{n} \frac{\mathrm{d}\log(h(u_i))}{\mathrm{d}u_i}\left[-\frac{2(y_i-\mu_i)}{\phi}\frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i}\frac{\partial \eta_i}{\partial \mathbf{f}}\right] - \lambda\mathbf{K}\mathbf{f} \\
&= \frac{1}{\phi}\sum_{i=1}^{n} v_i(y_i-\mu_i)\dot{a}_i\mathbf{n}_i - \lambda\mathbf{K}\mathbf{f}.
\end{aligned}$$

Finally, the first derivative of $L_p(\boldsymbol{\theta},\lambda)$, with respect to ϕ can be computed using the product and chain rules:

$$\begin{aligned}
U^p_\phi(\boldsymbol{\theta}) = \frac{\partial L_p(\boldsymbol{\theta},\lambda)}{\partial \phi} &= -\frac{n}{2}\frac{\partial \log\phi}{\partial \phi} + \sum_{i=1}^{n}\left\{\frac{\mathrm{d}\log(h(u_i))}{\mathrm{d}u_i}\frac{\partial u_i}{\partial \phi}\right\} \\
&= -\frac{n}{2\phi} + \sum_{i=1}^{n} \frac{\mathrm{d}\log(h(u_i))}{\mathrm{d}u_i}\left[-\frac{(y_i-\mu_i)^2}{\phi^2}\right] \\
&= -\frac{n}{2\phi} + \frac{1}{2\phi^2}\sum_{i=1}^{n} v_i(y_i-\mu_i)^2.
\end{aligned}$$

Appendix 2. Penalized hessian matrix

In this section, we calculate the second derivative of the logarithm of the penalized likelihood function $L_p(\boldsymbol{\theta},\lambda)$. The penalized hessian matrix, $\ddot{\mathbf{L}}^p(\boldsymbol{\theta})$, is equal to

$$\ddot{\mathbf{L}}^p(\boldsymbol{\theta}) = \frac{\partial^2 L_p(\boldsymbol{\theta},\lambda)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} = \begin{pmatrix} \ddot{\mathbf{L}}^p_{\beta\beta}(\boldsymbol{\theta}) & \ddot{\mathbf{L}}^p_{\beta f}(\boldsymbol{\theta}) & \ddot{\mathbf{L}}^p_{\beta\phi}(\boldsymbol{\theta}) \\ \ddot{\mathbf{L}}^p_{f\beta}(\boldsymbol{\theta}) & \ddot{\mathbf{L}}^p_{ff}(\boldsymbol{\theta}) & \ddot{\mathbf{L}}^p_{f\phi}(\boldsymbol{\theta}) \\ \ddot{\mathbf{L}}^p_{\phi\beta}(\boldsymbol{\theta}) & \ddot{\mathbf{L}}^p_{\phi f}(\boldsymbol{\theta}) & \ddot{L}^p_{\phi\phi}(\boldsymbol{\theta}) \end{pmatrix}.$$

The second derivative of $L_p(\boldsymbol{\theta}, \lambda)$, with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^T$ can be computed using the product and chain rules:

$$\begin{aligned}
\ddot{\mathbf{L}}^p_{\beta\beta}(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \frac{\partial}{\partial \boldsymbol{\beta}} \left[\frac{1}{\phi} \sum_{i=1}^n v_i (y_i - \mu_i) \dot{a}_i \mathbf{x}_i^T \right] \\
&= \frac{1}{\phi} \sum_{i=1}^n \left[\frac{\partial v_i}{\partial \boldsymbol{\beta}} (y_i - \mu_i) \dot{a}_i + v_i \frac{\partial (y_i - \mu_i) \dot{a}_i}{\partial \boldsymbol{\beta}} \right] \mathbf{x}_i^T \\
&= \frac{1}{\phi} \sum_{i=1}^n \left\{ -2 W_h'(u_i) \frac{\partial u_i}{\partial \boldsymbol{\beta}} (y_i - \mu_i) \dot{a}_i + v_i (y_i - \mu_i) \frac{\partial \dot{a}_i}{\partial \boldsymbol{\beta}} - v_i \dot{a}_i \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right\} \mathbf{x}_i^T \\
&= -\frac{1}{\phi} \sum_{i=1}^n [a_i \dot{a}_i^2 - v_i (y_i - \mu_i) \ddot{a}_i] \mathbf{x}_i \mathbf{x}_i^T,
\end{aligned}$$

where $a_i = [v_i - 4W_h'(u_i)u_i]$, $u_i = (y_i - \mu_i)^2/\phi$ and $\ddot{a}_i = \mathrm{d}^2\mu_i/\mathrm{d}\eta_i^2$. The second derivative of $L_p(\boldsymbol{\theta}, \lambda)$, with respect to $\mathbf{f}$ and $\mathbf{f}^T$ can be computed using the product and chain rules:

$$\begin{aligned}
\ddot{\mathbf{L}}^p_{ff}(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \mathbf{f} \partial \mathbf{f}^T} = \frac{\partial}{\partial \mathbf{f}} \left[\frac{1}{\phi} \sum_{i=1}^n v_i (y_i - \mu_i) \dot{a}_i \mathbf{n}_i^T - \lambda \mathbf{f}^T \mathbf{K} \right] \\
&= \frac{1}{\phi} \sum_{i=1}^n \left[\frac{\partial v_i}{\partial \mathbf{f}} (y_i - \mu_i) \dot{a}_i + v_i \frac{\partial (y_i - \mu_i) \dot{a}_i}{\partial \mathbf{f}} \right] \mathbf{n}_i^T - \lambda \frac{\partial \mathbf{f}^T \mathbf{K}}{\partial \mathbf{f}} \\
&= \frac{1}{\phi} \sum_{i=1}^n \left\{ -2 W_h'(u_i) \frac{\partial u_i}{\partial \mathbf{f}} (y_i - \mu_i) \dot{a}_i + v_i (y_i - \mu_i) \frac{\partial \dot{a}_i}{\partial \mathbf{f}} - v_i \dot{a}_i \frac{\partial \mu_i}{\partial \mathbf{f}} \right\} \mathbf{n}_i^T - \lambda \mathbf{K} \\
&= -\frac{1}{\phi} \sum_{i=1}^n [a_i \dot{a}_i^2 - v_i (y_i - \mu_i) \ddot{a}_i] \mathbf{n}_i \mathbf{n}_i^T - \lambda \mathbf{K}.
\end{aligned}$$

The second derivative of $L_p(\boldsymbol{\theta}, \lambda)$, with respect to ϕ and ϕ can be computed using the product and chain rules:

$$\begin{aligned}
\ddot{L}^p_{\phi\phi}(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \phi \partial \phi} = \frac{\partial}{\partial \phi} \left[-\frac{n}{2} \phi^{-1} + \frac{1}{2} \sum_{i=1}^n v_i (y_i - \mu_i)^2 \phi^{-2} \right] \\
&= \frac{n}{2\phi^2} + \frac{1}{2} \sum_{n=1}^n (y_i - \mu_i)^2 \left\{ \frac{\partial v_i}{\partial \phi} \phi^{-2} + v_i \frac{\partial \phi^{-2}}{\partial \phi} \right\} = \frac{n}{2\phi^2} + \frac{1}{\phi^2} \sum_{i=1}^n [c_i u_i^2 - v_i u_i],
\end{aligned}$$

where $c_i = W_h'(u_i) = \mathrm{d}W_h(u_i)/\mathrm{d}u_i$. The second derivative of $L_p(\boldsymbol{\theta}, \lambda)$, with respect to ϕ and $\boldsymbol{\beta}^T$ can be computed using the product and chain rules:

$$\begin{aligned}
\ddot{\mathbf{L}}^p_{\phi\beta}(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \phi \partial \boldsymbol{\beta}^T} = \frac{\partial}{\partial \phi} \left[\frac{1}{\phi} \sum_{i=1}^n v_i (y_i - \mu_i) \dot{a}_i \mathbf{x}_i^T \right] \\
&= \sum_{i=1}^n (y_i - \mu_i) \dot{a}_i \left[\frac{\partial v_i}{\partial \phi} \phi^{-1} + v_i \frac{\partial \phi^{-1}}{\partial \phi} \right] \mathbf{x}_i^T = \frac{2}{\phi^2} \sum_{i=1}^n b_i (y_i - \mu_i) \dot{a}_i \mathbf{x}_i^T,
\end{aligned}$$

where $b_i = [W_h(u_i) + W_h'(u_i)u_i]$. The second derivative of $L_p(\boldsymbol{\theta}, \lambda)$, with respect to ϕ and $\mathbf{f}^T$ can be computed using the product and chain rules:

$$\begin{aligned}\ddot{\mathbf{L}}^p_{\phi f}(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \phi \partial \mathbf{f}^T} = \frac{\partial}{\partial \phi}\left[\frac{1}{\phi}\sum_{i=1}^n v_i(y_i - \mu_i)\dot{a}_i \mathbf{n}_i^T - \lambda \mathbf{f}^T \mathbf{K}\right] \\ &= \sum_{i=1}^n (y_i - \mu_i)\dot{a}_i \left[\frac{\partial v_i}{\partial \phi}\phi^{-1} + v_i \frac{\partial \phi^{-1}}{\partial \phi}\right]\mathbf{n}_i^T = \frac{2}{\phi^2}\sum_{i=1}^n b_i(y_i - \mu_i)\dot{a}_i \mathbf{n}_i^T.\end{aligned}$$

The second derivative of $L_p(\boldsymbol{\theta}, \lambda)$, with respect to $\mathbf{f}$ and $\boldsymbol{\beta}^T$ can be computed using the product and chain rules:

$$\begin{aligned}\ddot{\mathbf{L}}^p_{f\beta}(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \mathbf{f} \partial \boldsymbol{\beta}^T} = \frac{\partial}{\partial \mathbf{f}}\left[\frac{1}{\phi}\sum_{i=1}^n v_i(y_i - \mu_i)\dot{a}_i \mathbf{x}_i^T\right] \\ &= \frac{1}{\phi}\sum_{i=1}^n \left[\frac{\partial v_i}{\partial \mathbf{f}}(y_i - \mu_i)\dot{a}_i + v_i \frac{\partial (y_i - \mu_i)\dot{a}_i}{\partial \mathbf{f}}\right]\mathbf{x}_i^T \\ &= \frac{1}{\phi}\sum_{i=1}^n \left\{-2W_h'(u_i)\frac{\partial u_i}{\partial \mathbf{f}}(y_i - \mu_i)\dot{a}_i + v_i(y_i - \mu_i)\frac{\partial \dot{a}_i}{\partial \mathbf{f}} - v_i \dot{a}_i \frac{\partial \mu_i}{\partial \mathbf{f}}\right\}\mathbf{x}_i^T \\ &= -\frac{1}{\phi}\sum_{i=1}^n [a_i \dot{a}_i^2 - v_i \ddot{a}_i(y_i - \mu_i)]\mathbf{n}_i {\mathbf{x}_i}^T,\end{aligned}$$