



# Influence measures in nonparametric regression model with symmetric random errors

Germán Ibacache-Pulgar<sup>1,5</sup>  · Cristian Villegas<sup>2</sup> ·  
Javier Linkolk López-Gonzales<sup>3</sup> · Magaly Moraga<sup>4</sup>

Accepted: 15 May 2022 / Published online: 25 June 2022  
© Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

In this paper we present several diagnostic measures for the class of nonparametric regression models with symmetric random errors, which includes all continuous and symmetric distributions. In particular, we derive some diagnostic measures of global influence such as residuals, leverage values, Cook's distance and the influence measure proposed by Peña (Technometrics 47(1):1–12, 2005) to measure the influence of an observation when it is influenced by the rest of the observations. A simulation study to evaluate the effectiveness of the diagnostic measures is presented. In addition, we develop the local influence measure to assess the sensitivity of the maximum penalized likelihood estimator of smooth function. Finally, an example with real data is given for illustration.

**Keywords** Nonparametric regression models · Local influence · Peña measured · Cook's distance · Cubic spline

---

✉ Germán Ibacache-Pulgar  
gibacachepulgar@gmail.com

Cristian Villegas  
clobos@usp.br

Javier Linkolk López-Gonzales  
javierlinkolk@gmail.com

Magaly Moraga  
magaly.moraga@uach.cl

- <sup>1</sup> Present Address: Faculty of Sciences, Institute of Statistics, University of Valparaíso, Valparaíso, Chile
- <sup>2</sup> Present Address: Department of Exact Sciences, University of São Paulo, São Paulo, Brazil
- <sup>3</sup> Present Address: Facultad de Ingeniería y Arquitectura, Universidad Peruana Unión, Lima, Perú
- <sup>4</sup> Present Address: Instituto de Estadística, Universidad Austral de Chile, Valdivia, Chile
- <sup>5</sup> Interdisciplinary Center for Atmospheric and Astro-Statistical Studies, Universidad de Valparaíso, Valparaíso, Chile

## 1 Introduction

As was noted by some authors in nonparametric regression analysis, the estimation of the nonparametric function is influenced when there are outlying observations. For this reason, diagnostic analysis is a fundamental process in the statistical modelling and have been largely investigated in the statistical literature. Some of the diagnostic techniques frequently used are global influence (elimination of cases) and local influence. The majority of the works have given emphasis in studying the effect of eliminating observations on the results from the fitted model.

This approach has also been extended to nonparametric and semiparametric models. For example, Eubank (1984, 1985) derives influence diagnostic measures based on the leverage and residuals for spline regression. Silverman (1985) discusses the application of residuals in spline regression. Eubank and Gunst (1986) derive some influence diagnostic measures for penalized least-squares estimates from a Bayesian perspective. Eubank and Thomas (1993) propose diagnostic tests and graphics for assessing heteroscedasticity in spline regression. Kim (1996) discusses the application of residuals, leverage and Cook-type distance for smoothing spline. Wei (2004) presents some influence diagnostic and robustness measures for smoothing spline. Kim et al. (2002) derive influence measures for the partial linear models based on residuals and leverage for the estimates of the regression coefficients and the nonparametric function suggested in Speckman (1988). Fung et al. (2002) studied influence diagnostics for normal semiparametric mixed models with longitudinal data. They consider single influential case or subject for the maximum penalized likelihood estimates suggested in Zhang et al. (1998). Peña (2005) defined for normal regression model a new way to measure the influence of an observation based on how this observation is being influenced by the rest of the data and Türkan and Toktamis (2013) extended these diagnostic measure for semiparametric regression model. Vanegas and Paula (2016) studied the log-symmetric regression models. In this context, the mean and skewness were modeled through nonparametric functions based on P-splines and cubic spline considering an arbitrary number of knots. In addition, they developed some diagnostic measures which were computerized in R-project. The main idea of the local influence technique, introduced by Cook (1986), is to evaluate the sensitivity of the parameters estimators when small perturbations are introduced in the assumptions of the model or in the data (for example, in the response or explanatory variables). This method has the advantage, with respect to elimination of cases, that it is not necessary to calculate the estimates of the parameters for each case excluded. Some of the works related to the technique of local influence are the following. Thomas (1991) developed local influence diagnostics for the smoothing parameter under nonparametric regression model. Ibacache-Pulgar et al. (2012, 2013) derived the local influence curvature for elliptical semiparametric mixed and symmetric semiparametric additive models, respectively. Ferreira and Paula (2017) extended the local influence technique for different perturbation schemes considering a skew-normal partially linear model and Emami (2018) applied the local influence analysis to the Liu penalized least squares estimators. Recently, Zhu et al. (2003) and Ibacache-Pulgar and Paula (2011) to

provide local influence measures under different perturbation schemes in normal and Student-t partially linear models, respectively. The aim of the work is to derive some diagnostic measures such as global and local influence in nonparametric regression model under symmetric random errors. Here, the nonparametric components is estimated by incorporating a quadratic penalty function in the log-likelihood function and whose solution leads to cubic spline functions. The work is presented as follows. In Sect. 2, we introduce the nonparametric regression models under symmetric random errors and a penalized log-likelihood function is proposed for the parameter estimation. In Sect. 3, we propose some diagnostic measures based on the global influence and we develop the local influence method for various perturbations schemes. In Sect. 4, a set of real data is used to illustrate the methodologies proposed in this paper. A simulation study to evaluate the effectiveness of the diagnostic measures derived in this paper is presented in Sect. 5. The last section deals with some concluding remarks.

## 2 Nonparametric regression model with symmetric random errors

In this section we present the nonparametric regression model under symmetrical random errors and penalized likelihood function, which is often required for maximizing the penalized likelihood function. The symmetric class (see, for instance, Fang et al. (1990)) includes all symmetric continuous distributions such as normal, Student-t, power exponential, logistics I and II, contaminated normal, among others. The variety of error distributions with different kurtosis coefficients gives more flexibility for modelling data sets from light- and heavy-tailed distributions.

### 2.1 The model

The nonparametric regression model is a powerful tool in statistical modelling due to their flexibility to model explanatory variable effect that can contribute in the nonparametric way, when there is no linear relationship between the variables. Such models assume the following relationship between the response and the explanatory variable values:

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where  $y_i$  denotes the  $i$ th response value at design point  $t_i$ ,  $f(\cdot)$  is an unknown smooth function that belongs to the Sobolev function space  $W_2^2[a, b] = \{f : f \in [a, b], f, f^{(1)} \text{ absolutely continuous and } f^{(2)} \in L^2[a, b]\}$ , where  $f^{(2)}(t) = \frac{d^2}{dt^2}f(t)$ , and  $\epsilon_i$  are independent random errors such that  $\epsilon_i$  follows a symmetric distribution with location parameter 0, scale parameter  $\phi > 0$  and density function, given by

$$h_{\epsilon_i}(\epsilon) = \frac{1}{\sqrt{\phi}} g(\epsilon^2/\phi), \quad (2)$$

where  $\epsilon \in \mathcal{R}$  and  $g : \mathcal{R} \rightarrow [0, \infty]$  is typically known as the density generator function that satisfies

$$\int_0^\infty g(u)du < \infty \quad \text{with} \quad u = \frac{(y - \mu)^2}{\phi}.$$

Let now  $\mathbf{f} = (\xi_1, \dots, \xi_r)^T$  a  $(r \times 1)$  vector such that  $\xi_j = f(t_j^0)$ , with  $t_j^0 (j = 1, \dots, r)$  denoting the distinct and ordered values of the explanatory variable  $t$  called knots, and  $\mathbf{N}$  an  $(n \times r)$  incidence matrix whose  $(i, j)$ th element equals the indicator function  $I(t_i = t_j^0)$  for  $j = 1, \dots, r$  and  $i = 1, \dots, n$ . Then, we may rewrite model (1) as follows:

$$\mathbf{y} = \mathbf{Nf} + \boldsymbol{\epsilon}, \quad (3)$$

where  $\mathbf{y}$  is the  $(n \times 1)$  vector of response values and  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$  is a  $(n \times 1)$  vector of random errors.

## 2.2 The penalized log-likelihood function

From (2) we have that  $y_i$  follows a symmetric distribution with location parameter  $\mu_i$  and scale parameter  $\phi$  whose density function is given by

$$h_y(y_i) = \frac{1}{\sqrt{\phi}} g(\delta_i),$$

where  $\delta_i = \phi^{-1}(y_i - \mu_i)^2$  and  $\mu_i = \mathbf{n}_i^T \mathbf{f}$ , with  $\mathbf{n}_i^T$  denoting the  $i$ th row of the incidence matrix  $\mathbf{N}$ . We have, when they exist, that  $E(y_i) = \mu_i = \mathbf{n}_i^T \mathbf{f}$  and  $\text{Var}(y_i) = \varrho\phi$ , where  $\varrho > 0$  is a constant that may be obtained from the derivative of the characteristic function (see, for instance, Fang et al. (1990)). Then, the log-likelihood function for  $\boldsymbol{\theta} = (\mathbf{f}^T, \phi)^T$  is given by

$$L(\boldsymbol{\theta}) = -\frac{n}{2} \log \phi + \sum_{i=1}^n \log \{g(\delta_i)\}, \quad (4)$$

where  $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathcal{R}^{p^*}$ , with  $p^* = r + 1$ . A well known procedure for estimating  $f$  is based on the idea of log-likelihood penalization and consists in incorporating a penalty function in the log-likelihood function. Following, for instance, Green and Silverman (1993), we will consider the penalty function given by

$$J(f) = -\frac{\alpha}{2} \int_a^b [f^{(2)}(t)]^2 dt = -\frac{\alpha}{2} \mathbf{f}^T \mathbf{Kf},$$

where  $\mathbf{K} \in \mathcal{R}^{r \times r}$  is a positive-definite matrix that depends only on the knots  $t_j^0$ . Therefore, the penalized log-likelihood function associated to  $\boldsymbol{\theta}$  can be expressed as

$$L_p(\boldsymbol{\theta}, \alpha) = L(\boldsymbol{\theta}) - \frac{\alpha}{2} \mathbf{f}^T \mathbf{Kf}, \quad (5)$$

where the smoothing parameter  $\alpha$  controls the tradeoff between goodness of fit, measured by large values of  $L(\boldsymbol{\theta})$ , and the smoothness estimated function, measured by small values of  $J(f)$ . It is important to note that large values of  $\alpha$  generate smoother

curves while smaller values produce more oscillating curves. In practice situations the smoothing parameter should be selected from the data. When a smoothing spline is used, for example, it is usual to consider the cross-validation method or the generalized cross-validation method Craven and Wahba (1978). Alternatively, this parameter may be selected by applying the Akaike information criterion (Akaike, 1973) or the Bayesian information criterion (Schwarz, 1978).

An alternative approach for estimating the smooth function is to use a set of base functions which are more local in their effects compared to other methods such as fourier expansion. For example, Eilers and Marx (1996) proposed P-splines, whose penalty is direct on the coefficients, instead the curve, and is based on finite differences of the coefficients of adjacent B-splines (Boor, 1978), allowing a good discrete approximation to the integrated square of the  $d$ th derivative. Such penalties are more flexible since it is independent of the degree of the polynomial used to construct the B-splines. The study of other types of base can be found, for instance, in Wood (2003).

### 2.3 Estimation process

Differentiating the penalized log-likelihood function (5) with respect to  $\theta = (\mathbf{f}^T, \phi)^T$  and equating the elements to zero, we have the following estimating equations:

$$\frac{1}{\phi} \mathbf{N}^T \mathbf{D}_v (\mathbf{y} - \mathbf{Nf}) - \alpha \mathbf{Kf} = 0$$

$$(2\phi)^{-1} \left\{ \phi^{-1} (\mathbf{y} - \mathbf{Nf})^T \mathbf{D}_v (\mathbf{y} - \mathbf{Nf}) - n \right\} = 0,$$

with  $\mathbf{D}_v = \text{diag}_{1 \leq i \leq n} (v_i)$  and  $v_i = -2\zeta_i$ , with  $\zeta_i = \frac{d \log g(\delta_i)}{d\delta_i}$ . The quantity  $v_i$ 's can be interpreted as weights since  $g(\delta_i)$  is, for the majority of the symmetrical distributions, a positive decreasing function. Exceptions are Kotz, generalized Kotz and double exponential distributions. Consequently, the solution to estimating equations leads to the following two-step iterative process:

**Step 1** Denoted by  $\theta^{(0)}$  and  $\theta^{(m+1)}$  the starting and current values of  $\theta$ , for  $m = 0, 1, \dots$  and  $\alpha$  fixed. In this stage,  $\mathbf{f}$  can be estimated by using the equation

$$\mathbf{f}^{(m+1)} = \left( \mathbf{N}^T \mathbf{D}_v^{(m)} \mathbf{N} + \alpha \phi^{(m)} \mathbf{K} \right)^{-1} \mathbf{N}^T \mathbf{D}_v^{(m)} \mathbf{y},$$

where  $\mathbf{D}_v^{(m)} = \text{diag}_{1 \leq i \leq n} (v_i) |_{\theta^{(m)}}$

**Step 2** Update  $\phi$  by using the following equation:

$$\phi^{(m+1)} = n^{-1} \left( \mathbf{y} - \mathbf{Nf}^{(m)} \right)^T \mathbf{D}_v^{(m)} \left( \mathbf{y} - \mathbf{Nf}^{(m)} \right).$$

Note that in first step of iterative process we maximize the penalized log-likelihood function with regard  $\mathbf{f}$  given  $\phi$  while in the second step we maximize the penalized log-likelihood function with regard  $\phi$  given  $\mathbf{f}$ . Thus, alternating between stages 1 and 2, this iterative process leads to the maximum penalized likelihood estimate (MPLE) of  $\theta$ .

## 2.4 Approximate standard errors

Analogously to the parametric regression model with normal random errors, the approximate covariance matrix of  $\hat{\boldsymbol{\theta}}$  is derived from the inverse of the expected information matrix. In fact,

$$\widehat{\text{Cov}}_{\text{approx}}(\hat{\boldsymbol{\theta}}) = J_p^{-1} \Big|_{\hat{\boldsymbol{\theta}}},$$

where the matrix  $J_p$  is defined by (see, for instance, Ibacache-Pulgar et al. (2013))

$$J_p = \begin{pmatrix} \mathbf{N}^T \mathbf{D}_d \mathbf{N} + \alpha \mathbf{K} & 0 \\ 0 & \frac{n(4f_g - 1)}{4\phi^2} \end{pmatrix},$$

where  $\mathbf{D}_d = \frac{4d_g}{\phi} \mathbf{I}$ , with  $d_g = E\left(\zeta_g^2(\epsilon_i^2)\epsilon_i^2\right)$ ,  $f_g = E\left(\zeta_g^2(\epsilon_i^2)\epsilon_i^4\right)$ ,  $\mathbf{I}$  denoting an  $(n \times n)$  identity matrix and  $\epsilon_i$  follows a symmetric distribution with location parameter 0, scale parameter  $\phi$  and generator function  $g$ ; see, for example, Ibacache-Pulgar and Paula (2011). In particular, if we are interested in drawing inferences for  $f$  and  $\phi$ , the approximate covariance matrices can be estimated by using the corresponding block-diagonal matrices obtained from  $J_p^{-1}$ , that is,

$$\widehat{\text{Cov}}_{\text{approx}}(\hat{\mathbf{f}}) = (\mathbf{N}^T \mathbf{D}_d \mathbf{N} + \alpha \mathbf{K})^{-1} \Big|_{\hat{\boldsymbol{\theta}}} \quad (6)$$

and

$$\widehat{\text{Cov}}_{\text{approx}}(\hat{\phi}) = \frac{4\phi^2}{n(4f_g - 1)} \Big|_{\hat{\boldsymbol{\theta}}}.$$

An approximate pointwise standard error band (SEB) for  $f(\cdot)$  which allows us to assess how accurate the estimator  $\hat{f}(\cdot)$  is at different locations within the range of interest it is given by  $\hat{f}(t_j^0) \pm 2\sqrt{\widehat{\text{Var}}_{\text{approx}}(\hat{f}(t_j^0))}$ , where  $\widehat{\text{Var}}_{\text{approx}}(\hat{f}(t_j^0))$  denotes the estimated approximate variance of  $\hat{f}(t_j^0)$ , for  $j = 1, \dots, r$ , obtained from Eq. (6).

## 3 Global and local influence measures

In this section, we propose some diagnostics measures such as: leverage, standardized residuals, Cook's distance, Peña measured and likelihood displacement for detecting misspecifications of the error distribution as well as the presence of outlying observations.

### 3.1 Leverage

Considering that in the iterative process the weights  $v_i$ 's are calculated based on previous iteration, it is not possible to obtain an explicit expression for the smoother matrix (hat matrix). However, for the final iteration, the estimates not depend on  $\theta$  and therefore the hat matrix has the same role as the hat matrix in linear regression that is detecting atypical observation for the predicted values of  $y$ . Considering  $\alpha$ ,  $\phi$  and  $\mathbf{D}_v$  fixed, the vector of predictions from the fitted model (3) is given by

$$\hat{\mathbf{y}} = \hat{\boldsymbol{\mu}} = \mathbf{H}(\alpha)\mathbf{y}, \tag{7}$$

where  $\mathbf{H}(\alpha) = \mathbf{N}(\mathbf{N}^T\mathbf{D}_v\mathbf{N} + \alpha\phi\mathbf{K})^{-1}\mathbf{N}^T\mathbf{D}_v = \{h_{ij}(\alpha)\}_{i,j=1}^n$  is called smoother or hat matrix and  $\hat{y}_i = \hat{\mu}_i = \sum_{j=1}^n h_{ij}(\alpha)y_j$ . The diagonal elements of the matrix defined as

$$h_{ii}(\alpha) = \mathbf{n}_i^T(\mathbf{N}^T\mathbf{D}_v\mathbf{N} + \alpha\phi\mathbf{K})^{-1}\mathbf{n}_i v_i$$

are called leverage values and indicate how much influence  $y_i$  has on the fit of  $y_i$ . In general, the hat matrix  $\mathbf{H}(\alpha)$  is not a projection operator, except under normally distributed random errors (see, for instance, Eubank 1984). However, for purposes of diagnostic analysis, this matrix has the same role as the hat matrix in classical linear regression. Then, considering that the hat matrix  $\mathbf{H}(\alpha)$  is not a projection operator, we have that the conditional covariance matrix of the prediction vector is given by

$$\text{Cov}_{\text{approx}}(\hat{\boldsymbol{\mu}}) = \varrho\phi\mathbf{H}(\alpha)\mathbf{H}^T(\alpha). \tag{8}$$

Note that  $\text{Var}_{\text{approx}}(\hat{\mu}_i) = \varrho\phi h_{ii}^2(\alpha)$ , for  $i = 1, \dots, n$ .

### 3.2 Effective degrees of freedom and smoothing parameter

In the literature concerning nonparametric regression models there are different definitions for the degrees of freedom, depending on the context in which they are used. A natural candidate to measure the contribution of the nonparametric component (see, for instance, Buja et al. 1989) is given by

$$\text{df}(\alpha) = \text{tr}\{\mathbf{H}(\alpha)\}, \tag{9}$$

where  $\mathbf{H}(\alpha)$  is defined above. According to Eilers and Marx (1996), if we consider  $\mathbf{Q}_N = \mathbf{N}^T\mathbf{D}_v\mathbf{N}$  and  $\mathbf{Q}_\alpha = \alpha\phi\mathbf{K}$ , the trace of  $\mathbf{H}(\alpha)$  can be written as

$$\begin{aligned} \text{tr}\{\mathbf{H}(\alpha)\} &= \text{tr}\{(\mathbf{Q}_N + \mathbf{Q}_\alpha)^{-1}\mathbf{Q}_N\} \\ &= \text{tr}\{\mathbf{Q}_N^{1/2}(\mathbf{Q}_N + \mathbf{Q}_\alpha)^{-1}\mathbf{Q}_N^{1/2}\} \\ &= \text{tr}\{(\mathbf{I}_{(r \times r)} + \mathbf{Q}_N^{-1/2}\mathbf{Q}_\alpha\mathbf{Q}_N^{-1/2})^{-1}\}. \end{aligned}$$

Consequently, this can be written as (see, for instance, Hastie and Tibshirani 1990)

$$\begin{aligned}
 \text{tr}\{\mathbf{H}(\alpha)\} &= \text{tr}\{(\mathbf{I}_{(r \times r)} + \mathbf{P}_{(r \times r)})^{-1}\} \\
 &= \sum_{j=1}^r \frac{1}{1 + \alpha\phi\ell_j} \\
 &\approx 2 + \sum_{j=3}^r \frac{1}{1 + \alpha\phi\ell_j},
 \end{aligned} \tag{10}$$

where  $\ell_j$ , for  $j = 1, \dots, r$ , are the eigenvalues of  $\mathbf{P} = \mathbf{Q}_N^{-1/2} \mathbf{Q}_\alpha \mathbf{Q}_N^{-1/2}$ . The calculation of  $\text{df}(\alpha)$  from Eq. (9) involves a matrix of order  $n$ , while the calculation from expression (10) involves matrices of order  $r$  ( $r < n$ ), which minimizes the computational cost. It is important to note that (i)  $\text{df}(\alpha) = \text{tr}\{\mathbf{H}(\alpha)\}$  is a monotonically decreasing function of  $\alpha$ , given  $\phi$ ; (ii)  $\text{df}(\alpha) \rightarrow 2 + r$  as  $\alpha \rightarrow 0$ ; (iii)  $\text{df}(\alpha) \rightarrow 2$  as  $\alpha \rightarrow \infty$ ; (iv)  $2 \leq \text{df}(\alpha) \leq 2 + r$ ; and (v) once we have obtained the eigenvalues  $\ell_j$ 's, the calculation of the degrees of freedom  $\text{df}(\alpha)$ , given  $\phi$  and  $\alpha$ , is computationally simple, and can be used to determine a value for the smoothing parameter given its decreasing monotonic relationship.

### 3.3 Residual

As an extension for the nonparametric regression model with symmetric random errors, we will consider the ordinary residual for checking the assumption of our model. It follows from (7) that the vector of ordinary residuals is given by

$$\mathbf{e} = \mathbf{y} - \widehat{\boldsymbol{\mu}} = (\mathbf{I}_n - \mathbf{H}(\alpha))\mathbf{y},$$

where the  $i$ th residual is  $e_i = y_i - \widehat{\mu}_i$ , for  $i = 1, \dots, n$ . From (8), follows that the approximate variance of the  $i$ th ordinary residual is given by

$$\text{Var}_{\text{approx}}(e_i) = \varrho\phi\{1 - h_{ii}(\alpha)\}^2.$$

Then, following Eubank (1985) and Silverman (1985), for  $\phi$  and  $\alpha$  fixed, we can use the standardized residuals,

$$r_i^2 = \frac{e_i^2}{\varrho\phi(1 - h_{ii}(\alpha))^2}$$

for assessing the quality of the fit to  $y_i$ . Alternatively, one might consider the versions of the studentized residuals given by

$$r_i^2 = \frac{e_i^2}{(\varrho\widehat{\phi}^*)(1 - h_{ii}(\alpha))^2} \quad \text{and} \quad t^{[i]} = \frac{e_i^2}{(\varrho\widehat{\phi}^{*[i]})(1 - h_{ii}(\alpha))^2},$$

where  $\widehat{\phi}^*$  is the estimator of  $\phi$  defined as (see, for instance, Wahba 1983)



$$\widehat{\phi}^* = \frac{(\mathbf{y} - \widehat{\boldsymbol{\mu}})^T \mathbf{D}_v (\mathbf{y} - \widehat{\boldsymbol{\mu}})}{\text{tr}\{\mathbf{I} - \mathbf{H}(\alpha)\}},$$

and  $\widehat{\phi}^{*[i]}$  is the estimator of  $\phi$  but without the  $i$ th observation (see, for instance, Eubank 1984), this is,

$$\widehat{\phi}^{*[i]} = \sum_{j=1, j \neq i}^n v_j \frac{[\mathbf{e}_j + v_i h_{ji}(\alpha) \mathbf{e}_i / (1 - v_i h_{ii}(\alpha))]^2}{[n - 1 - \text{tr}\{\mathbf{H}^{[i]}(\alpha)\}]},$$

with

$$\text{tr}\{\mathbf{H}^{[i]}(\alpha)\} = \sum_{j=1, j \neq i}^n \left[ h_{jj}(\alpha) + v_i h_{ji}^2(\alpha) / (1 - v_i h_{ii}(\alpha)) \right].$$

Alternatively, we can consider the quantile residuals proposed by Dunn and Smyth (1996) and which are defined as

$$r_{q_i} = \Phi^{-1}\{F_{Y_i}(y_i, \boldsymbol{\theta})\},$$

where  $\boldsymbol{\theta} = (\mathbf{f}^T, \phi)^T$ ,  $\Phi(\cdot)$  and  $F_{Y_i}(y_i, \boldsymbol{\theta})$  denote, respectively, the fda's of the standard normal distribution and the postulated symmetric distributions. Under the model one has that  $r_{q_i}$  are independent and equally distributed  $N(0, 1)$ , for  $i = 1, \dots, n$ .

### 3.4 Cook's distance

In ordinary linear regression models, Cook (1977) proposed measuring the influence of a point by the squared norm of the vector of forecast changes. Analogously, Hastie and Tibshirani (1990) defined a version of Cook's distance for measuring the influence of the  $i$ th observation on the fit of a nonparametric regression model. Based on those works, we define Cook's distance as follows:

$$D_i = \frac{(\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}}^{[i]})^T \text{Cov}_{\text{approx}}^{-1}(\widehat{\boldsymbol{\mu}}) (\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}}^{[i]})}{\text{tr}\{\mathbf{H}(\alpha)\}},$$

where  $\widehat{\boldsymbol{\mu}}^{[i]} = \widehat{\mathbf{N}}\mathbf{f}^{[i]}$  denote the cubic smoothing spline fit to the data when the  $i$ th observation,  $(t_i, y_i)$ , has been deleted, and by assuming that  $\alpha\mathbf{K}$  remains approximately constant removing the  $i$ th observation one has that

$$\widehat{\mathbf{f}}^{[i]} = \widehat{\mathbf{f}} - \frac{(\mathbf{N}^T \mathbf{D}_v \mathbf{N} + \alpha \phi \mathbf{K})^{-1} \mathbf{n}_i v_i \mathbf{e}_i}{(1 - h_{ii}(\alpha))}. \tag{11}$$

A more convenient way of writing  $D_i$  follows from relation (see, for instance, Craven and Wahba 1978),

$$\begin{aligned}
 \widehat{\mu}_i - \widehat{\mu}_i^{[i]} &= \mathbf{n}_i^T \widehat{\mathbf{f}} - \mathbf{n}_i^T \widehat{\mathbf{f}}^{[i]} \\
 &= \frac{\mathbf{n}_i^T (\mathbf{N}^T \mathbf{D}_v \mathbf{N} + \alpha \phi \mathbf{K})^{-1} \mathbf{n}_i \mathbf{v}_i \mathbf{e}_i}{(1 - h_{ii}(\alpha))} \\
 &= \frac{h_{ii}(\alpha) \mathbf{e}_i}{1 - h_{ii}(\alpha)}.
 \end{aligned}
 \tag{12}$$

Then, by using the Eqs. (11) and (12), we may write statistic  $D_i$  as

$$\begin{aligned}
 D_i &= \frac{(\widehat{\mathbf{f}} - \widehat{\mathbf{f}}^{[i]})^T \mathbf{n}_i \text{Cov}_{\text{approx}}^{-1}(\widehat{\boldsymbol{\mu}}) \mathbf{n}_i^T (\widehat{\mathbf{f}} - \widehat{\mathbf{f}}^{[i]})}{\text{tr}\{\mathbf{H}(\alpha)\}} \\
 &= \frac{1}{\text{tr}\{\mathbf{H}(\alpha)\}} \frac{1}{\varrho \phi h_{ii}^2(\alpha)} \frac{h_{ii}^2(\alpha) \mathbf{e}_i^2}{(1 - h_{ii}(\alpha))^2} \\
 &= \frac{1}{\varrho \phi \text{tr}\{\mathbf{H}(\alpha)\}} \frac{\mathbf{e}_i^2}{(1 - h_{ii}(\alpha))^2} \\
 &= \frac{1}{\text{tr}\{\mathbf{H}(\alpha)\}} r_i^2
 \end{aligned}$$

Then, motivated by practices in ordinary regression analysis, influential observations are detected by large values of their Cook’s distance; this is, when  $D_i > 2\bar{D}$ , with  $\bar{D} = \sum_{i=1}^n \frac{D_i}{n}$ . Note that under normality, this is  $v_i = 1$  and  $\text{Var}(y_i) = \phi h_{ii}(\alpha)$ , the statistic  $D_i$  agree with statistic proposed, for example, by Eubank (1985).

### 3.5 Peña measure

In the normal linear regression model, Peña (2005) proposed a statistic called *S-statistic*, to measure sensitive in the forecast of the  $i$ th observation when each observation is deleted. This type of influence analysis complements the traditional one and is able to indicate features in the data, such as clusters of high-leverage outliers. In addition, this new statistics is very simple to compute and with an intuitive interpretation, that can be a useful tool in regression analysis, mainly with large datasets in high dimension. As an extension on nonparametric regression model with symmetric random errors, such statistic can be defined analogously by

$$S_i = \frac{\|\mathbf{s}_i\|^2}{\text{tr}\{\mathbf{H}(\alpha)\} \mathbf{Var}_{\text{approx}}^{-1}(\widehat{\mu}_i)},$$

where

$$\mathbf{s}_i = \left( \widehat{\mu}_i - \widehat{\mu}_i^{[1]}, \dots, \widehat{\mu}_i - \widehat{\mu}_i^{[n]} \right)^T$$

Considering  $\text{Var}_{\text{approx}}(\widehat{\mu}_i) = \varrho \phi h_{ii}^2(\alpha)$  and by using the identity (see, for instance, Eubank 1985)

$$\widehat{\mu}_i - \widehat{\mu}_i^{[j]} = \frac{h_{ij}(\alpha)\mathbf{e}_j}{(1 - h_{jj}(\alpha))},$$

we can write statistic  $S_i$  as

$$\begin{aligned} S_i &= \frac{\|\mathbf{s}_i\|^2}{\text{tr}\{\mathbf{H}(\alpha)\} \varrho \widehat{\phi} h_{ii}^2(\alpha)} \\ &= \frac{1}{\varrho \widehat{\phi} \text{tr}\{\mathbf{H}(\alpha)\} h_{ii}^2(\alpha)} \sum_{j=1}^n \frac{h_{ij}^2(\alpha)\mathbf{e}_j^2}{(1 - h_{jj}(\alpha))^2} \\ &= \sum_{j=1}^n \rho_{ji}^2 D_j, \end{aligned}$$

where  $\rho_{ji} = \{h_{ij}(\alpha)/h_{ii}(\alpha)h_{jj}(\alpha)\} \leq 1$  is the correlation between forecast  $\widehat{\mu}_i - \widehat{\mu}_j$  and  $D_j = \frac{1}{\text{tr}\{\mathbf{H}(\alpha)\}} r_j^2$ . An alternative way of writing  $S_i$  is given by

$$S_i = \frac{1}{\varrho \widehat{\phi} \text{tr}\{\mathbf{H}(\alpha)\}} \sum_{j=1}^n w_{ji}(\alpha) \mathbf{e}_j^{[j]2},$$

where  $\mathbf{e}_j^{[j]} = \mathbf{e}_j / (1 - h_{jj}(\alpha))$  and  $w_{ji}(\alpha) = h_{ij}^2(\alpha) / h_{ii}^2(\alpha)$ . In this case, the cutoff point is considered as  $4.5MAD(S_i)$ , where  $MAD(S_i) = \text{median}|S_i - \text{med}(S)|$ , with  $\text{med}(S)$  being the median of the  $S_i$  values. An observation is considered heterogeneous if  $|S_i - \text{med}(S)| \geq \text{cutoff}$ . The properties of the  $S$ -statistic are discussed in Peña (2005).

### 3.6 Likelihood displacement

Let  $L_p(\boldsymbol{\theta}, \alpha)$  the penalized log-likelihood function under nonparametric regression model with symmetric random errors and  $\boldsymbol{\theta} = (\mathbf{f}^T, \phi)^T$ . Following Cook (1986), we define the penalized likelihood displacement as

$$LD_i(\boldsymbol{\theta}, \alpha) = 2 \left[ L_p(\widehat{\boldsymbol{\theta}}, \alpha) - L_p(\widehat{\boldsymbol{\theta}}^{[i]}, \alpha) \right],$$

where  $\widehat{\boldsymbol{\theta}}^{[i]} = \left( \widehat{\mathbf{f}}^{[i]T}, \widehat{\phi}^{[i]} \right)^T$  denotes the MPLE of  $\boldsymbol{\theta}$  by dropping the  $i$ th observation,  $L_p(\widehat{\boldsymbol{\theta}}, \alpha)$  is the penalized log-likelihood function evaluated in  $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ , and

$$L_p(\widehat{\boldsymbol{\theta}}^{[i]}, \alpha) = -\frac{n}{2} \log \widehat{\phi}^{[i]} + \sum_{i=1}^n \log \left( g \left( \widehat{\delta}_i^{[i]} \right) \right) - \frac{\alpha}{2} \widehat{\mathbf{f}}^{[i]T} \mathbf{K} \widehat{\mathbf{f}}^{[i]}.$$

Then, we have that likelihood displacement may be expressed in the form

$$LD_i(\boldsymbol{\theta}, \alpha) = n \log \left( \frac{\widehat{\phi}^{[i]}}{\widehat{\phi}} \right) + 2 \sum_{i=1}^n \log \left( \frac{g(\widehat{\delta}_i)}{g(\widehat{\delta}_i^{[i]})} \right) + \alpha \left( \mathbf{f}^{[i]T} \mathbf{K} \mathbf{f}^{[i]} - \widehat{\mathbf{f}}^T \mathbf{K} \widehat{\mathbf{f}} \right).$$

In particular, under normal random errors we have that the likelihood displacement may be expressed in the form

$$LD_i(\boldsymbol{\theta}, \alpha) = n \log \left[ \left( \frac{n}{n-1} \right) \frac{(n-1-p')}{(r_i^2 + n-1-p')} \right] + \frac{(n-1)(1-h_{ii}(\alpha))^{-1} r_i^2}{(n-1-p')} - 1 - \alpha \text{tr} \left\{ \widetilde{\mathbf{H}}(\alpha) \left( \mathbf{y} \mathbf{y}^T - \mathbf{y}^* \mathbf{y}^{*T} \right) \right\},$$

where  $p' = \text{tr} \{ \mathbf{H}^{[i]}(\alpha) \}$  and  $\widetilde{\mathbf{H}}(\alpha) = \mathbf{H}(\alpha)^T \mathbf{K} \mathbf{H}(\alpha)$ . According Green and Silverman (1993) (Lemma 3.1) the vector  $\widehat{\mathbf{f}}^{[i]}$  satisfies equality  $\widehat{\mathbf{f}}^{[i]} = \mathbf{H}(\alpha) \mathbf{y}^*$ , where the vector  $\mathbf{y}^*$  is defined by

$$y_j^* = y_j \quad \text{for } j \neq i$$

$$y_i^* = \widehat{f}^{[i]}(t_i).$$

Note that part of  $LD_i(\boldsymbol{\theta}, \alpha)$ , due to the nonparametric regression model, coincides with expression (5.2.9) given in Cook and Weisberg (1982) under normal linear regression model.

### 3.7 Local influence measure

Let  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$  be an  $(n \times 1)$  vector of perturbations restricted to some open subset  $\Omega \in \mathcal{R}^n$  and the logarithm of the perturbed penalized likelihood denoted by  $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})$ . Suppose that there is a point  $\omega_0 \in \Omega$  that represents no perturbation of the data so that  $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega}_0) = L_p(\boldsymbol{\theta}, \alpha)$ . To assess the influence of minor perturbations on  $\widehat{\boldsymbol{\theta}}$ , we consider the likelihood displacement  $LD(\boldsymbol{\omega}) = 2 \left[ L_p(\widehat{\boldsymbol{\theta}}, \alpha) - L_p(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}, \alpha) \right] \geq 0$ , where  $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$  is the MPLE under  $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})$ . The measure  $LD(\boldsymbol{\omega})$  is useful for assessing the distance between  $\widehat{\boldsymbol{\theta}}$  and  $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ . Cook (1986) suggested studying the local behavior of  $LD(\boldsymbol{\omega})$  around  $\boldsymbol{\omega}_0$ . The procedure consists in selecting a unit direction  $\ell \in \Omega (\|\ell\| = 1)$ , and then to consider the plot of  $LD(\boldsymbol{\omega}_0 + a\ell)$  against  $a$ , where  $a \in \mathcal{R}$ . This plot is called lifted line. Each lifted line can be characterized by considering the normal curvature  $C_\ell(\boldsymbol{\theta})$  around  $a = 0$ . The suggestion is to consider the direction  $\ell = \ell_{\max}$  corresponding to the largest curvature  $C_{\ell_{\max}}(\boldsymbol{\theta})$ . The index plot of  $\ell_{\max}$  may reveal those observations that under small perturbations exercise notable influence on  $LD(\boldsymbol{\omega})$ . According to Cook (1986) the normal curvature in the unitary direction  $\ell$  is given by  $C_\ell(\boldsymbol{\theta}) = -2 \left\{ \ell^T \mathbf{A}_p^T \mathbf{L}_p^{-1} \mathbf{A}_p \ell \right\}$ , where  $\mathbf{L}_p$  is the Hessian matrix and  $\mathbf{A}_p$  is the perturbation matrix. These matrices are defined in the appendix.

## 4 Application and results

To illustrate the methodology described in this paper we consider the house price data set that has been analyzed by various authors. The aim of the study is to assess the association of the house prices with the air quality of the neighborhood by using regression models. The response variable LMV (logarithm of the median house price in USD 1000) is related with 14 explanatory variables. Altogether there are 506 observations. Here we only consider the explanatory variable LSTAT (% lower status of the population) due to its non-linear relationship with the response variable; see Fig. 1 .

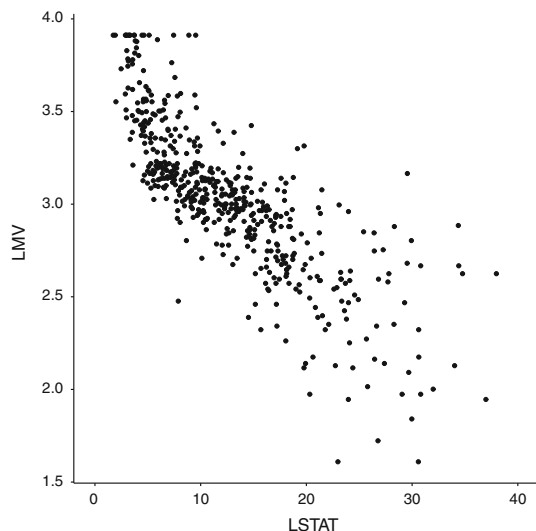
Indeed, we consider the nonparametric regression model

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, \dots, 506,$$

where  $y_i$  denotes the logarithm value of the median house price in USD 1000 and  $t_i$  denotes the value of % lower status of the population. For comparative purposes, we will assume that random errors  $\epsilon_i$  follow a normal, Student-t and power exponential distribution. The smoothing parameters associated with both models were selected such that the effective degrees of freedom are close to 3. The degrees of freedom  $\nu = 5$  and shape parameter ( $k = 0.3$ ) for the Student-t and power exponential models, respectively, were obtained by applying the Akaike information criterion. Table 1 shows a summary of the fit of models.

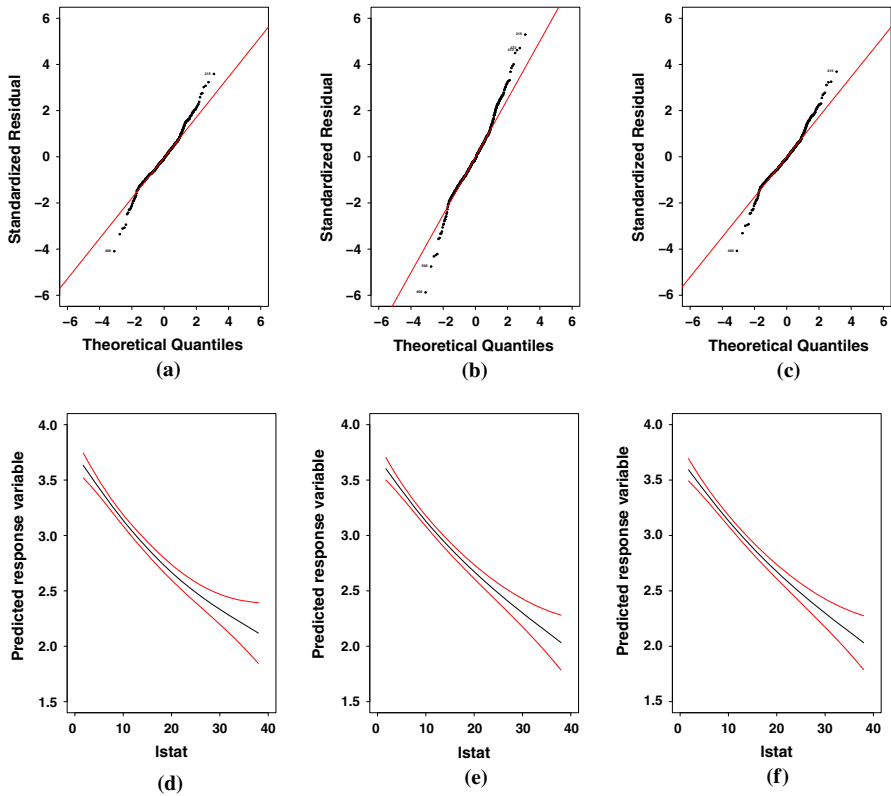
Comparing these results, we observe that the estimate  $\hat{\phi}$  under the power exponential model is lower (as is the standard error) with respect to the normal and Student-t models, however these are not comparable with each other. In addition, we may notice that the AIC value under the Student-t model is smaller compared to the normal and power exponential models, indicating a superiority of the heavy-tailed model, which is confirmed through the normal probability plots presented in Fig. 2a–

**Fig. 1** Scatter plot: LMV versus LSTAT



**Table 1** Maximum penalized likelihood estimates, estimated standard errors and AIC values under normal, Student-t ( $v = 5$ ) and power exponential ( $k = 0.3$ ) models fitted to house price data

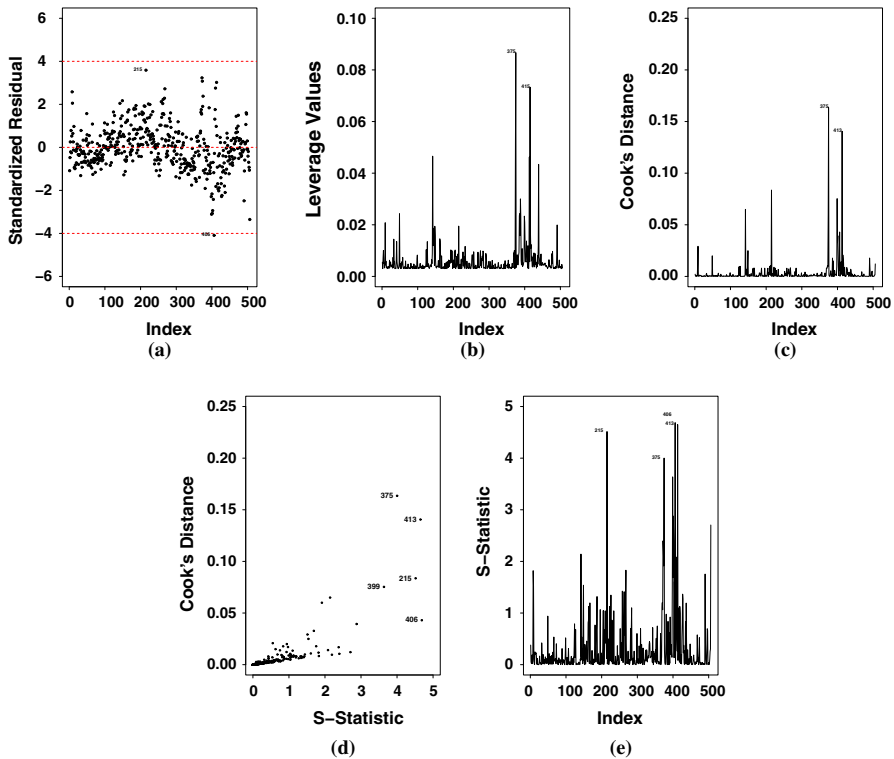
	Normal		Student-t		Power exponential	
	Estimate	SE	Estimate	SE	Estimate	SE
$\phi$	0.0542	0.0034	0.0356	0.0028	0.031	0.0022
AIC	-34.8231		-50.5265		-45.4996	



**Fig. 2** Scatter plots: **a** normal probability plot under normal errors, **b** normal probability plot under Student-t errors, **c** normal probability plot under power exponential errors, **d** estimated smooth function under normal errors, **e** estimated smooth function under Student-t errors, **f** estimated smooth function under power exponential errors, with its approximate pointwise standard error band denoted by red lines

c. The estimated smooth function under three models and their corresponding approximate standard error band are presented in Fig. 2d–f.

Figs. 3, 4 and 5 show some plots of residuals and global influence under normal, Student- t and power exponential errors, respectively. Specifically, the index plots of the residuals (left side of the first row), leverage values (right side of the first row) and Cook’s distance (left side of the second row), the scatterplot Cook’s distance

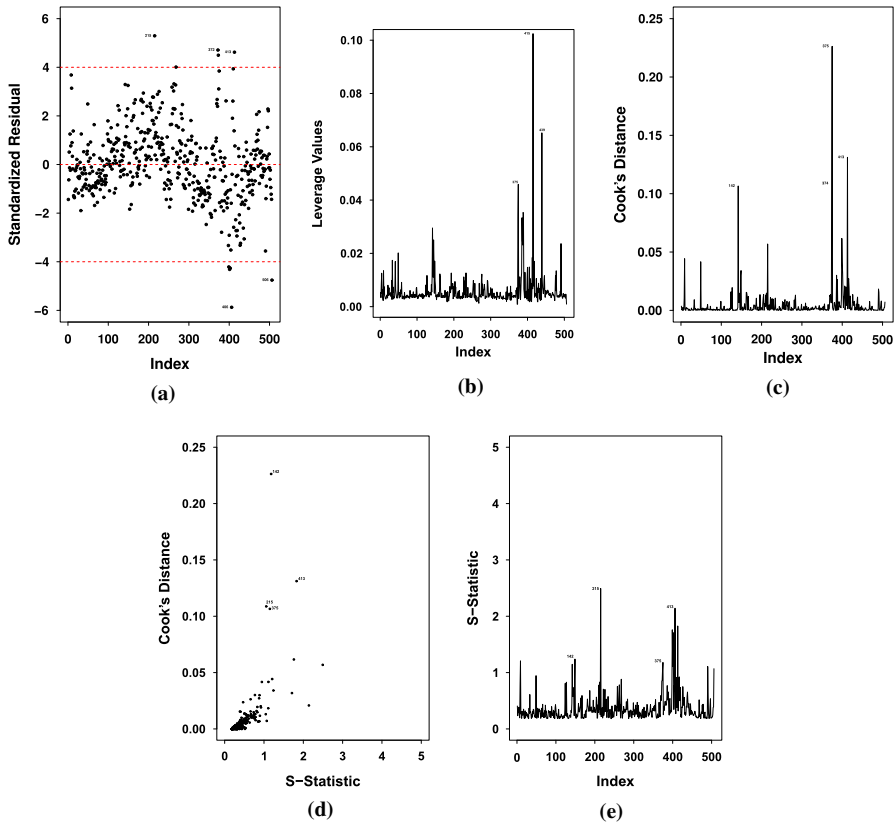


**Fig. 3** Global influence analysis of the Boston data under normal errors: **a** index plot of residuals, **b** index plot of leverage values, **c** index plot of Cook's distance, dispersion plot of *S*-statistics, **d** Cook's distance, **e** index plot of *S*-statistics

versus the *S*-statistic (right side of the second row), and the index plot of the *S*-statistic (bottom of the figure) considering the cutoff point proposed by Peña (2005).

In Fig. 3 (under normal errors), the plot of the residual does not show the presence of outliers, while the plot of the leverage values reveals that the observations #375 and #415 are intermediate-leverage outliers. For its part, Cook's distance plot shows that observations #375 and #413 are outliers. On the other hand, the scatterplot of Cook's distance versus the *S*-statistic clearly indicates the presence of some outliers; this is, the observations #375, #399, #406, #413 and #415. This is confirmed by comparing the values of *S*-statistic with the cutoff point mentioned above.

In Fig. 4 (under Student-*t* errors), the plot of the residual shows the presence of five outliers, observations #215, #372, #406, #413 and #506, while the plot of the leverage values reveals that the observation #375, #415 and #439 are intermediate-leverage outliers. Cook's distance plot shows that observations #142, #374, #375 and #413 are outliers, while the plot of dispersion between Cook's distance and the *S*-statistic indicates gain the presence of some outliers, specifically, observations #142, #215, #375 and #413 being confirmed, for some observations, by comparing the values of *S*-statistic.

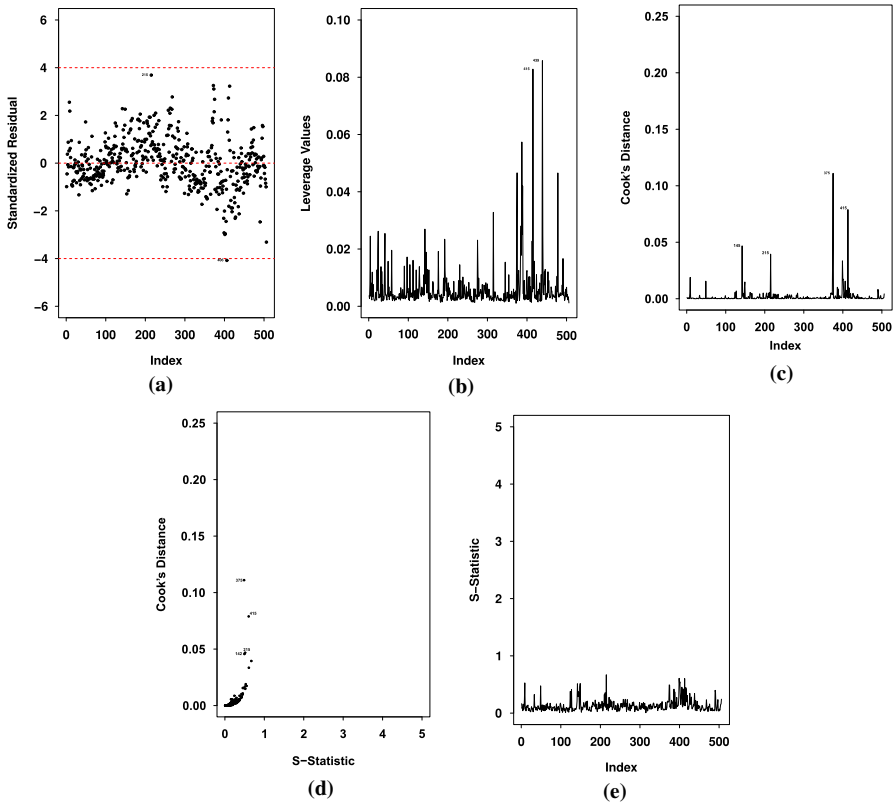


**Fig. 4** Influence analysis of the Boston data under Student-t errors: **a** index plot of residuals, **b** index plot of leverage values, **c** index plot of Cook's distance, dispersion plot of  $S$ -statistics, **d** Cook's distance, **e** index plot of  $S$ -statistics

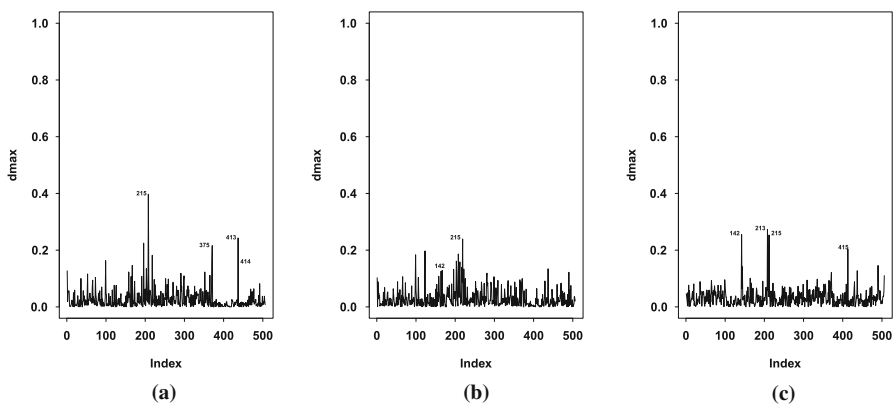
In Fig. 5 (under power exponential errors), the plot of the residual does not show the presence of outliers, while the plot of the leverage values reveals that the observations #415 and #439 are intermediate-leverage outliers. For its part, Cook's distance plot shows that observation #145, #215, #375 and #415 are outliers. On the other hand, the scatterplot of Cook's distance versus the  $S$ -statistic clearly indicates the presence of some outliers; this is, the observations #142, #215, #375 and #415. This is confirmed by comparing the values of  $S$ -statistic. It is important to mention that the observations #215, and #406 that appear as possible outliers under the Student-t model, the estimation process assigns them small weights, confirming the robust aspects of the MPLEs against outlying observations under heavier-tailed error models; see Fig. 8.

On the other hand, Figs. 6 and 7 show the local influence plots for the case-weight and response variable perturbation schemes assuming normal, Student-t and power exponential errors, respectively. The results of local influence for  $f$  obtained under the scale perturbation scheme are the same as those obtained for the case-weight scheme, and therefore are omitted. Based on Fig. 6 we notice that observations 215#, 375#, #413 and #414 are more influential under the normal model (left side) whereas the

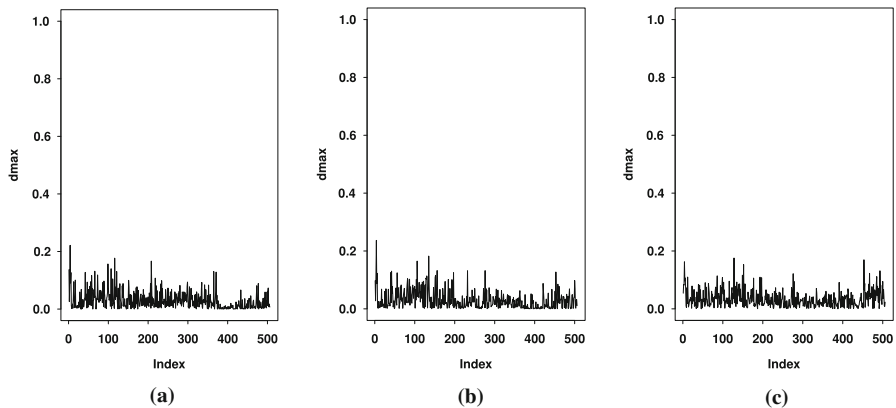




**Fig. 5** Influence analysis of the Boston data under power exponential errors: **a** index plot of residuals, **b** index plot of leverage values, **c** index plot of Cook's distance, dispersion plot of *S*-statistics, **d** Cook's distance, **e** index plot of *S*-statistics



**Fig. 6** Local influence analysis of the Boston data under case-weight perturbation: **a** normal errors, **b** Student-t errors, **c** power exponential errors



**Fig. 7** Local influence analysis of the Boston data under response variable perturbation: **a** normal errors, **b** Student-t errors, **c** power exponential errors

observations #215 and #142 appear slightly influential under Student-t model (middle). Under power exponential model, the 142#, 213#, #215 and #415 are more influential (right side). Looking at Fig. 7 we observe that no observation appears highly influential under the three models. Based on these local influence graphics we can conclude that  $\hat{f}$  appear to be less sensitive in the Student-t model under case-weight perturbation, whereas the sensitivity of  $\hat{f}$  appears to be similar for the three fitted models under response variable perturbation.

It is important to note that for Student-t model the current weight  $v_i^{(m)} = (v + 1) / (v + \delta_i^{(m)})$ , with  $\delta_i^{(m)} = (y_i - \mu_i^{(m)})^2 / \phi^{(m)}$ , is inversely proportional to the distance between the observed value  $y_i$  and its current predicted value  $\mu_i^{(m)}$ , so that outlying observations tend to have small weights in the estimation process. Therefore, we may expected that the MPLEs from the Student-t model are less sensitive to outlying observations than the MPLEs from normal models. Fig. 8 shows the plot between the standardized residual and estimated weights, and estimated weights under Student-t and power exponential models. We can be seen that observations #215 and #406 have a large and small residual, respectively, and a high estimated weight; see Fig. 8a–b. Looking Fig. 8c–d we can see that observations #215 and #406 have a large and small residual respectively, and a high estimated weight. Finally, it is important to note that the iterative process under Student-t model generates a reduction in the weights associated with the observations detected as discrepant. Hence such estimators present some characteristics of robustness similar to the associated with the weight function described by Huber (2004).

## 5 Simulation study

To evaluate further the effectiveness of the diagnostic measures derived in this paper, specifically Cook's distance and Peña Measure, we consider a simulated data set with two forced outliers. We simulated data from the model



$$y_i = \sin\left(\frac{2\pi i}{n}\right) + \epsilon_i \quad (i = 1, \dots, n)$$

where  $\epsilon_i$  follows a symmetric distribution. To perform the simulation, we will consider the models  $N(0, 0.1^2)$  and  $t(0, 0.1^2, 3)$ , and implement an R routine that calculates the proportion of times that forced outlier observations are detected by diagnostic measures  $D_i$  and  $S_i$ , for  $n \in \{25, 50, 10\}$  and 1000 replicates each. To obtain the two outlier observations we change the values of the responses for the 1th and  $n$ th observations in the form  $y_1^* = y_1 + 1.5$  and  $y_n^* = y_n - 1.5$ , respectively. A summary of our simulation study is presented in Table 2; see also Figs. 9, 10, 11 and 12. We can observe, for example, that for normal errors and sample size 50 in 100% of the simulations the 1th and  $n$ th observation is detected as influential for both measures, while for Student-t errors with the same sample size in the 100% and 96.6% of the simulations, the 1th and  $n$ th observation are detected as influential by the measures  $D_i$  and  $S_i$ , respectively. This confirmed in Figs. 13 and 14, respectively, where it is observed that the values of  $D_i$  and  $S_i$  associated with observations 1th and  $n$ th are the highest. In conclusion, in this simulation study, the two forced outliers are correctly identified by the diagnostic measures that we derive in Sect. 3. Similar results were found for the  $LD_i(\theta, \alpha)$  measure but were omitted in this study.

## 6 Concluding remarks

In this paper we consider some diagnostic measures of global and local influence for the nonparametric regression model with symmetric random errors. In the context of global influence, we derive expressions for the residuals, leverage values and Cook's distance similar to those obtained in classical regression under the assumption of normality. In addition, we extend the influence measure proposed by Peña (2005) in order to measure the influence of an observation when it is influenced by the rest of the observations and we use the Boston house price as an illustration. In the context of local influence, we derive the normal curvature for three perturbation schemes, with the purpose of evaluating the sensitivity of the maximum penalized likelihood estimator of smooth function. Both global and local influence measures were illustrated through a set of data. The study provides evidence on the robust aspects of the MPLEs from Student-t with small degrees of freedom against outlying observations. However, these robust aspects do not seem to be extended to all perturbation schemes of the local influence approach, indicating the usefulness to the

**Table 2** Number of times (in %) in which the forced outlier observations were detected as influential through the measures  $D_i$  and  $S_i$

sample size	Normal		Student-t	
	$D_i$	$S_i$	$D_i$	$S_i$
$n = 25$	100	100	99.6	86.7
$n = 50$	100	100	100	96.6
$n = 100$	100	100	99.9	97.1

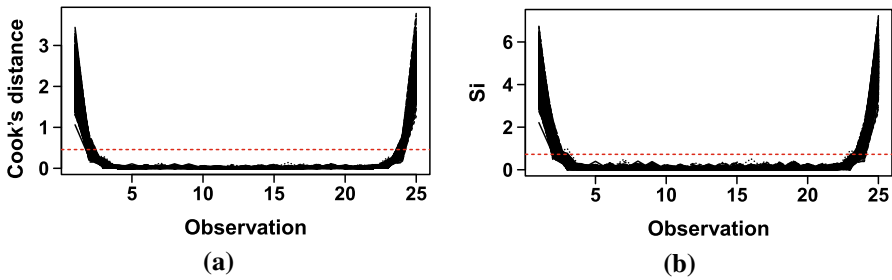


Fig. 9 Index plots of the  $D_i$  and  $S_i$  measures for  $n = 25$  and 1000 simulations under normal errors

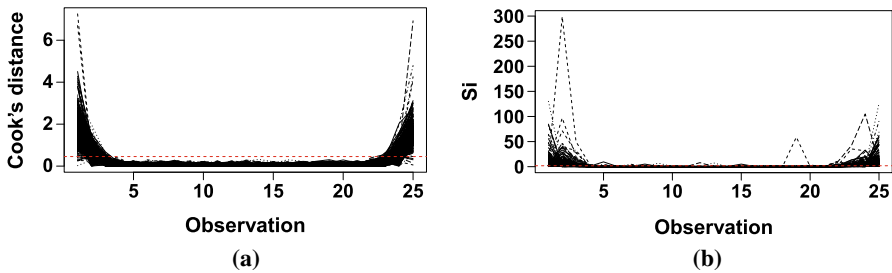


Fig. 10 Index plots of the  $D_i$  and  $S_i$  measures for  $n = 25$  and 1000 simulations under Student-t errors

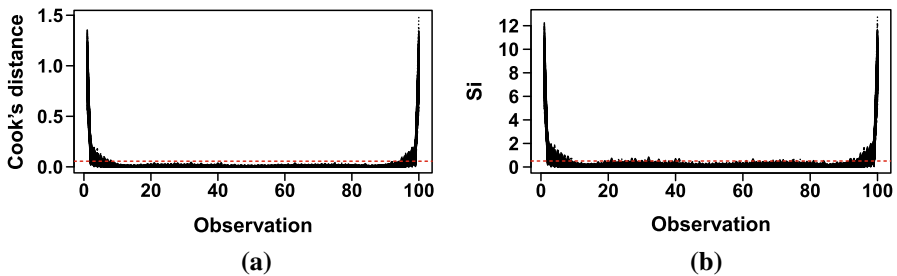


Fig. 11 Index plots of the  $D_i$  and  $S_i$  measures for  $n = 100$  and 1000 simulations under normal errors

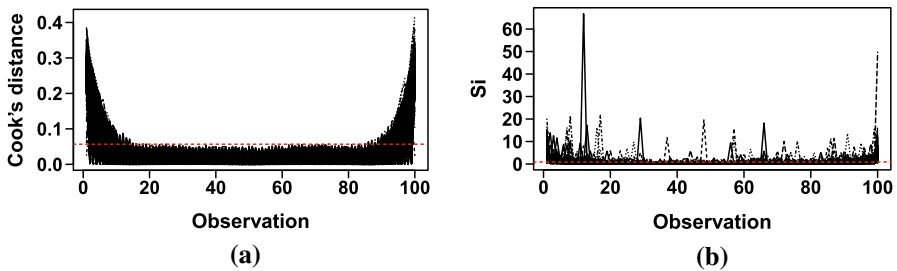


Fig. 12 Index plots of the  $D_i$  and  $S_i$  measures for  $n = 100$  and 1000 simulations under Student-t errors

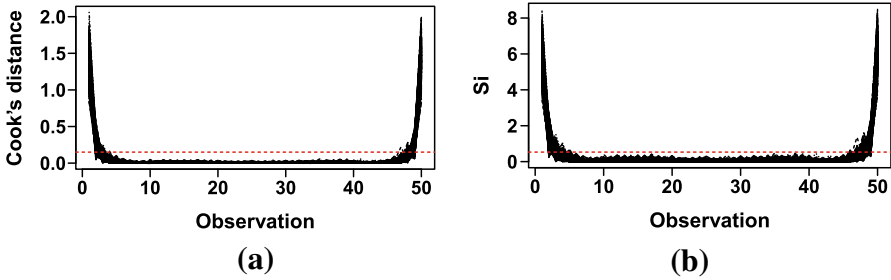


Fig. 13 Index plots of the  $D_i$  and  $S_i$  measures for  $n = 50$  and 1000 simulations under normal errors

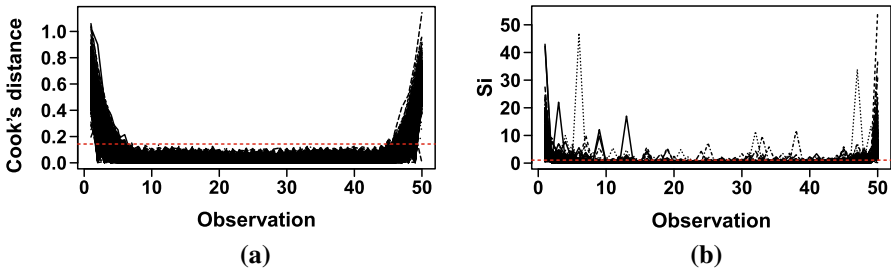


Fig. 14 Index plots of the  $D_i$  and  $S_i$  measures for  $n = 50$  and 1000 simulations under Student-t errors

normal curvature derived in this paper for assessing the sensitivity of the MPLEs from the symmetric nonparametric regression model.

### Appendix

#### Hessian matrix

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

$$\begin{aligned} \frac{\partial^2 L_p(\theta, \alpha)}{\partial \mathbf{f} \partial \mathbf{f}^T} &= -\frac{1}{\phi} \mathbf{N}^T \mathbf{D}_a \mathbf{N} - \alpha \mathbf{K}, \\ \frac{\partial^2 L_p(\theta, \alpha)}{\partial \phi^2} &= \frac{1}{\phi^2} \left\{ \frac{n}{2} + \boldsymbol{\delta}^T \mathbf{D}_{\zeta'} \boldsymbol{\delta} - \frac{1}{\phi} \boldsymbol{\epsilon}^T \mathbf{D}_v \boldsymbol{\epsilon} \right\} \text{ and} \\ \frac{\partial^2 L_p(\theta, \alpha)}{\partial \mathbf{f} \partial \phi} &= \frac{2}{\phi^2} \mathbf{N}^T \mathbf{b}, \end{aligned}$$

where

$$\mathbf{D}_a = \text{diag}_{1 \leq i \leq n}(a_i), \mathbf{D}_{\zeta'} = \text{diag}_{1 \leq i \leq n}(\zeta'_i), \mathbf{b} = (b_1, \dots, b_n)^T, \boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^T, \boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T, a_i = -2(\zeta_i + 2\zeta'_i \delta_i), b_i = (\zeta_i + \zeta'_i \delta_i) \epsilon_i, \epsilon_i = y_i - \mu_i, \zeta'_i = \frac{d\zeta_i}{d\delta_i} \text{ and } \delta_i = \phi^{-1} \epsilon_i^2.$$

### Cases-weight perturbation

Let us consider cases-weight perturbation for the observations in the penalized log-likelihood function as

$$L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega}) = \sum_{i=1}^n \omega_i L_i(\boldsymbol{\theta}) - \frac{\alpha}{2} \mathbf{f}^T \mathbf{K} \mathbf{f},$$

where  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$  is the vector of weights, with  $0 \leq \omega_i \leq 1$ , for  $i = 1, \dots, n$ . In this case, the vector of no perturbation is given by  $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$ . Differentiating  $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})$  with respect to the elements of  $\boldsymbol{\theta}$  and  $\omega_i$ , we obtain after some algebraic manipulation

$$\begin{aligned} \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \omega=\boldsymbol{\omega}_0} &= -\frac{2}{\hat{\phi}} \hat{\zeta}_i \hat{\epsilon}_i \mathbf{n}_i \quad \text{and} \\ \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \phi \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \omega=\boldsymbol{\omega}_0} &= -\frac{1}{2\hat{\phi}} - \frac{1}{2\hat{\phi}} \hat{\zeta}_i \hat{\delta}_i, \end{aligned}$$

where  $\hat{\epsilon}_i = y_i - \hat{\mu}_i$ , for  $i = 1, \dots, n$ .

### Scale perturbation

Under scale parameter perturbation scheme it is assumed that  $y_i \sim S(\mu_i, \omega_i^{-1} \phi, g)$ , where  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$  is the vector of perturbations, with  $\omega_i > 0$ , for  $i = 1, \dots, n$ . In this case, the vector of no perturbation is given by  $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$  such that  $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega} = \boldsymbol{\omega}_0) = L_p(\boldsymbol{\theta}, \alpha)$ . Taking differentials of  $L_p(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})$  with respect to the elements of  $\boldsymbol{\theta}$  and  $\omega_i$ , we obtain after some algebraic manipulation

$$\begin{aligned} \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \omega=\boldsymbol{\omega}_0} &= -\frac{2}{\hat{\phi}} \left\{ \hat{\zeta}_i' \hat{\delta}_i + \hat{\zeta}_i \right\} \hat{\epsilon}_i \mathbf{n}_i^T \\ \left. \frac{\partial^2 L_{p_i}(\boldsymbol{\theta}, \alpha | \boldsymbol{\omega})}{\partial \phi \partial \omega_i} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \omega=\boldsymbol{\omega}_0} &= -\frac{1}{2\hat{\phi}} - \frac{1}{\hat{\phi}} \left\{ \hat{\zeta}_i' \hat{\delta}_i + \hat{\zeta}_i \right\} \hat{\delta}_i, \end{aligned}$$

where  $\hat{\epsilon}_i = y_i - \hat{\mu}_i$ , for  $i = 1, \dots, n$ .

### Response variable perturbation

To perturb the response variable values we consider  $y_{i\omega} = y_i + \omega_i$ , for  $i = 1, \dots, n$ , where  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$  is the vector of perturbations. Here, the vector of no perturbation is given by  $\boldsymbol{\omega}_0 = (0, \dots, 0)^T$  and the perturbed penalized log-likelihood function is constructed from (5) with  $y_i$  replaced by  $y_{i\omega}$ , that is,

$$L_p(\boldsymbol{\theta}, \alpha|\boldsymbol{\omega}) = L(\boldsymbol{\theta}|\boldsymbol{\omega}) - \frac{\alpha}{2} \mathbf{f}^T \mathbf{K} \mathbf{f},$$

where  $L(\cdot)$  is given by (4) with  $\delta_{i\omega} = \phi^{-1}(y_{i\omega} - \mu_i)^2$  in the place of  $\delta_i$ . Differentiating  $L_p(\boldsymbol{\theta}, \alpha|\boldsymbol{\omega})$  with respect to the elements of  $\boldsymbol{\theta}$  and  $\omega_i$ , we obtain, after some algebraic manipulation, that

$$\begin{aligned} \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \alpha|\boldsymbol{\omega})}{\partial \mathbf{f} \partial \omega_i} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} &= -\frac{1}{\hat{\phi}} \left\{ 2\hat{\zeta}_i' \hat{\delta}_i + 2\hat{\zeta}_i \right\} \quad \text{and} \\ \frac{\partial^2 L_{P_i}(\boldsymbol{\theta}, \alpha|\boldsymbol{\omega})}{\partial \phi \partial \omega_i} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} &= -\frac{\hat{\epsilon}_i}{\hat{\phi}^2} \left\{ 2\hat{\zeta}_i' \hat{\delta}_i + 2\hat{\zeta}_i \right\}, \end{aligned}$$

where  $\hat{\epsilon}_i = y_i - \hat{\mu}_i$ , for  $i = 1, \dots, n$ .

**Acknowledgements** This research was funded by FONDECYT 11130704, Chile, and DID S-2017-32, Universidad Austral de Chile, grant.

## References

- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Second international symposium on information theory, pp 267–281
- Boor C (1978) A practical guide to splines. Springer Verlag, Berlin
- Buja A, Hastie T, Tibshirani R (1989) Linear smoothers and additive models. *Ann Stat* 17:453–510
- Cook RD (1977) Detection of influential observation in linear regression. *Technometrics* 19(1):15–18
- Cook RD (1986) Assessment of local influence. *J Roy Stat Soc: Ser B (Methodol)* 48(2):133–155
- Cook RD, Weisberg S (1982) Residuals and influence in regression. Chapman and Hall, New York
- Craven P, Wahba G (1978) Smoothing noisy data with spline functions. *Numer Math* 31(4):377–403
- Dunn P, Smyth G (1996) Randomized quantile residuals. *J Comput Graph Stat* 52(5):236–244
- Eilers P, Marx B (1996) Flexible smoothing with b-splines and penalties. *Stat Sci* 11:89–121
- Emami H (2018) Local influence for Liu estimators in semiparametric linear models. *Stat Pap* 59(2):529–544
- Eubank R (1984) The hat matrix for smoothing splines. *Stat Probab Lett* 2(1):9–14
- Eubank R (1985) Diagnostics for smoothing splines. *J R Stat Soc Ser B (Methodological)* 47:332–341
- Eubank R, Gunst R (1986) Diagnostics for penalized least-squares estimators. *Stat Probab Lett* 4(5):265–272
- Eubank RL, Thomas W (1993) Detecting heteroscedasticity in nonparametric regression. *J Roy Stat Soc: Ser B (Methodol)* 55(1):145–155
- Fang K, Kotz S, Ng KW (1990) Symmetric multivariate and related distributions. Chapman and Hall, London
- Ferreira CS, Paula GA (2017) Estimation and diagnostic for skew-normal partially linear models. *J Appl Stat* 44(16):3033–3053
- Fung WK, Zhu ZY, Wei BC, He X (2002) Influence diagnostics and outlier tests for semiparametric mixed models. *J R Stat Soc Ser B (Statistical Methodology)* 64(3):565–579
- Green PJ, Silverman BW (1993) Nonparametric regression and generalized linear models: a roughness penalty approach. CRC Press, Boca Raton
- Hastie T, Tibshirani R (1990) Generalized additive models, vol 4343. CRC Press, Boca Raton
- Huber PJ (2004) Robust statistics, vol 523. John Wiley & Sons, New Jersey
- Ibacache-Pulgar G, Paula GA (2011) Local influence for student-t partially linear models. *Comput Stat Data Anal* 55(3):1462–1478
- Ibacache-Pulgar G, Paula GA, Galea M (2012) Influence diagnostics for elliptical semiparametric mixed models. *Stat Model* 12(2):165–193
- Ibacache-Pulgar G, Paula GA, Cysneiros FJA (2013) Semiparametric additive models under symmetric distributions. *TEST* 22(1):103–121



- Kim C (1996) Cook's distance in spline smoothing. *Stat Probab Lett* 31(2):139–144
- Kim C, Park BU, Kim W (2002) Influence diagnostics in semiparametric regression models. *Stat Probab Lett* 60(1):49–58
- Peña D (2005) A new statistic for influence in linear regression. *Technometrics* 47(1):1–12
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464
- Silverman BW (1985) Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J Roy Stat Soc: Ser B (Methodol)* 47(1):1–21
- Speckman P (1988) Kernel smoothing in partial linear models. *J Roy Stat Soc: Ser B (Methodol)* 50(3):413–436
- Thomas W (1991) Influence diagnostics for the cross-validated smoothing parameter in spline smoothing. *J Am Stat Assoc* 86(415):693–698
- Türkan S, Toktamis Ö (2013) Detection of influential observations in semiparametric regression model. *Revista Colombiana de Estadística* 36(2):271–284
- Vanegas L, Paula G (2016) An extension of log-symmetric regression models: R codes and applications. *J Stat Comput Simul* 86(9):1709–1735
- Wahba G (1983) Bayesian “confidence intervals” for the cross-validated smoothing spline. *J Roy Stat Soc: Ser B (Methodol)* 45(1):133–150
- Wei WH (2004) Derivatives diagnostics and robustness for smoothing splines. *Comput Stat Data Anal* 46(2):335–356
- Wood SN (2003) Thin plate regression splines. *J R Stat Soc Ser B (Statistical Methodology)* 65(1):95–114
- Zhang D, Lin X, Raz J, Sowers M (1998) Semiparametric stochastic mixed models for longitudinal data. *J Am Stat Assoc* 93(442):710–719
- Zhu ZY, He X, Fung WK (2003) Local influence analysis for penalized gaussian likelihood estimators in partially linear models. *Scand J Stat* 30(4):767–780

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.