Statistics and Its Interface Volume 14 (2021) 165–182

Estimation and diagnostics for partially linear censored regression models based on heavy-tailed distributions

Marcela Nuñez Lemus, Victor H. Lachos,
Christian E. Galarza, and Larissa A. Matos*

In many studies, limited or censored data are collected. This occurs, in several practical situations, for reasons such as limitations of measuring instruments or due to experimental design. So, the responses can be either left, interval or right censored. On the other hand, partially linear models are considered as a flexible generalizations of linear regression models by including a nonparametric component of some covariates in the linear predictor. In this paper, we discuss estimation and diagnostic procedures in partially linear censored regression models with errors following a scale mixture of normal (SMN) distributions. This family of distributions contains a group of well-known heavy-tailed distributions that are often used for robust inference of symmetrical data, such as Student-t, slash and contaminated normal, among others. A simple EM-type algorithm for iteratively computing maximum penalized likelihood (MPL) estimates of the parameters is presented. To examine the performance of the proposed model, case-deletion and local influence techniques are developed to show its robustness against outlying and influential observations. This is performed by sensitivity analysis of the maximum penalized likelihood estimates under some usual perturbation schemes, either in the model or in the data, and by inspecting some proposed diagnostic graphs. We evaluate the finite sample performance of the algorithm and the asymptotic properties of the MPL estimates through empirical experiments. An application to a real dataset is presented to illustrate the effectiveness of the proposed methods. Both estimation procedure and diagnostic tools were implemented in the R PartCensReg package.

Keywords and phrases: Censored regression model, EM-type algorithm, Partially linear models, Local influence, Scale mixtures of normal distributions.

1. INTRODUCTION

The problem of estimation of a regression model where the dependent variable is censored has been studied in different fields, such as econometric analysis and clinical testing, among many others. For example, in AIDS research, the viral load measures may be subject to some lower and upper detection limits, below or above which they are not quantifiable. As a result, the viral load responses are either left or right censored depending on the diagnostic assays used [see, for instance, 32].

In the framework of censored regression (CR) models, the random errors are routinely assumed to follow a normal distribution for mathematical convenience. However, if the random error distribution is non-normal, in particular, if its tails are heavier than normal ones, then the accuracy of the ordinary least squares solutions is lost, introducing biases in the parameter estimates. For more accurate models, a large number of parametric models to extend well-known distributions and to provide flexibility in modeling data have been investigated in recent years. For instance, [1] advocated the use of the Student-t distribution in the context of CR models. More recently, [23] developed diagnostic measures for CR models using the Student-t distribution, including the implementation of an interesting (and simple) expectation-maximization (EM) algorithm for maximum likelihood (ML) estimation. [8, 9] proposed a CR model with observational errors following a SMN distribution (SMN-CR model) from Bayesian and likelihood based perspectives, respectively. They demonstrated the robustness of the SMN-CR model against outliers through extensive simulations.

Partially linear regression (PLR) models belong to the class of semiparametric regression models [see, for instance, 13]. They are quite flexible since the nonparametric component can model nonlinear behavior introduced by some covariates in the model. Linear regression models can be seen as a limiting case of PLR models when the nonparametric component is not considered. Comprehensive surveys are available in [12] and [13]. In the past few years, several works on PLR under flexible error distributions have been published. For instance, [16] developed diagnostic measures for PLR models using the Student-t distribution, [30] derived an iterative estimation process and some diagnostic procedures in PLR with AR(1) symmetrical errors. [7] proposed a PLR model allowing the errors to follow a skew-normal [2]

---

*Corresponding author.

distribution. In the context of partial linear censored regression (PCR) models, [31] proposed the log-symmetric regression model, where the presence of non-informative censored observations is admitted. [3] advocated the use of the SMN class of distributions in PCR (SMN-PCR) models and adopted a Bayesian framework to carry out posterior inference. Although some works involving PCR models with symmetrical distributions have been published recently, so far, to the best of our knowledge, no attempt has been made to study the SMN-PCR model from a likelihood based perspective. In this paper, a fully likelihood-based approach is presented, including the implementation of an efficient EM-type algorithm, the ECME algorithm, for iteratively computing the MPL estimates of the parameters.

Since the classic normal model is very sensitive to outlying observations, the assessment of robustness of the parameter estimates is an important concern. The deletion method, which consists of studying the impact on the parameter estimates after dropping individual observations, is probably the most employed technique to detect influential observations (see [6] and the references therein). Nevertheless, research on the influence of small perturbations in the model(or data) on the parameter estimates has received increasing attention in recent years. This can be achieved by performing local influence analysis, a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs. This research area has received considerable attention in the statistical literature for linear regression models since the seminal work of [5]. However, for the SMN-PCR model, the marginal log-likelihood function is too complex for many applications and a direct application of Cook's approach may be cumbersome, since first and second partial derivatives of this function are involved. [33] presented an approach to perform local influence analysis for general statistical models with missing (or incomplete) data by working with a Q-displacement function, closely related to the conditional expectation of the complete-data log-likelihood used at the *E-step* of the ECME algorithm. This approach produces results very similar to those obtained from Cook's method. Moreover, case-deletion can be also studied by using the Q-displacement function following the approach of [34]. These methods, and their variants, have been applied successfully to perform influence analysis in several CR models, as seen in [25] and [23], among others. In this work, we develop a local influence method using this approach for the SMN-PCR model, showing that it leads to simple influence measures. The proposed estimation and diagnostic method are implemented in the R package `PartCensReg` [20] available in the CRAN repository.

The rest of the paper is organized as follows. The SMN-PCR model is defined in Section 2 after a brief introduction of the SMN class and related properties. In Section 3, we develop an ECME algorithm [21] to perform MPL estimation for the parameters of the proposed model. In Section 4 we present some influence diagnostic techniques, based on case deletion and local influence approaches. Sections 5 and 6 are dedicated to the analysis of real and simulated datasets, respectively. Section 7 concludes with a short discussion of the issues raised by our study and some avenues directions for future research.

2. FORMULATION OF THE MODEL

2.1 Preliminaries

We begin by defining some notation and presenting the basic concepts which are used throughout this work. A normal distribution with mean μ and variance σ^2 is denoted by $\mathrm{N}(\mu, \sigma^2)$,where $\phi\left(\cdot|\mu, \sigma^2\right)$ denotes its probability density function (pdf). Also, $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the pdf and the cumulative distribution function (cdf) of the standard normal distribution; and $F_{SMN}(.)$, $F_{PVII}(\cdot)$, $F_{SL}(\cdot)$ and $F_{CN}(\cdot)$ represent the cdf of the standard SMN distribution, standard slash, standard Pearson type *VII* and the standard contaminated normal distribution, respectively. When X follows a $\mathrm{Gamma}(a, b)$ distribution, we will consider the shape-rate parameterization, i.e., with mean a/b and variance a/b^2, where $a > 0$ and $b > 0$. We use the traditional convention of denoting a random variable (or a random vector) by an upper-case letter and its realization by the corresponding lower-case letter. Random vectors and matrices are denoted by boldface letters. I denotes the indicator function.

A random variable Y is said to have a SMN distribution with location parameter $\mu \in \mathcal{R}$, scale parameter $\sigma^2 \in (0, \infty)$ and an auxiliary vector of parameters $\boldsymbol{\nu} \in \mathcal{R}^k$, denoted by $Y \sim \mathrm{SMN}(\mu, \sigma^2, \boldsymbol{\nu})$, if it has the following stochastic representation:

$$Y \overset{d}{=} \mu + U^{-1/2} Z, \tag{1}$$

where Z and U are independent random variables, $Z \sim \mathrm{N}(0, \sigma^2)$, U is a mixing positive random variable with cdf $H(\cdot|\boldsymbol{\nu})$, with $\boldsymbol{\nu}$ being a scalar or vector parameter indexing the distribution of U and $\overset{d}{=}$ means "has the same distribution as". It is easy to see from (1) that $Y|U = u \sim \mathrm{N}(\mu, u^{-1}\sigma^2)$. Using conditional distribution, the pdf of (1) is:

$$f_{SMN}(y|\mu, \sigma^2, \boldsymbol{\nu}) \\ = (2\pi\sigma^2)^{-1/2} \int_0^\infty u^{1/2} \exp\left[-u\frac{(\mathbf{y}-\mu)^2}{2\sigma^2}\right] dH(u|\boldsymbol{\nu}).$$

The following result is very important for the development of our proposed ECME algorithm. It was provided and proved by Garay et al. [9, Proposition 1], and is an extension of Theorem 1 and Corollary 1 in Genç [10]. Let $Y \sim \mathrm{SMN}(0, 1, \boldsymbol{\nu})$ with scale factor U and mixture distribution $H(\cdot|\boldsymbol{\nu})$. Thus for $a < b$, $\mathrm{E}\left[U^r Y^s | Y \in \mathscr{A}\right]$ for $r \geq 1$,

Table 1. $E_\phi(r,h)$ and $E_\Phi(r,h)$ for some members of the SMN family of distributions

Distribution	$\mathrm{E}_\phi(r,h)$
Student-t	$\frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi}}\left(\frac{\nu}{2}\right)^{\nu/2}\left(\frac{h^2+\nu}{2}\right)^{-\frac{(\nu+2r)}{2}}$
Slash	$\frac{\nu}{\sqrt{2\pi}}\left(\frac{h^2}{2}\right)^{-(\nu+1)}\Gamma\left(\nu+r,\frac{h^2}{2}\right)$
Contaminated Normal	$\varphi\gamma^r\phi(h\sqrt{\gamma})+(1-\varphi)\phi(h)$

Distribution	$\mathrm{E}_\Phi(r,h)$
Student-t	$\frac{\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu}{2}\right)^{-r}F_{PVII}(h\|\nu+2r,\nu)$
Slash	$\left(\frac{\nu}{\nu+1}\right)F_{SL}(h\|\nu+r)$
Contaminated Normal	$\gamma^r F_{CN}(h\|\varphi,\gamma)+(1-\gamma^r)\Phi(h)$

$\mathscr{A}=(a,b)$ and $s=0,1,2$ are given by:

$$\mathrm{E}\left[U^r|Y\in\mathscr{A}\right]=\tau(a,b)\left[\mathrm{E}_\Phi(r,b)-\mathrm{E}_\Phi(r,a)\right], \tag{2}$$

$$\mathrm{E}\left[U^rY|Y\in\mathscr{A}\right]=\tau(a,b)\left[\mathrm{E}_\phi(r-0.5,a)-\mathrm{E}_\phi(r-0.5,b)\right], \tag{3}$$

$$\begin{aligned}\mathrm{E}\left[U^rY^2|Y\in\mathscr{A}\right]=\tau(a,b)\,&[\mathrm{E}_\Phi(r-1,b)-\mathrm{E}_\Phi(r-1,a)\\&+a\mathrm{E}_\phi(r-0.5,a)-b\mathrm{E}_\phi(r-0.5,b)],\end{aligned} \tag{4}$$

with $\tau(a,b)=(F_{SMN}(b)-F_{SMN}(a))^{-1}$ and

$$\mathrm{E}_\phi(r,h)=\mathrm{E}\left[U^r\phi\left(h\,U^{0.5}\right)\right]=\int_0^\infty u^r\phi\left(h\,u^{0.5}\right)dH\left(u|\boldsymbol{\nu}\right),$$

$$\mathrm{E}_\Phi(r,h)=\mathrm{E}\left[U^r\Phi\left(h\,U^{0.5}\right)\right]=\int_0^\infty u^r\Phi\left(h\,u^{0.5}\right)dH\left(u|\boldsymbol{\nu}\right).$$

The class of SMN distributions includes as particular cases well-known distributions like the Pearson type VII, Student-t, slash, contaminated normal distributions, among others. The normal distribution is a special case when U is degenerate in 1, i.e., $\mathrm{P}(U=1)=1$. Therefore, the calculation of $\mathrm{E}_\phi(r,h)$ and $\mathrm{E}_\Phi(r,h)$ will depend on the type of distribution (see Table 1). We refer to [9] for proofs and additional properties.

2.2 The model

Let us consider a partially linear model, where the responses $Y_1,\ldots,Y_n$ are random variables with independent and identically distributed errors according to a SMN distribution. To be more precise, let us write:

$$\begin{aligned}Y_i&=\mathbf{x}_i^\top\boldsymbol{\beta}+f(t_i)+\varepsilon_i,\\\varepsilon_i&\overset{\text{iid}}{\sim}\mathrm{SMN}(0,\sigma^2,\boldsymbol{\nu}),\end{aligned} \tag{5}$$

where Y_i is the response for subject i, $i=1,\ldots,n$, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)^\top$ is a vector of regression parameters of dimension $(p\times 1)$, $\mathbf{x}_i=(x_{1_i},\ldots,x_{p_i})^\top$ is a vector of explanatory variable values $p\times 1$, t_i is a scalar that can represent a value of a continuous covariate, for example time, and $f(\cdot)$ is a smoothing function.

In this work, we are interested in the case where left-censored observations might occur. That is, the observations are of the form:

$$Y_{\text{obs}_i}=\begin{cases}\kappa_i & \text{if}\quad Y_i\le\kappa_i;\\ Y_i & \text{if}\quad Y_i>\kappa_i,\end{cases} \tag{6}$$

$i=1,\ldots,n$, for some threshold point κ_i for the i-th subject. We have chosen to work with the left censored case, but the results are easily extendable to other censoring types. We call the model defined in Equations (5)–(6) the SMN-PCR model.

Alternatively, the model (5) can be written as:

$$Y_i=\mathbf{x}_i^\top\boldsymbol{\beta}+\mathbf{n}_i^\top\mathbf{f}+\varepsilon_i, \tag{7}$$

where $\mathbf{f}=(f(t_1^0),\ldots,f(t_r^0))^\top$ is an $r\times 1$ vector with $t_1^0,\ldots,t_r^0$ being the distinct and ordered values of t_i; $\mathbf{n}_i$ is an $(r\times 1)$ incidence vector with the s-th element equal to the indicator function $I(t_i=t_s^0)$ for $s=1,\ldots,r$. In matrix form, model (7), can be written as:

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{N}\mathbf{f}+\boldsymbol{\varepsilon}, \tag{8}$$

where $\mathbf{Y}=(Y_1,\ldots,Y_n)^\top$ is the response vector of dimension $n\times 1$, $\mathbf{X}$ is an $(n\times p)$ design matrix, $\mathbf{N}$ is an $(n\times r)$ incidence matrix with the (i,s)-th element equal to the indicator function $I(t_i=t_s^0)$, for $s=1,\ldots,r$ and $\boldsymbol{\varepsilon}=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)^\top$ is an $(n\times 1)$ vector of random errors with elements belonging to the SMN class of distributions.

2.3 The log-likelihood function

Let $\mathbf{y}_{\text{obs}}=(y_1,\ldots,y_n)^\top$ be an observed sample of $\mathbf{Y}=(Y_1,\ldots,Y_n)^\top$, containing m censored values of the characteristic of interest. We can partition the observed sample $\mathbf{y}_{\text{obs}}$ into two subsamples of m censored and $n-m$ uncensored values, such that $\mathbf{y}_{\text{obs}}=\{\kappa_1,\ldots,\kappa_m,y_{m+1},\ldots,y_n\}$. Then, the log-likelihood function of the parameter vector $\boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\mathbf{f}^\top,\sigma^2,\boldsymbol{\nu}^\top)^\top$ considering left-censored observations is given by:

$$\begin{aligned}\ell(\boldsymbol{\theta})&=\log\left[\prod_{i=1}^n\left[F_{SMN}\left(\frac{\kappa_i-\mu_i}{\sigma}\right)\right]^{I_i}\left[f_{SMN}(y_i|\mu_i,\sigma^2,\boldsymbol{\nu})\right]^{1-I_i}\right],\\&=\sum_{i=1}^m\log\left[F_{SMN}\left(\frac{\kappa_i-\mu_i}{\sigma}\right)\right]\\&\quad+\sum_{i=m+1}^n\log\left[f_{SMN}(y_i|\mu_i,\sigma^2,\boldsymbol{\nu})\right],\end{aligned} \tag{9}$$

where $\mu_i=\mathbf{x}_i^\top\boldsymbol{\beta}+\mathbf{n}_i^\top\mathbf{f}$, and $I_i=1$ if $y_i\le\kappa_i$ and $I_i=0$ otherwise.

Maximization of Equation (9) without imposing restrictions on the function $\mathbf{f}(\cdot)$ may cause over-fitting and non-identification of $\boldsymbol{\beta}$ [see, for instance, 11]. A well-known procedure is based on the penalized log-likelihood, which consists of incorporating a penalty function in the log-likelihood function, such that:

$$\ell_p(\boldsymbol{\theta}, \alpha) = \ell(\boldsymbol{\theta}) - \frac{\alpha}{2} J(\mathbf{f}), \tag{10}$$

where $\ell_p(\boldsymbol{\theta}, \alpha)$ denotes the penalized log-likelihood function, $J(\mathbf{f})$ is the penalty function over $\mathbf{f}(\cdot)$ and α is a smoothing parameter that controls the tradeoff between goodness-of-fit and the estimated function's smoothness. The maximum penalized likelihood (MPL) estimates are obtained by maximizing the penalized log-likelihood defined in (10). The MPL estimation problem based on an efficient ECME algorithm is considered in the next section.

3. PARAMETER ESTIMATION VIA AN ECME ALGORITHM

To implement the EM method, we require a representation of the model in terms of missing data. First, observe that by Equation (1), if $Y_i \sim \mathrm{SMN}(\mu_i, \sigma^2, \boldsymbol{\nu})$ then:

$$\begin{aligned} Y_i | U_i = u_i &\sim \mathrm{N}(\mu_i, u_i^{-1}\sigma^2), \\ U_i &\sim \mathrm{H}(\cdot|\boldsymbol{\nu}). \end{aligned} \tag{11}$$

This relationship is a convenient hierarchical representation of the SMN-PCR model, and will be useful in the *E-step* of the algorithm.

The key to the development of our ECME algorithm is to consider the augmented dataset $\boldsymbol{z} = \{\kappa_1, \ldots, \kappa_m, y_{m+1}, \ldots, y_n, u_1, \ldots, u_n\}$. As a consequence, we can use the representation in (11) to obtain the complete-data penalized log-likelihood, given by:

$$\ell_{cp}(\boldsymbol{\theta}|\boldsymbol{z}) = -\frac{n}{2}\log\sigma^2 + \frac{1}{2}\sum_{i=1}^{n}\log u_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i(y_i - \mu_i)^2 + \sum_{i=1}^{n}\log h(u_i|\nu) - \frac{\alpha}{2}J(\mathbf{f}) + cte,$$

where $\ell_{cp}(\boldsymbol{\theta}|\boldsymbol{z})$ is the complete penalized log-likelihood function, $h(\cdot|\nu)$ is the density of the mixing variable U and *cte* is a constant independent of the parameter vector $\boldsymbol{\theta}$. In the *E-step* of the algorithm, we must obtain the so-called Q-function,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\ell_{cp}(\boldsymbol{\theta}|\boldsymbol{z})|\mathbf{y}_{\text{obs}}],$$

in which the superscript (k) indicates the estimate of the related parameter at stage k of the algorithm and $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}$ is the conditional expectation of the complete penalized log-likelihood function given the current estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$. Like [16] and [7], we consider the following penalty function:

$$J(\mathbf{f}) = \int_a^b [f''(t)]^2 dt,$$

where $[f''(t)]$ denotes the second derivative of $f(t)$ with [a, b] containing the values t_j^0, for all $j = 1, \ldots, r$. As in [12], we use the natural cubic spline as a solution for the smoothing function $f(\cdot)$, therefore $J(\mathbf{f}) = \mathbf{f}^\top \mathbf{K}\mathbf{f}$, where $\mathbf{K} \in \mathcal{R}^{r\times r}$ is a non-negative definite matrix that depends only on the knot differences. A complete expression of $\mathbf{K}$ may be found, for instance, in [12].

Thus, dropping the constants and given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, the Q-function can be written as:

$$\begin{aligned} &Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) \\ &\quad\propto -\frac{n}{2}\log\widehat{\sigma^2}^{(k)} - \frac{1}{2\widehat{\sigma^2}^{(k)}}\sum_{i=1}^{n}\Big[\xi_{2_i}(\widehat{\boldsymbol{\theta}}^{(k)}) - 2\xi_{1_i}(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\mu}_i^{(k)} \\ &\qquad + \xi_{0_i}(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\mu}_i^{(k)^2}\Big] - \frac{\alpha}{2}\widehat{\mathbf{f}}^{(k)^\top}\mathbf{K}\widehat{\mathbf{f}}^{(k)} \\ &\quad\propto -\frac{n}{2}\log\widehat{\sigma^2}^{(k)} - \frac{1}{2\widehat{\sigma^2}^{(k)}}(\mathbf{1}_n^\top\boldsymbol{\xi}_2^{(k)} - 2\widehat{\boldsymbol{\mu}}^{(k)\top}\boldsymbol{\xi}_1^{(k)} \\ &\qquad + \widehat{\boldsymbol{\mu}}^{(k)\top}\boldsymbol{\Omega}^{(k)}\widehat{\boldsymbol{\mu}}^{(k)}) - \frac{\alpha}{2}\widehat{\mathbf{f}}^{(k)^\top}\mathbf{K}\widehat{\mathbf{f}}^{(k)}, \end{aligned} \tag{12}$$

where $\boldsymbol{\Omega}$ is a diagonal matrix with elements $\xi_{0_i}(\widehat{\boldsymbol{\theta}}^{(k)})$ of dimension $n \times n$, $\boldsymbol{\xi}_1^{(k)}$ $\boldsymbol{\xi}_2^{(k)}$ are vectors of dimension $n \times 1$ with elements $\xi_{1_i}(\widehat{\boldsymbol{\theta}}^{(k)})$, $\xi_{2_i}(\widehat{\boldsymbol{\theta}}^{(k)})$ respectively, $\widehat{\boldsymbol{\mu}}^{(k+1)}$ is the $n \times 1$ vector of means at the k-th iteration and $\mathbf{1}_n$ a $(n \times 1)$ vector of ones.

Therefore, it is clear that the expression of the Q-function depends completely on the knowledge of the expectations

$$\xi_{s_i}(\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i Y_i^s | y_{\text{obs}_i}], \quad s = 0, 1, 2.$$

Thus, for a censored observation i, we have

$$\xi_{s_i}(\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i Y_i^s | Y_i \le \kappa_i], \tag{13}$$

which was obtained in Garay et al. [9, Proposition 1] with expressions given by. On the other hand, for an uncensored observation i, we have

$$\xi_{s_i}(\boldsymbol{\theta}^{(k)}) = y_i^s \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i | Y_i]. \tag{14}$$

The values of $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i|Y_i]$ were computed before by Osorio et al. [27] and are presented in Table 2, with $d(\boldsymbol{\theta}^{(k)}, y_i) = (y_i - \mu_i^{(k)})/\sigma^{(k)}$. Here, $\boldsymbol{\xi}_s^{(k)}$ will denote the vector containing the $\xi_{s_i}(\widehat{\boldsymbol{\theta}}^{(k)})$ elements in which, if the observation i is censored, it will be computed using $\xi_{s_i}(\widehat{\boldsymbol{\theta}}^{(k)})$ given in (13), or else using $\xi_{s_i}(\widehat{\boldsymbol{\theta}}^{(k)})$ as in (14), for $s = 0, 1, 2$. Note that,

Table 2. $E_{\boldsymbol{\theta}^{(k)}}[U_i|Y_i]$ for some members of the SMN family of distributions

Distribution	$E_{\boldsymbol{\theta}^{(k)}}[U_i \mid y_i]$
Student-t	$\dfrac{(\nu+1)}{\nu + d^2(\boldsymbol{\theta}^{(k)}, y_i)}$
Slash	$\dfrac{\Gamma\left(\nu+1.5, d^2(\boldsymbol{\theta}^{(k)}, y_i)/2\right)}{\Gamma\left(\nu+0.5, d^2(\boldsymbol{\theta}^{(k)}, y_i)/2\right)}$
Contaminated normal	$\dfrac{1-\varphi+\varphi\gamma^{1.5}e^{0.5(1-\gamma)d^2(\boldsymbol{\theta}^{(k)},y_i)}}{1-\varphi+\varphi\gamma^{0.5}e^{0.5(1-\gamma)d^2(\boldsymbol{\theta}^{(k)},y_i)}}$

$E_{\boldsymbol{\theta}^{(k)}}[\log h(U_i|\boldsymbol{\nu})|y_{\text{obs}_i}]$ and $E_{\boldsymbol{\theta}^{(k)}}[\log (U_i)\,|y_{\text{obs}_i}]$ depend only on $\boldsymbol{\nu}$, which is assumed known at this stage.

Thus, the proposed ECME algorithm can be summarized in the following steps:

1. ***E-step***: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, compute $\xi_{s_i}(\widehat{\boldsymbol{\theta}}^{(k)})$ for $s = 0, 1, 2$.
2. ***CM-step***: Update $\widehat{\boldsymbol{\theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ over $\boldsymbol{\theta}$, which leads to the following expressions:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left[\sum_{i=1}^{n} \xi_{0_i}(\widehat{\boldsymbol{\theta}}^{(k)})\mathbf{x}_i\mathbf{x}_i^\top\right]^{-1} \\
&\quad \times \sum_{i=1}^{n} \mathbf{x}_i \left[\xi_{1_i}(\widehat{\boldsymbol{\theta}}^{(k)}) - \xi_{0_i}(\widehat{\boldsymbol{\theta}}^{(k)})\mathbf{n}_i^\top \widehat{\mathbf{f}}^{(k)}\right] \\
&= \left(\mathbf{X}^\top \boldsymbol{\Omega}^{(k)}\mathbf{X}\right)^{-1} \mathbf{X}^\top \left(\boldsymbol{\xi}_1^{(k)} - \boldsymbol{\Omega}^{(k)}\mathbf{N}^\top\widehat{\mathbf{f}}^{(k)}\right), \\
\widehat{\mathbf{f}}^{(k+1)} &= \left[\sum_{i=1}^{n} \xi_{0_i}(\widehat{\boldsymbol{\theta}}^{(k)})\mathbf{n}_i\mathbf{n}_i^\top + \widehat{\alpha}^{(k)}\widehat{\sigma^2}^{(k)}\mathbf{K}\right]^{-1} \\
&\quad \times \sum_{i=1}^{n} \mathbf{n}_i \left[\xi_{1_i}(\widehat{\boldsymbol{\theta}}^{(k)}) - \xi_{0_i}(\widehat{\boldsymbol{\theta}}^{(k)})\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}^{(k+1)}\right] \\
&= \left(\mathbf{N}^\top \boldsymbol{\Omega}^{(k)}\mathbf{N} + \widehat{\alpha}^{(k)}\widehat{\sigma^2}^{(k)}\mathbf{K}\right)^{-1} \\
&\quad \times \mathbf{N}^\top \left(\boldsymbol{\xi}_1^{(k)} - \boldsymbol{\Omega}^{(k)}\mathbf{X}\widehat{\boldsymbol{\beta}}^{(k+1)}\right), \\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\Big[\xi_{2_i}(\widehat{\boldsymbol{\theta}}^{(k)}) - 2\xi_{1_i}(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\mu}_i^{(k+1)} \\
&\quad + \xi_{0_i}(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\mu}_i^{(k+1)^2}\Big] \\
&= \frac{1}{n}\left(\mathbf{1}_n^\top\boldsymbol{\xi}_2^{(k)} - 2\widehat{\boldsymbol{\mu}}^{(k+1)\top}\boldsymbol{\xi}_1^{(k)} + \widehat{\boldsymbol{\mu}}^{(k+1)\top}\boldsymbol{\Omega}^{(k)}\widehat{\boldsymbol{\mu}}^{(k+1)}\right),
\end{aligned}
$$

3. ***CML-step***: Update $\boldsymbol{\nu}^{(k)}$ by maximizing the actual marginal log-likelihood function, obtaining

(15)

$$
\begin{aligned}
\boldsymbol{\nu}^{(k+1)} = \arg\max_{\boldsymbol{\nu}} &\left\{\sum_{i=1}^{m} \log\left[F_{SMN}\left(\frac{\kappa_i - \widehat{\mu}_i^{(k+1)}}{\widehat{\sigma}^{(k+1)}}\right)\right]\right. \\
&\left. + \sum_{i=m+1}^{n} \log\left[f_{SMN}(y_i|\widehat{\mu}_i^{(k+1)}, \widehat{\sigma^2}^{(k+1)}, \boldsymbol{\nu})\right]\right\}.
\end{aligned}
$$

The vector of parameters $\boldsymbol{\nu}$ is just a scalar (degrees of freedom) for the Student-t and slash cases, while $\boldsymbol{\nu} = (\varphi, \gamma)^\top$ for the contaminated normal case. A more efficient *CML-step* (15) can be easily accomplished by using, for instance, the `optimize` or `optimx` routines in the R software [29]. The algorithm iterates between the *E*- and *CML-steps* until reaching convergence, i.e., until some distance involving two successive evaluations of the actual log-likelihood, like $||\ell(\boldsymbol{\theta}^{(k+1)}) - \ell(\boldsymbol{\theta}^{(k)})||$ or $||\ell(\boldsymbol{\theta}^{(k+1)})/\ell(\boldsymbol{\theta}^{(k)}) - 1||$, is small enough. A set of reasonable starting values can be obtained by computing $\widehat{\boldsymbol{\beta}}^{(0)}$ and $\widehat{\sigma^2}^{(0)}$ as the solution of the least squares regression model of $\mathbf{Y}$ on $\mathbf{X}$, considering the censoring values as observed and $\widehat{\mathbf{f}}^{(0)} = (\mathbf{N}^\top\mathbf{N} + \alpha\widehat{\sigma^2}^{(0)}\mathbf{K})^{-1}\mathbf{N}^\top(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(0)})$.

3.1 Estimation of smoothing parameter α

In additive models, the Akaike information criterion (AIC) can be applied to select an appropriate α. Following [7], the AIC for PLR models is defined by:

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

where p is the dimension of the regression parameters $\boldsymbol{\beta}$, q the number of parameters of the SMN distribution being considered and $\ell_{cp}(\widehat{\boldsymbol{\theta}}, \alpha)$ is evaluated at $\widehat{\boldsymbol{\theta}}$ for a fixed α. The degrees of freedom (df) is defined as the number of effective parameters involved in modeling the nonparametric effects and can be approximated by [14]:

$$
\text{df}(\alpha) = tr\{\mathbf{I}_r + \alpha\mathcal{L}\},
$$

where $\mathcal{L} = \widehat{\sigma^2}\mathbf{B}^{-1/2}\mathbf{K}\mathbf{B}^{-1/2}$, with $\mathbf{B} = \mathbf{N}^\top\mathbf{N}$.

3.2 Standard error approximation

Analogously to the parametric case, the approximate variance-covariance matrix of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \mathbf{f}^\top, \sigma^2)^\top$ is derived from the inverse of the observed information matrix [22]. In effect, $\widehat{Var}(\widehat{\boldsymbol{\theta}}) = \mathrm{I}_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}|y)|_{\widehat{\boldsymbol{\theta}}}$, where $\mathrm{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}|y) = -\sum_{i=1}^{n}\frac{\partial^2\ell_{cp_i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$ and $\ell_{cp_i}(\boldsymbol{\theta})$ is the penalized log-likelihood function of the SMN-PCR model, given by:

$$
\begin{aligned}
\ell_{cp}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \ell_{cp_i}(\boldsymbol{\theta}) \\
&= \sum_{i=1}^{m} \log\left[\Psi_i(\boldsymbol{\theta})\right] \\
&\quad + \sum_{i=m+1}^{n}\left\{-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 + \log[\psi_i(\boldsymbol{\theta})]\right\} \\
&\quad - \frac{\alpha}{2}\mathbf{f}^\top\mathbf{K}\mathbf{f},
\end{aligned}
$$

with $\Psi_i(\boldsymbol{\theta}) = \int_0^\infty \Phi\left[k^{-1/2}(u_i)D_i\right]dH(u_i|\boldsymbol{\nu})$ and $\psi_i(\boldsymbol{\theta}) = \int_0^\infty k^{-1/2}(u_i)\exp\left[\frac{-k^{-1}(u_i)d_i}{2}\right]dH(u_i|\boldsymbol{\nu})$, where $d_i = \frac{(y_i-\mu_i)^2}{\sigma^2}$

and $D_i = \sqrt{d_i}$. Thus, the matrix of second derivatives $\mathrm{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}|y)$ can be represented as:

$$\mathrm{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}|y) = -\sum_{i=1}^{n} \frac{\partial^2 \ell_{cp_i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \mathrm{I}^1(\boldsymbol{\theta}) + \mathrm{I}^2(\boldsymbol{\theta}) + \mathrm{I}^3(\boldsymbol{\theta}),$$

where

$$\begin{aligned}
\mathrm{I}^1(\boldsymbol{\theta}) &= -\sum_{i=1}^{m} \left\{ \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \log[\Psi_i(\boldsymbol{\theta})] \right\} \\
&= \sum_{i=1}^{m} \left[\frac{1}{\Psi_i^2(\boldsymbol{\theta})} \frac{\partial\Psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \frac{\partial\Psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} - \frac{1}{\Psi_i(\boldsymbol{\theta})} \frac{\partial^2\Psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \right], \\
\mathrm{I}^2(\boldsymbol{\theta}) &= -\sum_{i=m+1}^{n} \left\{ \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) \right.\right. \\
&\qquad \left.\left. - \frac{\alpha}{2(n-m)} \mathbf{f}^\top \mathbf{K} \mathbf{f} \right] \right\}, \\
\mathrm{I}^3(\boldsymbol{\theta}) &= -\sum_{i=m+1}^{n} \left\{ \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \log[\psi_i(\boldsymbol{\theta})] \right\} \\
&= \sum_{i=1}^{m} \left[\frac{1}{\psi_i^2(\boldsymbol{\theta})} \frac{\partial\psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \frac{\partial\psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} - \frac{1}{\psi_i(\boldsymbol{\theta})} \frac{\partial^2\psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \right].
\end{aligned}$$

The calculation of $\psi_i(\boldsymbol{\theta})$ and $\Psi_i(\boldsymbol{\theta})$ involves, respectively, the pdf and cdf of the normal, Student-t, slash and contaminated normal distributions, thus

$$\frac{\partial\Psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \mathbb{I}_i^{\Phi}(1/2)\frac{\partial D_i}{\partial\boldsymbol{\theta}}, \tag{16}$$

$$\frac{\partial^2\Psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = -\frac{1}{2}\mathbb{I}_i^{\Phi}(3/2)\frac{\partial D_i^2}{\partial\boldsymbol{\theta}}\frac{\partial D_i}{\partial\boldsymbol{\theta}} + \mathbb{I}_i^{\Phi}(1/2)\frac{\partial^2 D_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top},$$

$$\frac{\partial\psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -\frac{1}{2}\mathbb{I}_i^{\phi}(3/2)\frac{\partial d_i}{\partial\boldsymbol{\theta}}, \tag{17}$$

$$\frac{\partial^2\psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \frac{1}{4}\mathbb{I}_i^{\phi}(5/2)\frac{\partial d_i}{\partial\boldsymbol{\theta}}\frac{\partial d_i}{\partial\boldsymbol{\theta}^\top} - \frac{1}{2}\mathbb{I}_i^{\phi}(3/2)\frac{\partial^2 d_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}.$$

See Appendix A for the calculation of derivatives. Using the same notation as in Lachos et al. [17], we have that:

$$\mathbb{I}_i^{\phi}(\omega) = \int_0^\infty k^{-\omega}(u_i)\exp\left[\frac{-k^{-1}(u_i)d_i}{2}\right] dH(u_i|\boldsymbol{\nu}). \tag{18}$$

Since $\mathbb{I}_i^{\Phi}(\omega) = \frac{1}{\sqrt{2\pi}}\mathbb{I}_i^{\phi}(\omega)$, for each distribution considered, the integral defined in (18) can be written as:

- Student-t distribution

$$\mathbb{I}_i^{\phi}(\omega) = \frac{\nu^{\nu/2}2^{\omega}\Gamma(\omega + \frac{\nu}{2})}{\Gamma\left(\frac{\nu}{2}\right)(\nu + d_i)^{\omega+\nu/2}};$$

- Slash distribution

$$\mathbb{I}_i^{\phi}(\omega) = \nu\int_0^1 u_i^{\omega+\nu-1}\exp\left(-\frac{u_i}{2}d_i\right) du_i;$$

- Contaminated normal distribution

$$\begin{aligned}
\mathbb{I}_i^{\phi}(\omega) = \sqrt{2\pi}\Big[&\varphi\gamma^{\omega-1/2}\phi(\sqrt{d_i};0,1/\gamma) \\
&+ (1-\varphi)\phi(\sqrt{d_i};0,1)\Big].
\end{aligned}$$

4. DIAGNOSTIC ANALYSIS

After the estimation procedure, the next step is evaluation of model results to detect outlying and influential observations in addition to possible deviations in the model, because in some cases the character of the regression can be determined only by a few observations. For example, in the context of simple linear regression, it is well known that inferences based on ordinary least squares regression can be strongly influenced by only a few outlying observations in the data. In these circumstances, [6] stated there are two alternatives to handle this situation. The first consists of the development of robust estimation methods that require few assumptions and the second is related to the development of diagnostic tools to detect possible influential observations. So, we have the case-deletion approach [4], a traditional method for identifying influential observations, and the local influence approach, with the aim of investigating the behavior of some influence measures when we introduce small perturbations in the data and then monitor their impact on the outcome of the analysis.

4.1 Case deletion

Case-deletion is a widely used approach that studies the effect on the final inferential results of dropping the i-th case from the dataset. Hereafter, any subscript $[-i]$ refers to the original dataset with the i-th case deleted. In general, we consider $y_{[-i]} = (y_1, y_2, .., y_{n-1})^\top$ as the complete dataset with the i-th observation deleted. The complete-data penalized log-likelihood calculated after eliminating the i-th observation is denoted by $\ell_{cp}(\boldsymbol{\theta}|\boldsymbol{z}_{[-i]})$, therefore let $\widehat{\boldsymbol{\theta}}_{[-i]} = (\widehat{\boldsymbol{\beta}}_{[-i]}^\top, \widehat{\mathbf{f}}_{[-i]}^\top, \widehat{\sigma^2}_{[-i]})^\top$ be the argument that maximizes the function $Q_{[-i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_{cp}(\boldsymbol{\theta}|\boldsymbol{z}_{[-i]})|y_{obs[-i]}, \widehat{\boldsymbol{\theta}}]$, where $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\mathbf{f}}^\top, \widehat{\sigma^2})^\top$ are the MPL estimates obtained through ECME algorithm for $\boldsymbol{\theta}$.

To measure the influence of i-th observation in the MPL estimates of $\boldsymbol{\theta}$, we compare the difference between $\widehat{\boldsymbol{\theta}}_{[-i]}$ and $\widehat{\boldsymbol{\theta}}$. If this difference is large, then the i-th case can be considered influential, so it will require special attention. Since $\boldsymbol{\theta}_{[-i]}$ must be performed considering each individual separately for $i = 1, \ldots, n$, the computational effort can be high for large sample sizes. To circumvent this, [33] proposed the following one-step pseudo approximation:

$$\widehat{\boldsymbol{\theta}}^*_{[-i]} = \widehat{\boldsymbol{\theta}} + \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[-i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \tag{19}$$

where

$$\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial_{\boldsymbol{\theta}^\top}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \quad \text{and} \tag{20}$$

$$\dot{Q}_{[-i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial Q_{[-i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \tag{21}$$

are the Hessian matrix and the individual score vector evaluated at $\widehat{\boldsymbol{\theta}}$ respectively. Thus, $\dot{Q}_{[-i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = (\dot{Q}_{[-i]_{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \dot{Q}_{[-i]_{\mathbf{f}}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \dot{Q}_{[-i]_{\sigma^2}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}))^\top$ has the elements as:

$$\begin{aligned}
\dot{Q}_{[-i]_{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial Q_{[-i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\
&= \frac{1}{\widehat{\sigma^2}}\sum_{i\neq j}\left[\xi_{1_j}(\widehat{\boldsymbol{\theta}})\mathbf{x}_j - \xi_{0_j}(\widehat{\boldsymbol{\theta}})\mathbf{x}_j\widehat{\mu}_j\right], \\
\dot{Q}_{[-i]_{\mathbf{f}}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial Q_{[-i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\
&= \frac{1}{\widehat{\sigma^2}}\sum_{i\neq j}\left[\xi_{1_j}(\widehat{\boldsymbol{\theta}})\mathbf{n}_j - \xi_{0_j}(\widehat{\boldsymbol{\theta}})\mathbf{n}_j\widehat{\mu}_j\right] - \frac{\widehat{\alpha}}{n}\mathbf{K}\mathbf{f}, \\
\dot{Q}_{[-i]_{\sigma^2}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial Q_{[-i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\
&= -\frac{1}{2\widehat{\sigma^2}}\sum_{i\neq j}\Big\{1 - \frac{1}{\widehat{\sigma^2}}\Big[\xi_{2_j}(\widehat{\boldsymbol{\theta}}) - 2\xi_{1_j}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_j \\
&\quad + \xi_{0_j}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_j^2\Big]\Big\}.
\end{aligned}$$

Following [33], to measure the distance between $\widehat{\boldsymbol{\theta}}_{[-i]}$ and $\widehat{\boldsymbol{\theta}}$ and therefore to assess influential observations, we compute the *generalized Cook's distance* as follows:

$$GD_i = \left(\widehat{\boldsymbol{\theta}}_{[-i]} - \widehat{\boldsymbol{\theta}}\right)^\top\left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}\left(\widehat{\boldsymbol{\theta}}_{[-i]} - \widehat{\boldsymbol{\theta}}\right), \quad i = 1, \ldots, n, \tag{22}$$

and by substituting Equation (22) into (19), we obtain the approximation of the *generalized Cook's distance*

$$GD_i^1 = \dot{Q}_{[-i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top\left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1}\dot{Q}_{[-i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \quad i = 1, \ldots, n. \tag{23}$$

The Hessian matrix, $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$

After some rearrangement of terms and evaluation of the derivatives at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, we obtain the Hessian matrix $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ with elements given by:

$$\begin{aligned}
\ddot{Q}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial_{\boldsymbol{\beta}^\top}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\frac{1}{\widehat{\sigma^2}}\sum_{i=1}^{n}\xi_{0_i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i\mathbf{x}_i^\top, \\
\ddot{Q}_{\mathbf{f}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}\partial\mathbf{f}^\top}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\frac{1}{\widehat{\sigma^2}}\sum_{i=1}^{n}\xi_{0_i}(\widehat{\boldsymbol{\theta}})\mathbf{n}_i\mathbf{n}_i^\top - \widehat{\alpha}\mathbf{K}, \\
\ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\sigma^2}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\
&= -\frac{n}{2\widehat{\sigma^2}} + \frac{1}{(\widehat{\sigma^2})^3}\sum_{i=1}^{n}\Big[\xi_{2_i}(\widehat{\boldsymbol{\theta}}) - 2\xi_{1_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i + \xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i^2\Big], \\
\ddot{Q}_{\boldsymbol{\beta}\mathbf{f}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\mathbf{f}^\top}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\frac{1}{\widehat{\sigma^2}}\sum_{i=1}^{n}\xi_{0_i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i\mathbf{n}_i^\top, \\
\ddot{Q}_{\boldsymbol{\beta}\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma^2}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\
&= -\frac{1}{(\widehat{\sigma^2})^2}\sum_{i=1}^{n}\Big[\xi_{1_i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i - \xi_{0_i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i\widehat{\mu}_i\Big], \\
\ddot{Q}_{\mathbf{f}\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}\partial\sigma^2}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\
&= -\frac{1}{(\widehat{\sigma^2})^2}\sum_{i=1}^{n}\Big[\xi_{1_i}(\widehat{\boldsymbol{\theta}})\mathbf{n}_i - \xi_{0_i}(\widehat{\boldsymbol{\theta}})\mathbf{n}_i\widehat{\mu}_i\Big].
\end{aligned}$$

4.2 Local influence

Local influence analysis seeks to verify if small perturbations in the model or in the data affect the parameter estimates. Hence, to study the behavior of some influence measures, we will follow the approach proposed by [33], where the Q-function is perturbed to assess the influence of this perturbation on the estimation. [15] and [7] applied this method successfully in the context of PLR models.

Consider a perturbation vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ restricted to some open subset $\boldsymbol{\Omega} \in \mathcal{R}^n$. Let $\ell_{cp}(\boldsymbol{\theta}, \boldsymbol{\omega}|\boldsymbol{z})$ be the complete-data penalized log-likelihood function of the perturbed model. Thus, we assume that a $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ exists such that $\ell_{cp}(\boldsymbol{\theta}, \boldsymbol{\omega}|\boldsymbol{z}) = \ell_{cp}(\boldsymbol{\theta}|\boldsymbol{z})$ for all $\boldsymbol{\theta}$. Also, let $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}) = (\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})^\top, \widehat{\mathbf{f}}(\boldsymbol{\omega})^\top, \widehat{\sigma^2}(\boldsymbol{\omega}))^\top$ denote the maximum of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_{cp}(\boldsymbol{\theta}, \boldsymbol{\omega}|\boldsymbol{z})|\mathbf{Y}_{obs}, \widehat{\boldsymbol{\theta}}]$. Then, the influence graph is defined as $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$, where $f_Q(\boldsymbol{\omega})$ is the Q-displacement function, defined as:

$$f_Q(\boldsymbol{\omega}) = 2\Big[Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}})\Big].$$

To approximate the Q-displacement function, the normal curvature $C_{f_Q,\boldsymbol{h}}(\boldsymbol{\theta})$ of $\boldsymbol{\alpha}(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of a unit vector $\boldsymbol{h}$ ($||\ \boldsymbol{h}\ || = 1$) is used to summarize the local behavior of $f_Q(\boldsymbol{\omega})$. It can be shown that:

$$C_{f_Q,\boldsymbol{h}}(\boldsymbol{\theta}) = -2\boldsymbol{h}^\top\ddot{Q}_{\boldsymbol{\omega}_0}\boldsymbol{h} \ = \ 2\boldsymbol{h}^\top\Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}^\top\{\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}\boldsymbol{h},$$

leading to

$$-\ddot{Q}_{\boldsymbol{\omega}_0} = \Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}^\top\{\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}.$$

Additionally, $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is the Hessian matrix of dimension $(p+r+1)\times(p+r+1)$ and $\Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0} = \partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top = (\Delta_{\boldsymbol{\beta},\boldsymbol{\omega}_0}^\top, \Delta_{\mathbf{f},\boldsymbol{\omega}_0}^\top, \Delta_{\sigma^2,\boldsymbol{\omega}_0}^\top)^\top$ is the matrix of dimension $(p + r + 1) \times n$ evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. The information provided by $-\ddot{Q}_{\boldsymbol{\omega}_0}$ is fundamental for detecting influential observations [5]. From the spectral decomposition of a symmetric matrix

$$-2\ddot{Q}_{\boldsymbol{\omega}_0} = \sum_{k=1}^{n} \lambda_k \mathbf{v}_k \mathbf{v}_k^\top,$$

where $(\lambda_1, \mathbf{v}_1), \cdots, (\lambda_n, \mathbf{v}_n)$ are the eigenvalue-eigenvector pairs of $-2\ddot{Q}_{\boldsymbol{\omega}_0}$ with $\lambda_1 \geq \ldots \geq \lambda_r$, $\lambda_{r+1} = \ldots = \lambda_n = 0$ and orthonormal eigenvectors $\{\mathbf{v}_k, k = 1, .., n\}$. [33] and [19] proposed to examine all eigenvectors corresponding to nonzero eigenvalues to capture more information. For this end, we have the expressions

$$\tilde{\lambda}_k = \frac{\lambda_k}{\lambda_1 + \ldots + \lambda_r}, \quad \mathbf{v}_k^2 = (v_{k1}^2, \ldots, v_{kn}^2)^\top \quad \text{and}$$

$$M(0) = \sum_{k=1}^{r} \tilde{\lambda}_k \mathbf{v}_k^2.$$

Let $M(0)_l = \sum_{k=1}^r \tilde{\lambda}_k v_{kl}^2$ denote the l-th component of $M(0)$. The evaluation of influential observations is based on the visual inspection of $M(0)_l$ plotted against the l-th index, for $l = 1, \ldots, n$. There are some disadvantages in using the normal curvature for influence analysis, since this measure may assume any value (not bounded), meaning it is not invariant under uniform scaling changes. Instead, we use the conformal normal curvature [28], given by:

(24)

$$B_{f_Q,\boldsymbol{h}}(\boldsymbol{\theta}) = \frac{C_{f_Q,\boldsymbol{h}}(\boldsymbol{\theta})}{\operatorname{tr}\left[-2\ddot{Q}_{\boldsymbol{\omega}_0}\right]}$$

$$\Rightarrow \quad B_{f_Q,\boldsymbol{h}_l}(\boldsymbol{\theta}) = \frac{\Delta_{l_{\boldsymbol{\theta},\boldsymbol{\omega}_0}}^\top \{\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \Delta_{l_{\boldsymbol{\theta},\boldsymbol{\omega}_0}}}{\operatorname{tr}\left[\Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}^\top \{\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}\right]}, \quad l = 1, \ldots, n,$$

where $\boldsymbol{h}_l$ is a column vector in $\mathcal{R}^n$ with the l-th entry equal to one, and zeros in the remaining positions. Here, $\Delta_{l_{\boldsymbol{\theta},\boldsymbol{\omega}_0}}^\top$ corresponds to $\boldsymbol{h}_l^\top \Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}^\top$. The conformal normal curvature defined in (24) has the property that $0 \leq B_{f_Q,\boldsymbol{h}_l}(\boldsymbol{\theta}) \leq 1$ and the calculation is computationally easier. Based on the work of [33], $M(0)_l$ can be obtained via $B_{f_Q,\boldsymbol{h}_l}(\boldsymbol{\theta})$ for all l.

Currently, there is no general rule for determining a benchmark value to indicate whether an observation is influential or not. Let $\overline{M(0)}$ and $SM(0)$ be the mean and standard error of $\overline{\{M(0)_l, l = 1, \ldots, n\}}$ respectively. [33] showed that $\overline{M(0)} = 1/n$ and proposed $\overline{M(0)} + 2SM(0)$ as benchmark value for $M(0)_l$. On the other hand, [19] presented a generalization and also proposed to use $M(0)_l > \overline{M(0)} + c^* SM(0)$, with c^* being a selected constant greater than 2. The choice of c^* depends on the context of the application [7], where the l-th case is considered as influential if $M(0)_l$ is larger than the benchmark.

4.2.1 Perturbation schemes

We now evaluate the matrix $\Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ under four different perturbation schemes for the SMN-PCR model.

(a) Case-weight perturbation

This case of perturbation is appropriate to detect observations with large contribution to the penalized log-likelihood function and that may exercise strong influence on the maximum penalized likelihood estimates. From Equation (12), the so-called perturbed Q-function, considering an arbitrary attribution of weights, we have:

(25)
$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) &= \sum_{i=1}^{n} \omega_i \mathrm{E}[\ell_{cp_i}(\boldsymbol{\theta}|\boldsymbol{z})|\boldsymbol{Y}_{obs}, \widehat{\boldsymbol{\theta}}] \\ &= \sum_{i=1}^{n} \omega_i Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) - \frac{\widehat{\alpha}}{2} \widehat{\mathbf{f}}^\top \mathbf{K} \widehat{\mathbf{f}}, \end{aligned}$$

where $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n)^\top \in \mathcal{R}^n$. The original expected value of the penalized complete-data log-likelihood corresponds to $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. In this perturbation scheme, the matrix $\Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ derived from (25) has elements given by:

$$\begin{aligned} \Delta_{\boldsymbol{\beta},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \frac{\mathbf{x}_i}{\widehat{\sigma^2}} \left[\xi_{1_i}(\widehat{\boldsymbol{\theta}}) - \xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i\right], \\ \Delta_{\mathbf{f},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \mathbf{f} \partial \omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \frac{\mathbf{n}_i}{\widehat{\sigma^2}} \left[\xi_{1_i}(\widehat{\boldsymbol{\theta}}) - \xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i\right] \quad \text{and} \\ \Delta_{\sigma^2,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \\ &= -\frac{1}{2\widehat{\sigma^2}} + \frac{1}{2(\widehat{\sigma^2})^2} \left[\xi_{2_i}(\widehat{\boldsymbol{\theta}}) - 2\xi_{1_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i + \xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i^2\right]. \end{aligned}$$

(b) Scale perturbation

In order to study the behavior of the estimates when there are possible deviations from the assumption of homogeneity, we assume that $Y_i \sim \mathrm{SMN}(\mu_i, \sigma^2(\omega_i), \boldsymbol{\nu})$, with $\sigma^2(\omega_i) = \omega_i^{-1}\sigma^2$, $\omega_i > 0$ for $i = 1, \ldots, n$. The perturbed Q-function under this scheme is expressed as:

(26)
$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \Bigg\{ &-\frac{1}{2} \log\left(\frac{\widehat{\sigma^2}}{\omega_i}\right) \\ &- \frac{\omega_i}{2\widehat{\sigma^2}} \left[\xi_{2_i}(\widehat{\boldsymbol{\theta}}) - 2\xi_{1_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i + \xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i^2\right] \Bigg\} \\ &- \frac{\widehat{\alpha}}{2} \widehat{\mathbf{f}}^\top \mathbf{K} \widehat{\mathbf{f}}, \end{aligned}$$

where $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n)^\top \in \mathcal{R}^n$ is the vector of perturbations such that the non-perturbed model is obtained when $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. After some algebraic manipulations in (26), we obtain the elements of the matrix $\Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ as:

$$\begin{aligned} \Delta_{\boldsymbol{\beta},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \frac{\mathbf{x}_i}{\widehat{\sigma^2}} \left[\xi_{1_i}(\widehat{\boldsymbol{\theta}}) - \xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i\right], \\ \Delta_{\mathbf{f},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \mathbf{f} \partial \omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \frac{\mathbf{n}_i}{\widehat{\sigma^2}} \left[\xi_{1_i}(\widehat{\boldsymbol{\theta}}) - \xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i\right] \quad \text{and} \\ \Delta_{\sigma^2,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \end{aligned}$$

$$= \frac{1}{2(\widehat{\sigma^2})^2}\left[\xi_{2_i}(\widehat{\boldsymbol{\theta}}) - 2\xi_{1_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i + \xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i^2\right].$$

(c) Explanatory variable perturbation

Here, we are interested in perturbing a specific continuous explanatory variable. Let $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n)^\top$ be the vector of perturbations, $i = 1, \ldots, n$. So, the r-th explanatory variable of the design matrix is perturbed as $\mathbf{x}_{i_\omega}^\top = \mathbf{x}_i^\top + \omega_i \mathcal{S}_r \mathbf{e}_r^\top$ for $r = 1, \ldots, p$, where $\mathcal{S}_r$ is the standard deviation of the r-th explanatory variable and $\mathbf{e}_r$ is a vector of dimension $p \times 1$, with a one in the r-th position and zero elsewhere. Under this scheme, we have that the perturbed Q-function is

(27)

$$\begin{aligned} &Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) \\ &\quad = -\frac{n}{2}\log\widehat{\sigma^2} - \frac{1}{2\widehat{\sigma^2}}\sum_{i=1}^{n}\left[\xi_{2_i}(\widehat{\boldsymbol{\theta}}) - 2\xi_{1_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i^* + \xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i^{2*}\right] \\ &\qquad - \frac{\widehat{\alpha}}{2}\widehat{\mathbf{f}}^\top \mathbf{K}\widehat{\mathbf{f}}, \end{aligned}$$

where $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n)^\top \in \mathcal{R}^n$ and $\widehat{\mu}_i^* = \mathbf{x}_{i_\omega}^\top \widehat{\boldsymbol{\beta}} + \mathbf{n}_i^\top \widehat{\mathbf{f}}$. The vector of non-perturbations is $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top \in \mathcal{R}^n$. Taking the second derivative of (27) we find

$$\begin{aligned} \Delta_{\boldsymbol{\beta},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \\ &= \frac{\mathcal{S}_r}{\widehat{\sigma^2}}\left\{\xi_{1_i}(\widehat{\boldsymbol{\theta}})\mathbf{e}_r - \xi_{0_i}(\widehat{\boldsymbol{\theta}})\left[\widehat{\mu}_i\mathbf{e}_r + \mathbf{e}_r^\top\widehat{\boldsymbol{\beta}}\mathbf{x}_i\right]\right\}, \\ \Delta_{\mathbf{f},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}\partial\omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\frac{\mathcal{S}_r}{\widehat{\sigma^2}}\xi_{0_i}(\widehat{\boldsymbol{\theta}})\mathbf{e}_r^\top\widehat{\boldsymbol{\beta}}\ \mathbf{n}_i \quad \text{and} \\ \Delta_{\sigma^2,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \\ &= \frac{\mathcal{S}_r}{(\widehat{\sigma^2})^2}\left[\xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i\mathbf{e}_r^\top\widehat{\boldsymbol{\beta}} - \xi_{1_i}(\widehat{\boldsymbol{\theta}})\mathbf{e}_r^\top\widehat{\boldsymbol{\beta}}\right]. \end{aligned}$$

(d) Response variable perturbation

In this case, to perturb the response variable values we replace Y_{obs_i} by $Y_{\text{obs}_i}(\omega_i) = Y_{\text{obs}_i} + \mathcal{S}_y\omega_i$ for $i = 1, \ldots, n$, where $\mathcal{S}_y$ is the standard deviation of Y_{obs_i}. For the SMN-PCR model presented in Equations (5) and (6) we have:

$$Y_{\text{obs}_i}(\omega_i) = \begin{cases} \kappa_i(\omega_i) & \text{if} \quad Y_i \le \kappa_i; \\ Y_i(\omega_i) & \text{if} \quad Y_i > \kappa_i. \end{cases}$$

Therefore $Y_i(\omega_i) = Y_i - \mathcal{S}_y\omega_i$ [25]. So, the perturbed Q-function it is obtained replacing Y_{obs_i} values by $Y_{\text{obs}_i}(\omega_i)$, where $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n)^\top \in \mathcal{R}^n$ denotes the vector of perturbations and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$ is the corresponding non-perturbation vector such that $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = Q(\boldsymbol{\theta}, |\widehat{\boldsymbol{\theta}})$, with

(28)

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = -\frac{n}{2}\log(\widehat{\sigma^2}) - \frac{1}{2\widehat{\sigma^2}}\sum_{i=1}^{n}\Big[&\xi_{2_i}(\widehat{\boldsymbol{\theta}}) - 2\xi_{1_i}(\widehat{\boldsymbol{\theta}})\mathcal{S}_y\omega_i \\ &+ \xi_{0_i}(\widehat{\boldsymbol{\theta}})\mathcal{S}_y^2\omega_i^2 - 2\xi_{1_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i + 2\xi_{0_i}(\widehat{\boldsymbol{\theta}})\mathcal{S}_y\omega_i\widehat{\mu}_i \\ &+ \xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i^2\Big] - \frac{\widehat{\alpha}}{2}\widehat{\mathbf{f}}^\top\mathbf{K}\widehat{\mathbf{f}}. \end{aligned}$$

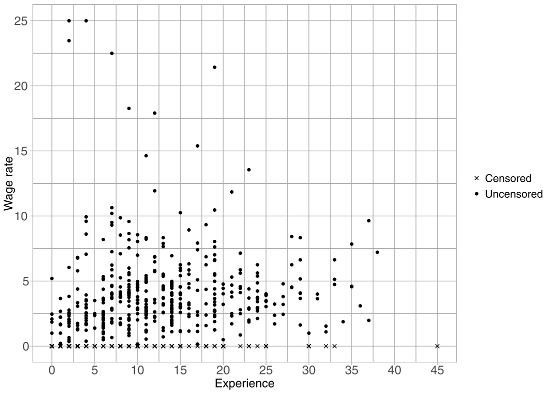

Figure 1. PSID-1975 dataset. Wage rates vs. number of years worked (Experience).

The matrix $\Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ has elements given by:

$$\begin{aligned} \Delta_{\boldsymbol{\beta},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\frac{\mathcal{S}_y}{\widehat{\sigma^2}}\ \xi_{0_i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i, \\ \Delta_{\mathbf{f},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}\partial\omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\frac{\mathcal{S}_y}{\widehat{\sigma^2}}\ \xi_{0_i}(\widehat{\boldsymbol{\theta}})\mathbf{n}_i \quad \text{and} \\ \Delta_{\sigma^2,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\omega_i}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \frac{\mathcal{S}_y}{(\widehat{\sigma^2})^2}\ \left[\xi_{0_i}(\widehat{\boldsymbol{\theta}})\widehat{\mu}_i - \xi_{1_i}(\widehat{\boldsymbol{\theta}})\right]. \end{aligned}$$

5. APPLICATION: WAGE RATE DATA

In this section, we illustrate the performance of the proposed method by analyzing the wage rate dataset described in Mroz [26], previously analyzed by Arellano-Valle et al. [1], Castro et al. [3] and [23].

The dataset comes from the University of Michigan Panel Study of Income Dynamics (PSID) and describes the average hourly earnings or wage rates (the dependent variable used in this application) of 753 married white women between the ages of 30 and 60, with 428 working at some time during the year 1975. For those who did not work in 1975, the wage rate is zero, so the variable can be classified as censored-uncensored, i.e., it follows Equation (13) with $\kappa_i = 0$ for $i = 1, \ldots, 753$. This dataset presents left censored observations, since we can only observe its real value if a woman worked for pay during 1975. Thus, our purpose is to model the wage rate as a function of a set of control

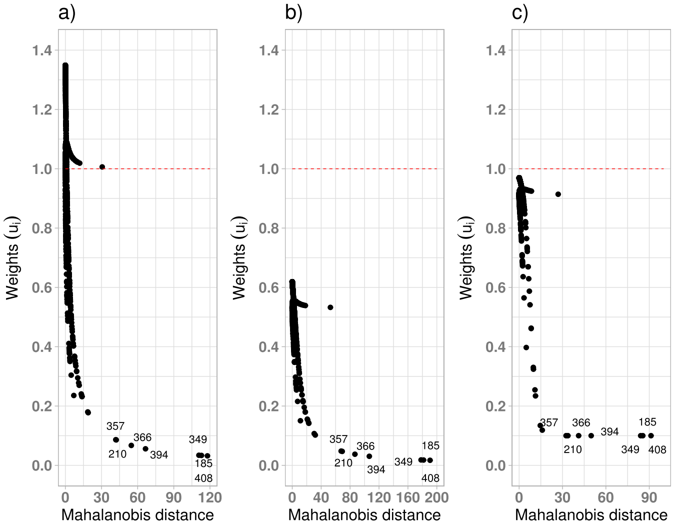

Figure 2. PSID-1975 dataset. Estimated weights u_i vs Mahalanobis distance d_i^2 for: a) T-PCR, b) SL-PCR and c) CN-PCR models, respectively.

variables such as the wife's age (x_{1_i}), her years of schooling (x_{2_i}), husband's hours worked (x_{3_i}), husband's wage in dollars (x_{4_i}), tax rate faced by the wife (x_{5_i}), number of children younger than six years old in the household (x_{6_i}) and number of children between the ages of six and nineteen (x_{7_i}). We also consider a nonlinear relation between the wage rates and the number of years the wife worked since age 18 (see Figure 1).

5.1 Analyses of the fitted models

The data were analyzed using the SMN-PCR models considering the Student-t (T-PCR), slash (SL-PCR), contaminated normal (CN-PCR), and of course the normal distribution (N-PCR) for comparative purposes.

Table 3 contains the MPL estimates of the parameters for the four fitted models, together with their corresponding standard errors calculated via the observed information matrix as presented in Subsection 3.2. Note from this table that the estimated values of ν are small, indicating a heavy-tailed behavior and consequently the lack of adequacy of the N-PCR model for this dataset. Further, we compare the results among the SMN-PCR models using the AIC value defined in Subsection 3.1 and the log-likelihood values $(\ell(\widehat{\boldsymbol{\theta}}))$. As expected, we can see that the SMN distributions with heavy tails have a better performance compared with the normal one, evidencing once again a clear departure from the normality assumption, with the SL-PCR model significantly better. Finally, the values of the respective standard errors (SE) of the heavy tails models are smaller than those under the normal assumption, indicating that the SMN-PCR models produce more precise estimates.

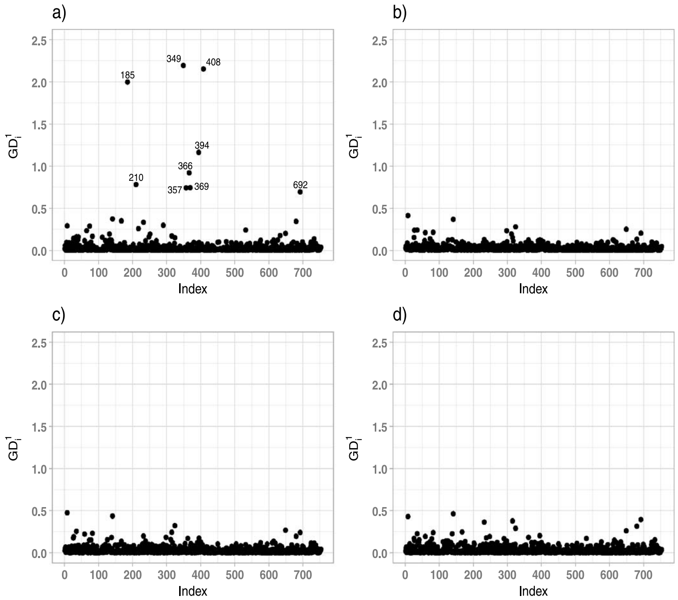

Figure 3. PSID-1975 dataset. Approximate generalized Cook's distance GD_i^1. a) N-PCR, b) T-PCR, c) SL-PCR and d) CN-PCR models respectively.

5.2 Diagnostics analysis

For the purpose of identifying possible observations that can affect the MPL estimates, we use the diagnostic measures presented in Section 4 for the PSID-1975 dataset. From the results of the ECME algorithm, Figure 2 shows the estimated weights $u_i = \xi_{0_i}(\widehat{\boldsymbol{\theta}})$, $i = 1, \ldots, 753$, versus the Mahalanobis distance, which is defined by $d_i^2 = (y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}} - \mathbf{n}_i^\top \widehat{\mathbf{f}})^2 / \widehat{\sigma^2}$. For the normal case, we have that $u_i = 1$, $\forall i$ (segmented red lines). We can observe from this figure that u_i is inversely proportional to d_i^2, i.e., large d_i^2 values imply smaller u_i weights. Hence, using distributions with heavier tails than the normal leads to smaller weights being attributed to possible influential observations (see also Figure 7, Appendix B.1).

To identify influential observations in a global context and following the approach described in Subsection 4.1, the index plots for the generalized Cook's distance GD_i^1 are shown in Figure 3. High values of GD_i^1 suggest that the i-th observation has an impact on the MPL estimates. We can note that, women #185, #210, #349, #357, #366, #369, #394, #408 and #692 are potentially influential in the MPL estimates under the N-PCR (panel a), but for distributions with heavy tails these women are no longer influential (panels b to d). Comparison of Figures 2 and 3 reveals, women who were considered influential in the normal case obtained small weights in the Student-t, slash and contaminated normal cases.

Next, we study the local influence based on $M(0)$ from the conformal normal curvature $B_{f_Q,\boldsymbol{h}_l}(\boldsymbol{\theta})$ defined in Equation (24), considering the four perturbation schemes: case-

Table 3. *PSID-1975 dataset. Parameter estimates and standard errors (SE) for various fits of the SMN-PCR models*

	Model							
	N-PCR		**T-PCR**		**SL-PCR**		**CN-PCR**	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_1	0.7688	(0.0881)	0.6672	(0.0783)	0.6620	(0.0757)	0.6682	(0.0730)
β_2	-0.0634	(0.0269)	-0.0743	(0.0204)	-0.0745	(0.0201)	-0.0746	(0.0207)
β_3	-0.0008	(0.0004)	-0.0004	(0.0003)	-0.0004	(0.0003)	-0.0005	(0.0003)
β_4	-0.1801	(0.0742)	-0.1321	(0.0642)	-0.1403	(0.0622)	-0.1527	(0.0605)
β_5	-8.6192	(3.8726)	-5.7045	(3.3936)	-6.0435	(3.2619)	-6.3471	(3.1808)
β_6	-1.8132	(0.4056)	-1.7666	(0.3136)	-1.7829	(0.3103)	-1.7874	(0.3254)
β_7	0.3257	(0.1456)	0.1986	(0.1136)	0.2038	(0.1098)	0.2145	(0.1126)
σ^2	15.8854	(1.2320)	5.4806	(0.6375)	3.4223	(0.3788)	7.2270	(0.7089)
ν	–	–	2.8655	–	1.1248	–	–	–
φ	–	–	–	–	–	–	0.1	–
γ	–	–	–	–	–	–	0.1	–
$\ell(\widehat{\boldsymbol{\theta}})$	-1378.2	–	-1305.8	–	-1303.3	–	-1303.8	–
AIC	2852.2	–	2709.6	–	2704.5	–	2707.6	–

weight perturbation, scale perturbation, explanatory variable perturbation and response variable perturbation. We compute the matrix $\Delta_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ for each perturbation scheme to analyze the respective local influence measures for $\widehat{\boldsymbol{\theta}}$ obtained from the best fitted model, that is the SL-PCR model. As benchmark, we use the criterion $M(0)_l > \overline{M(0)} + c^*SM(0)$, with $c^* = 4$, to classify i-th observation as potentially influential.

Examining Figure 4, we have that for the three first perturbation schemes (see panels a to f) women #185, #210, #349, #357, #366, #369, #394, #408 and #692 in the N-PCR model are considered influential. In addition, it is noteworthy that observations that were considered influential in the case-weight perturbation were also found influential with scale perturbation. No observation has a significant influence on the MPL estimates under the SL-PCR model, indicating the robustness of the MPL estimates against potentially influential observations. However, for the response variable perturbation (panels g to h), it was found that women #27, #55, #57, #87, #271, #298, #397 and #598 have a moderate influence in both models (normal and slash). We infer that this is because the "*experience*" values in those cases are high, which naturally leads to a moderate effect on the estimates of $M(0)_l$. For comparison, we use normal benchmarks for all local influence graphs. The influence analyses for the remaining fitted models are shown in Figure 8 (Appendix B.1).

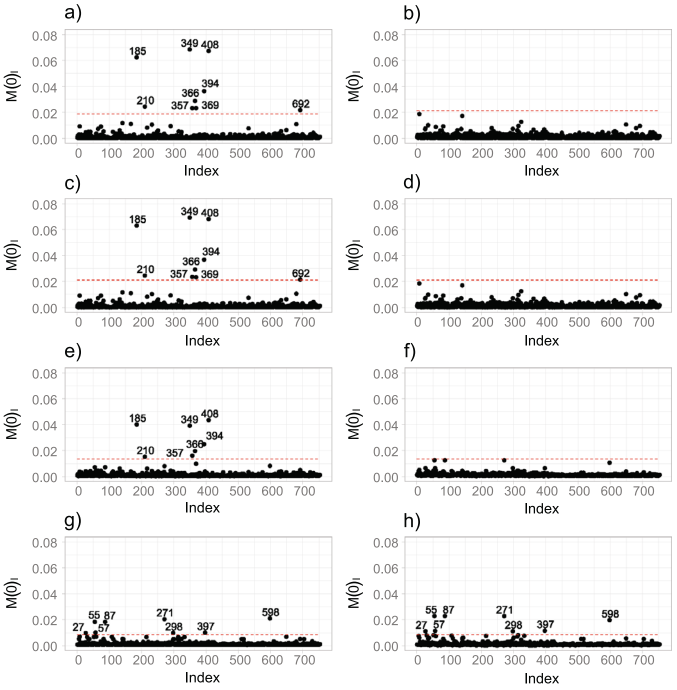

Figure 4. *PSID-1975 dataset. Index plots of $M(0)_l$ for assessing local influence. Different perturbations schemes (case-weight, scale, explanatory variable and response variable perturbation) are shown in the rows from top to bottom. The N-PCR and SL-PCR models correspond to the columns from left to right.*

5.3 Relative change in the MPL estimates

To detect the impact of the observations identified as potentially influential on the MPL estimates, we define the relative change (RC) as

$$RC_j(\widehat{\boldsymbol{\theta}}) = \left|\frac{\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{[-\mathcal{I}_j]}}{\widehat{\boldsymbol{\theta}}}\right| \times 100\%,$$

where $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2})$, and $\widehat{\boldsymbol{\theta}}_{[-\mathcal{I}_j]}$ is the estimate of the parameters after the observations indexed by the set $\mathcal{I}_j$ have been removed. From all possible combinations, we drop influential cases, one by one and from all observations at the same time. Then, the sets of interest are $\mathcal{I}_j = \{e_j\}$, for $j = 1, \ldots, J$ and $\mathcal{I}_{J+1} = E$, where E is the set of all J influential case indexes.

Table 4. PSID-1975 dataset. Relative change (%) of maximum penalized likelihood estimates of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma^2}$ in N-PCR and SL-PCR models

Model	Dropped	$RC_{\widehat{\beta}_1}$	$RC_{\widehat{\beta}_2}$	$RC_{\widehat{\beta}_3}$	$RC_{\widehat{\beta}_4}$	$RC_{\widehat{\beta}_5}$	$RC_{\widehat{\beta}_6}$	$RC_{\widehat{\beta}_7}$	$RC_{\widehat{\sigma^2}}$
N-PCR	185	1.908	7.551	11.70	7.446	2.109	6.693	4.575	8.718
	210	0.855	2.244	3.378	0.163	0.216	3.675	6.568	3.235
	349	1.218	20.28	6.792	2.789	7.210	0.485	12.48	8.291
	357	2.260	9.032	4.944	4.104	0.002	3.061	6.400	3.677
	366	0.030	14.11	5.085	11.98	4.974	0.348	4.735	4.481
	369	0.425	8.584	5.543	13.18	2.333	3.420	3.465	1.614
	394	3.525	8.433	3.675	2.474	0.257	2.077	2.860	5.467
	408	0.335	3.442	2.979	1.401	5.536	7.793	13.77	9.524
	692	0.709	0.713	1.587	1.687	0.390	0.231	0.921	0.576
	E	11.24	1.558	19.32	6.239	14.59	10.10	10.33	46.95
SL-PCR	185	0.015	0.182	7.375	1.812	0.923	0.280	0.537	4.376
	210	0.138	0.179	6.497	0.733	0.421	0.564	1.806	2.526
	349	0.047	1.004	7.762	1.114	0.277	0.026	0.163	4.353
	357	0.396	1.195	8.373	1.625	0.503	0.458	2.160	2.235
	366	0.137	1.293	8.811	0.662	0.130	0.062	1.466	2.854
	369	0.262	2.375	9.181	5.238	1.036	1.063	2.470	0.883
	394	0.355	0.678	8.331	0.735	0.566	0.309	1.166	3.046
	408	0.073	0.257	8.894	1.216	0.412	0.468	0.205	4.542
	692	0.358	0.213	5.980	1.120	0.103	0.044	0.534	0.355
	E	1.445	8.358	42.01	25.55	12.68	5.076	24.97	10.29

For our example, Table 4 presents the relative change of the MPL estimates after removing the observations indexed by $\mathcal{I}_j$ and refitting the N-PCR and SL-PCR models, respectively. Note that the biggest changes occur in the N-PCR model, particularly for the parameter σ^2. As expected, the results indicate smaller changes in the MPL estimates under SL-PCR model, confirming the robust aspects when distributions with heavier tails than the normal one are used. Although some relative changes are significant, particularly with the normal distribution, most of times the SL-PCR model presented much smaller RCs. In addition, the RC for the set of observations identified as influential under the response variable perturbation schema were small and quite similar in both models.

6. SIMULATION STUDY

This section reports the results of a Monte Carlo (MC) experiment designed to evaluate the performance of the proposed model to analyze the behavior of MPL estimates, the approximated standard errors, as well as its asymptotic properties. These computational procedures were implemented using the `R` software [29]. In particular, we consider the following PLR model:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + f(t_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{29}$$

where $\varepsilon_i \overset{\text{iid}}{\sim} \text{SMN}(0, \sigma^2, \boldsymbol{\nu})$. We generated left-censored samples from the model given in (29) considering censoring levels 0%, 10%, 20% and 30% and sample sizes $n = 200, 300, 400$ and 600. For each combination of censoring level and sample size, we generated 500 samples from the SMN-PCR model, in four different situations: N-PCR, T-PCR ($\nu = 4$), SL-PCR ($\nu = 2$) and CN-PCR ($\boldsymbol{\nu}^\top = (0.1, 0.1)$).

We performed all MC simulations setting $\boldsymbol{\beta} = (2, 4)^\top$ and $\sigma^2 = 2$, with $\mathbf{x}_i = (x_{1_i}, x_{2_i})^\top$ generated independently from uniform distributions on the intervals $U(0,1)$ and $U(1,2)$, respectively. The true nonparametric function was chosen as $f(t) = 10\sin(2\pi t)$, with $t \in (0, 1.5)$. We also assumed different values for t, so the incidence matrix $\mathbf{N}$ was the identity matrix of order $n \times n$. We computed for each parameter θ_k, its MC mean $(\overline{\overline{\widehat{\theta}}}_k)$ and MC standard deviation (MC-SD) for its MPL estimates. Also, the average values of the approximate standard errors obtained following subsection 3.2 (OM-SD), were recorded for comparison purposes. So, we have that:

$$\overline{\overline{\widehat{\theta}}}_k = \frac{1}{500}\sum_{j=1}^{500} \widehat{\theta}_k^{(j)}, \ \text{MC-SD} = \sqrt{\frac{\sum_{j=1}^{500}(\widehat{\theta}_k^{(j)} - \overline{\overline{\widehat{\theta}}}_k)^2}{499}},$$

$$\text{OM-SD} = \frac{1}{500}\sum_{j=1}^{500} \text{SE}(\widehat{\theta}_k^{(j)}).$$

Table 5 presents the MPL estimates of θ_k in the different scenarios considered. This table shows that the model fits provide estimates that are close to the true values of the parameters and are less sensitive to the variation of the censoring level. Besides this, the empirical standard deviations (MC-SD) are close to the MC standard errors (OM-SD) and the difference tends to zero as the sample size increases, indicating that the result of the standard errors (Subsection 3.2) is reliable.

Table 5. Simulated data. MC Mean value, MC standard deviation (MC-SD) and mean approximated standard errors

		N-PCR			**T-PCR**			**SL-PCR**			**CN-PCR**		
Parameter	C.L.	$\widehat{\theta}_k$	MC-SD	OM-SD	$\widehat{\theta}_k$	MC-SD	OM-SD	$\widehat{\theta}_k$	MC-SD	OM-SD	$\widehat{\theta}_k$	MC-SD	OM-SD
							$n = 200$						
β_1	0%	1.986	0.373	0.339	2.060	0.448	0.405	2.008	0.459	0.449	2.045	0.421	0.381
	10%	1.983	0.384	0.347	2.070	0.446	0.412	2.010	0.459	0.456	2.035	0.434	0.390
	20%	1.990	0.410	0.367	2.080	0.487	0.438	2.013	0.490	0.483	2.055	0.470	0.414
	30%	2.008	0.429	0.393	2.080	0.506	0.468	2.028	0.535	0.516	2.082	0.493	0.441
β_2	0%	4.072	0.360	0.327	4.134	0.420	0.391	4.087	0.444	0.435	4.115	0.397	0.369
	10%	4.080	0.385	0.345	4.146	0.457	0.412	4.086	0.462	0.454	4.114	0.420	0.389
	20%	4.087	0.406	0.359	4.174	0.468	0.428	4.117	0.492	0.476	4.151	0.446	0.405
	30%	4.116	0.402	0.385	4.224	0.471	0.460	4.168	0.504	0.507	4.188	0.433	0.434
σ^2	0%	1.712	0.199	0.175	1.689	0.232	0.231	1.740	0.231	0.210	1.687	0.238	0.212
	10%	1.702	0.213	0.183	1.674	0.243	0.239	1.727	0.241	0.219	1.677	0.242	0.221
	20%	1.691	0.218	0.192	1.663	0.253	0.252	1.727	0.252	0.231	1.673	0.254	0.234
	30%	1.704	0.233	0.206	1.672	0.263	0.270	1.721	0.259	0.246	1.675	0.273	0.251
							$n = 300$						
β_1	0%	1.985	0.290	0.279	2.048	0.355	0.333	2.000	0.353	0.369	2.032	0.335	0.315
	10%	1.980	0.291	0.286	2.052	0.360	0.341	2.001	0.365	0.378	2.033	0.339	0.323
	20%	1.983	0.304	0.302	2.054	0.387	0.362	2.016	0.390	0.397	2.037	0.354	0.342
	30%	1.998	0.325	0.323	2.084	0.402	0.385	2.041	0.412	0.425	2.063	0.374	0.365
β_2	0%	4.096	0.294	0.271	4.104	0.348	0.323	4.082	0.394	0.358	4.079	0.327	0.305
	10%	4.093	0.312	0.283	4.116	0.366	0.337	4.084	0.416	0.371	4.085	0.353	0.319
	20%	4.113	0.327	0.299	4.142	0.388	0.357	4.098	0.444	0.392	4.101	0.384	0.338
	30%	4.140	0.321	0.317	4.200	0.375	0.378	4.151	0.431	0.416	4.141	0.378	0.358
σ^2	0%	1.796	0.163	0.149	1.788	0.197	0.198	1.815	0.188	0.177	1.785	0.190	0.181
	10%	1.788	0.171	0.156	1.781	0.208	0.206	1.810	0.198	0.186	1.784	0.203	0.191
	20%	1.782	0.181	0.164	1.787	0.224	0.218	1.802	0.209	0.196	1.786	0.215	0.202
	30%	1.790	0.195	0.177	1.785	0.244	0.248	1.805	0.225	0.209	1.786	0.230	0.217
							$n = 400$						
β_1	0%	1.998	0.247	0.238	2.039	0.298	0.284	2.011	0.305	0.315	2.033	0.275	0.268
	10%	1.997	0.253	0.245	2.041	0.298	0.291	2.008	0.309	0.323	2.026	0.280	0.275
	20%	2.004	0.268	0.261	2.066	0.327	0.311	2.022	0.336	0.344	2.050	0.298	0.294
	30%	2.012	0.292	0.280	2.072	0.365	0.333	2.053	0.361	0.368	2.065	0.318	0.315
β_2	0%	4.087	0.279	0.249	4.127	0.316	0.296	4.115	0.355	0.328	4.065	0.296	0.280
	10%	4.080	0.289	0.260	4.133	0.319	0.309	4.104	0.370	0.341	4.081	0.304	0.292
	20%	4.113	0.305	0.277	4.184	0.338	0.329	4.117	0.397	0.363	4.117	0.323	0.312
	30%	4.123	0.300	0.293	4.229	0.314	0.347	4.186	0.390	0.384	4.152	0.314	0.330
σ^2	0%	1.842	0.138	0.132	1.832	0.165	0.174	1.855	0.163	0.156	1.829	0.164	0.160
	10%	1.839	0.147	0.138	1.827	0.173	0.182	1.851	0.170	0.164	1.818	0.166	0.167
	20%	1.836	0.157	0.147	1.823	0.184	0.192	1.849	0.178	0.173	1.824	0.186	0.178
	30%	1.839	0.167	0.157	1.837	0.197	0.207	1.848	0.195	0.185	1.830	0.194	0.191
							$n = 600$						
β_1	0%	1.996	0.206	0.196	2.027	0.239	0.232	2.021	0.253	0.258	2.012	0.214	0.220
	10%	2.001	0.214	0.204	2.040	0.238	0.241	2.018	0.262	0.266	2.006	0.226	0.230
	20%	2.004	0.226	0.217	2.061	0.264	0.259	2.030	0.277	0.284	2.028	0.244	0.244
	30%	2.010	0.244	0.232	2.062	0.287	0.275	2.053	0.295	0.304	2.027	0.261	0.261
β_2	0%	4.061	0.220	0.203	4.076	0.257	0.241	4.085	0.277	0.268	4.087	0.230	0.228
	10%	4.062	0.233	0.211	4.091	0.265	0.252	4.096	0.301	0.278	4.076	0.242	0.236
	20%	4.105	0.255	0.225	4.110	0.275	0.266	4.109	0.319	0.297	4.111	0.253	0.256
	30%	4.118	0.250	0.240	4.164	0.277	0.286	4.160	0.327	0.317	4.151	0.263	0.270
σ^2	0%	1.891	0.117	0.110	1.880	0.138	0.145	1.892	0.134	0.130	1.882	0.136	0.134
	10%	1.882	0.121	0.116	1.882	0.150	0.152	1.885	0.138	0.136	1.870	0.144	0.140
	20%	1.881	0.129	0.122	1.873	0.156	0.161	1.878	0.146	0.143	1.881	0.156	0.149
	30%	1.885	0.136	0.131	1.883	0.165	0.172	1.882	0.156	0.153	1.867	0.162	0.158

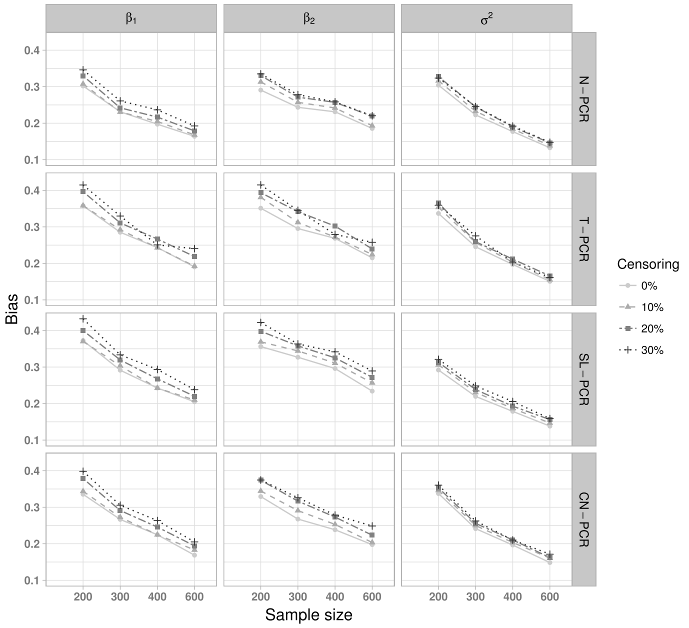

Figure 5. Simulated data. Asymptotic properties. MC mean of bias for β_1, β_2 and σ^2 for different sample sizes and levels of censoring in SMN-PCR models.

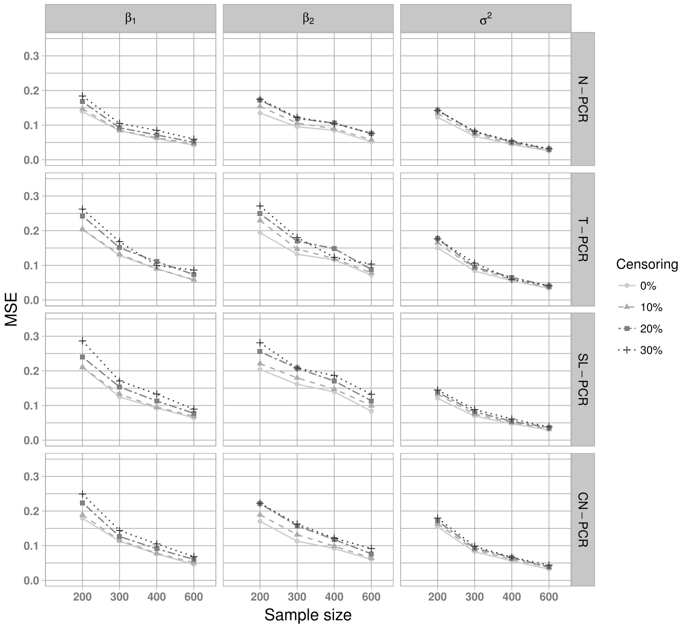

Figure 6. Simulated data. Asymptotic properties. MC mean of the Mean square error (MSE) for β_1, β_2 and σ^2 for different sample sizes and levels of censoring in SMN-PCR models.

The results of the coverage probability are presented in Table 7 (see Appendix B.2), where 95% confidence intervals were computed for each scenario using the OM-SD. As expected, the "coverage probability" is stable and around 90% for the regression parameters β_1 and β_2, but the percentage of coverage is somewhat impaired for the intervals built for σ^2. Overall, the ECME algorithm produces satisfactory estimates for the SMN-PCR fitted models. For the estimates of the nonparametric component, Figure 9 (Appendix B.2) presents the behavior of the 500 MC samples under the T-PCR model. We can note from this figure that when the censoring level increases, the variability among the nonparametric estimates functions increases, however as the sample size increases, the variability decreases. Similar results were obtained for the other models.

Now, to provide empirical evidence about the consistency of the MPL estimates, we analyze the absolute bias (Bias) and mean square error (MSE) of MPL estimates obtained from the fitted models in the simulation study. The Bias and MSE measures are given by:

$$\mathrm{Bias}(\theta_k) = \frac{1}{500}\sum_{j=1}^{500} |\widehat{\theta}_k^{(j)} - \theta_k| \quad \text{and}$$

$$\mathrm{MSE}(\theta_k) = \frac{1}{500}\sum_{j=1}^{500} (\widehat{\theta}_k^{(j)} - \theta_k)^2,$$

where $\widehat{\theta}_k^{(j)}$ is the MPL estimate of the parameter θ_k for the j-th sample, $j = 1, \ldots, 500$. From Figures 5 and 6 we can see that Bias and MSE decrease as the sample size increases. Also, we can see that the estimates obtained between the uncensored and most censored (30%) cases are quite similar. This gap is tolerably small and it becomes smaller when simple size increases. For instance, σ^2 estimates for a 600 sample size, are almost equal regardless of the censoring level and model. Hence, as a general rule the MPL estimates based on the proposed ECME algorithm for the SMN-PCR models, do show desirable asymptotic properties.

Additionally, we illustrate the capacity of the proposed diagnostic measures to identify possible influential observations. The diagnostic measures were computed from 200 MC simulations from a N-PCR model considering a sample size $n = 200$, 10% censoring level, same nonparametric function $f(t)$ as before and the benchmark setting at $c^* = 3.5$. In this scenario, we contaminated observation #82 by replacing in the parametric component $\boldsymbol{\beta}$ with $2\boldsymbol{\beta}$, $4\boldsymbol{\beta}$ and $8\boldsymbol{\beta}$ in order to generate three new contaminated responses y_{82}.

Table 6 presents the percentage of times that observation #82 was correctly identified as the most influential under different perturbation schemes considering the N-PCR model, and the percentage of times that a lower weight u_{82} was assigned with heavy-tailed models, such as, Student-t (T), slash (SL) and contaminated normal (CN). As expected, all percentages increase for higher contamination rates. First group (success percentages) represents the ability to identify influential observations in the normal model and the second one (preference percentages) indicates the robustness of heavy-tailed distributions, since a smaller

Table 6. Simulated data. Success percentages for different perturbation schemes in the N-PCR model and preference percentages under the T, SL and CN models, for different contamination schemes

	Normal				Heavy-tailed		
Contamination	Case-weight	Scale	Explanatory	Response	T	SL	CN
$2\boldsymbol{\beta}$	71.5	71.5	72.5	70.5	71.5	70.5	71.5
$4\boldsymbol{\beta}$	95.5	95.5	94.5	89.0	94.0	94.0	94.0
$8\boldsymbol{\beta}$	97.5	97.5	96.5	96.5	96.5	96.5	95.5

weight is attributed to influential observations and it increases due to the high contamination rate.

7. CONCLUSIONS

This article proposes a maximum penalized likelihood implementation for partially censored linear models considering heavy-tailed errors following a scale mixtures of normal distribution. Thus, this paper generalizes the work by [9] and [23] by incorporating a nonparametric component in the model that permits an easy and flexible modeling of possible nonlinear patterns introduced by some covariate. We also propose influence diagnostic tools for detecting influential observations in the context of PCR models with heavy-tailed distribution errors. The diagnostic analysis is based on local influence techniques presented in [33] and [34]. We carried out extensive simulation studies, and the results showed that the proposed model has good asymptotic properties and is very robust to outlying observations, outperforming the traditional normal errors model. Further, we applied our method to a real dataset to illustrate how the procedure developed can be used to evaluate model assumptions, identify outliers and obtain robust parameter estimates. To the best of our knowledge, this paper provides a first attempt to incorporate censoring in the context of partially linear models with heavy-tailed distributions from a likelihood-based perspective.

A natural extension would be to incorporate skewness and heavy tailedness simultaneously using scale mixtures of skew-normal (SMSN) distributions, as proposed in [18]. The proposed model can be extended in a full nonparametric way by using, for instance, Dirichlet process mixtures (DPMs) with a SMN or SMSN basal distribution. Other extensions of the current work include considering semiparametric mixed effects models with censored data, following the same lines of ideas proposed by [25] and [24]

APPENDIX A. COMPLEMENTARY RESULTS: STANDARD ERROR APPROXIMATION

Let $\mathbf{x}_i^* = (\mathbf{x}_i^\top, \mathbf{n}_i^\top)^\top$, $\boldsymbol{\eta} = (\boldsymbol{\beta}^\top, \mathbf{f}^\top)^\top$, for the Equations (16) and (17), the first and second derivatives of D_i, d_i for $\boldsymbol{\theta} = (\boldsymbol{\eta}, \sigma^2)$, using the notation $\ddot{D}_{\boldsymbol{\theta}}(D_i) = \frac{\partial D_i}{\partial \boldsymbol{\theta}}$, $D_{\boldsymbol{\theta}}(d_i) = \frac{\partial d_i}{\partial \boldsymbol{\theta}}$, $\ddot{D}^2_{\boldsymbol{\theta}\boldsymbol{\theta}^\top}(D_i) = \frac{\partial^2 D_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$ and $D^2_{\boldsymbol{\theta}\boldsymbol{\theta}^\top}(d_i) = \frac{\partial^2 d_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$, are given by:

$$\ddot{D}_{\eta}(D_i) = -\frac{\mathbf{x}_i^*}{\sigma}, \quad \ddot{D}_{\sigma^2}(D_i) = \frac{1}{2(\sigma^2)^{3/2}}(y_i - \mathbf{x}_i^{*\top}\boldsymbol{\eta}),$$

$$\ddot{D}^2_{\eta\eta^\top}(D_i) = \mathbf{0}, \quad \ddot{D}^2_{\sigma^2\sigma^2}(D_i) = \frac{3}{4(\sigma^2)^{5/2}}(y_i - \mathbf{x}_i^{*\top}\boldsymbol{\eta}),$$

$$\ddot{D}^2_{\eta\sigma^2}(D_i) = \frac{\mathbf{x}_i^*}{(\sigma^2)^{3/2}}, \quad D_{\eta}(d_i) = -\frac{1}{\sigma^2}(\mathbf{x}_i^* y_i - \mathbf{x}_i^*\mathbf{x}_i^{*\top}\boldsymbol{\eta}),$$

$$D_{\sigma^2}(d_i) = -\frac{1}{(\sigma^2)^2}(y_i - \mathbf{x}_i^{*\top}\boldsymbol{\eta})^2, \quad D^2_{\eta\eta^\top}(d_i) = \frac{2}{\sigma^2}\mathbf{x}_i^*\mathbf{x}_i^{*\top},$$

$$D^2_{\sigma^2\sigma^2}(d_i) = \frac{2}{(\sigma^2)^3}(y_i - \mathbf{x}_i^{*\top}\boldsymbol{\eta})^2 \quad \text{and}$$

$$D^2_{\eta\sigma^2}(d_i) = \frac{2}{(\sigma^2)^2}(\mathbf{x}_i^* y_i - \mathbf{x}_i^*\mathbf{x}_i^{*\top}\boldsymbol{\eta}).$$

For $\mathrm{I}^2(\boldsymbol{\theta})$, it is straightforward to find that $\mathrm{I}^2_{\boldsymbol{\eta}\boldsymbol{\eta}^\top} = \alpha\mathbf{K}^*$, $\mathrm{I}^2_{\boldsymbol{\eta}\sigma^2} = \mathbf{0}$ and $\mathrm{I}^2_{\sigma^2\sigma^2} = -\frac{n-m}{2(\sigma^2)^2}$, where $\mathbf{K}^*$ is a block diagonal matrix of dimension $(p+r) \times (p+r)$, given by:

$$\mathbf{K}^* = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{K} \end{bmatrix}.$$

APPENDIX B. ADDITIONAL RESULTS

B.1 Application: wage rate data

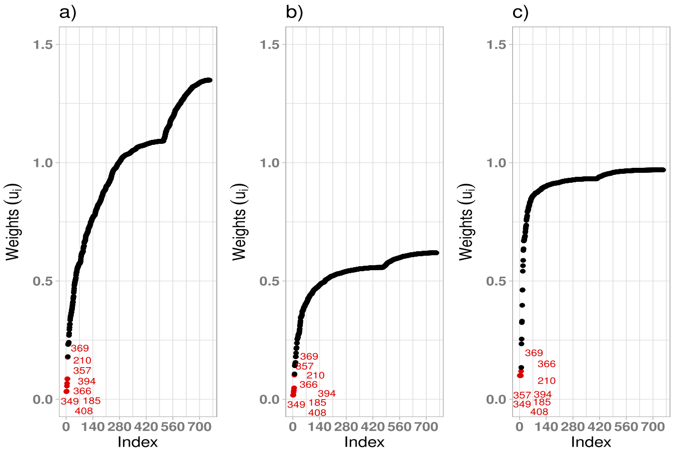

Figure 7. PSID-1975 dataset. Estimated weights u_i for: a) T-PCR, b) SL-PCR and c) CN-PCR models.

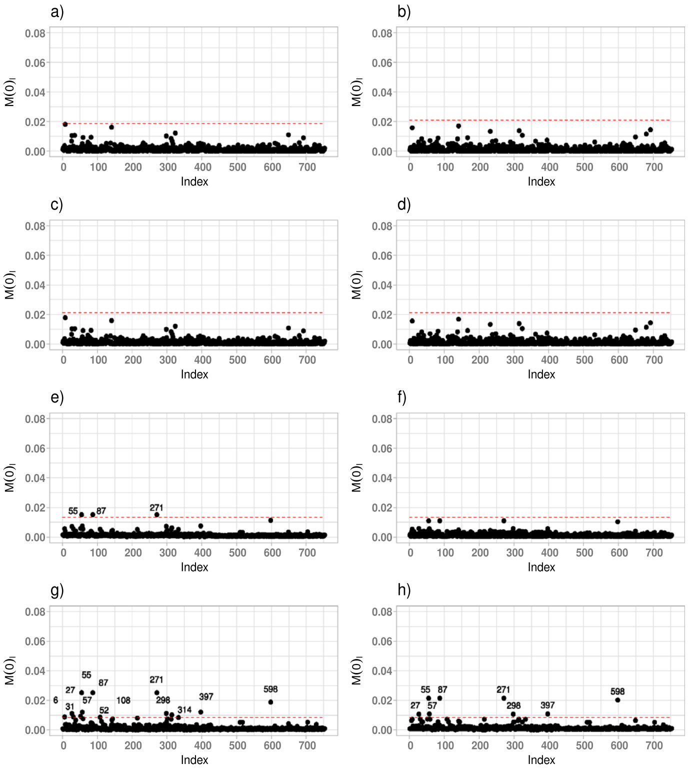

Figure 8. PSID-1975 dataset. Index plots of $M(0)_l$ for assessing local influence. Different perturbations schemes (case-weight, scale, explanatory variable and response variable perturbation) are shown in the rows from top to bottom. The T-PCR and CN-PCR model correspond to the columns from left to right.

B.2 Simulation study

Table 7. Simulated data. Coverage probability (%) based on 500 samples from the SMN-PCR model, considering different left censoring levels (C.L)

		Model											
		N-PCR			**T-PCR**			**SL-PCR**			**CN-PCR**		
n	C.L.	β_1	β_2	σ^2	β_1	β_2	σ^2	β_1	β_2	σ^2	β_1	β_2	σ^2
200	0%	92.4	91.2	88.8	93.4	92.6	89.0	95.6	94.2	89.6	93.0	92.0	81.8
	10%	93.0	90.6	88.0	93.2	90.6	87.0	95.2	93.8	88.0	92.2	92.6	82.8
	20%	93.4	91.0	88.2	91.8	91.0	86.0	95.6	91.2	89.8	92.8	91.6	85.0
	30%	93.2	93.2	83.0	93.2	93.0	80.6	94.8	92.0	81.6	92.2	93.0	87.0
300	0%	94.2	92.2	88.8	93.2	93.0	86.6	97.2	92.2	85.8	93.2	93.0	84.8
	10%	93.8	91.4	90.2	94.0	92.0	88.4	96.8	92.0	86.8	94.6	92.0	84.6
	20%	95.2	91.4	81.2	93.2	91.6	89.2	95.8	91.4	88.0	95.2	90.0	85.0
	30%	96.2	93.2	84.0	93.0	92.6	88.4	96.6	92.2	87.0	94.6	92.8	87.6
400	0%	94.6	91.0	83.4	94.6	90.4	93.6	95.6	91.2	91.8	94.8	93.2	88.0
	10%	95.0	90.8	85.0	94.8	93.8	92.2	95.2	92.6	90.6	95.4	94.0	87.2
	20%	96.0	90.0	84.6	92.6	90.6	90.6	96.2	92.2	91.6	94.8	93.0	88.6
	30%	95.2	92.4	87.6	93.2	91.0	90.8	95.2	92.2	91.4	95.6	93.6	91.2
600	0%	94.6	92.6	92.4	95.2	91.8	90.5	96.2	94.0	93.6	96.0	95.2	92.2
	10%	95.8	92.2	90.2	96.6	94.4	95.6	96.2	92.8	92.2	95.8	93.4	91.8
	20%	93.6	91.0	88.8	94.8	93.4	95.8	96.6	92.8	95.2	94.6	94.4	92.4
	30%	92.6	91.8	91.4	95.2	92.8	90.4	95.6	92.0	94.6	94.2	93.6	94.8

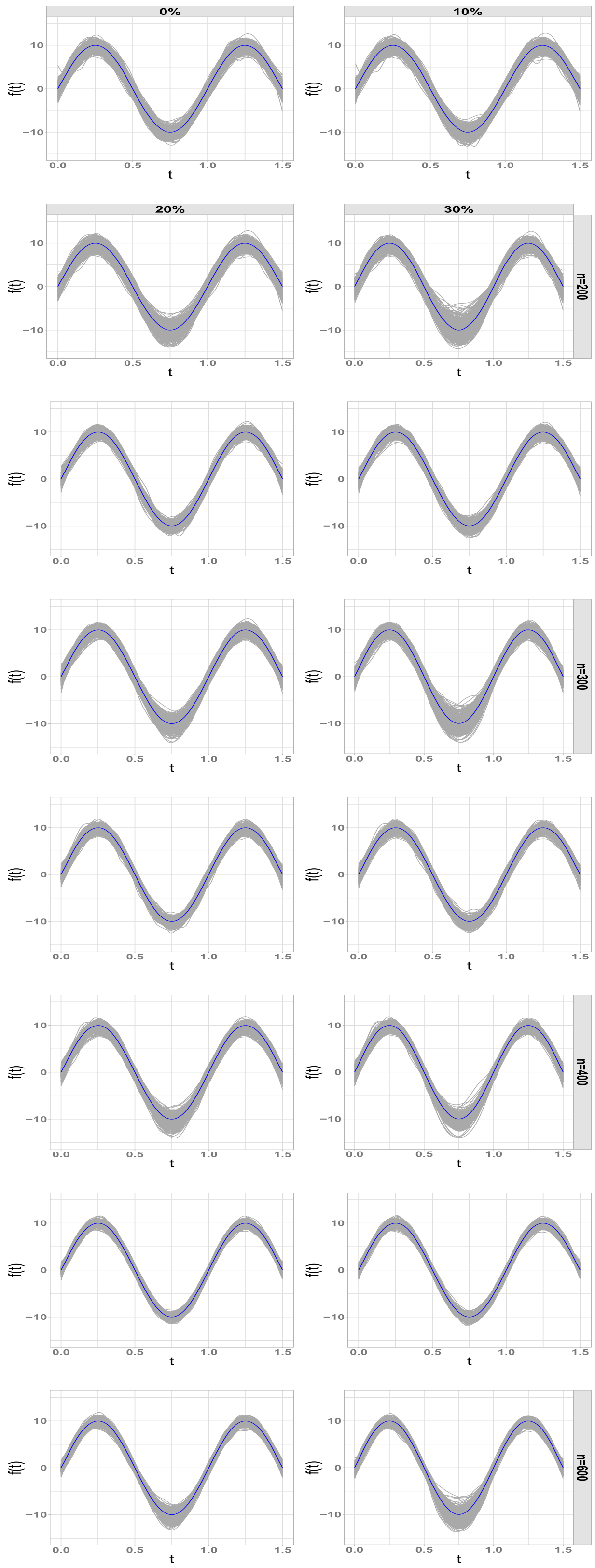

Figure 9. Simulated data. Behavior of the nonparametric component based on 500 samples from the T-PCR model. True curve (blue line) and adjusted curves (gray lines).

Author contributions

The codes implemented in this work, were organized in the R package `PartCensReg` and give computational support for estimation procedure and diagnostic analysis. The package is available in the CRAN repository [20].

Financial disclosure

The first author is partially supported by CAPES, Brazil. The research of C.E. Galarza was supported by Grant 2015/17110-9 from FAPESP-Brazil.

Conflict of interest

The authors declare no potential conflicts of interest.

Received 7 September 2019

REFERENCES

[1] Arellano-Valle, R.B., Castro, L.M., González-Farías, G., Muñoz-Gajardo, K.A., 2012. Student-t censored regression model: properties and inference. *Statistical Methods & Applications* **21**, 453–473. MR2992913

[2] Azzalini, A., 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178. MR0808153

[3] Castro, L.M., Lachos, V.H., Ferreira, G.P., Arellano-Valle, R.B., 2014. Partially linear censored regression models using heavy-tailed distributions: A Bayesian approach. *Statistical Methodology* **18**, 345–352. MR3151861

[4] Cook, R.D., 1977. Detection of influential observation in linear regression. *Technometrics* **19**, 15–18. MR0436478

[5] Cook, R.D., 1986. Assessment of local influence (with discussion). *Journal of the Royal Statistical Society, Series B* **48**, 133–169. MR0867994

[6] Cook, R.D., Weisberg, S., 1982. *Residuals and Influence in Regression*. Chapman & Hall/CRC, Boca Raton, FL. MR0675263

[7] Ferreira, C.S., Paula, G.A., 2017. Estimation and diagnostic for skew-normal partially linear models. *Journal of Applied Statistics* **44**, 3033–3053. MR3721088

[8] Garay, A.M., Bolfarine, H., Lachos, V.H., Cabral, C.R.B., 2015. Bayesian analysis of censored linear regression models with scale mixtures of normal distributions. *Journal of Applied Statistics* **42**, 2694–2714. MR3428840

[9] Garay, A.M., Lachos, V.H., Bolfarine, H., Cabral, C.R.B., 2017. Linear censored regression models with scale mixtures of normal distributions. *Statistical Papers* **58**, 247–278. MR3610167

[10] Genç, A.İ., 2013. Moments of truncated normal/independent distributions. *Statistical Papers* **54**, 741–764. MR3072898

[11] Green, P.J., 1987. Penalized likelihood for general semi-parametric regression models. *International Statistical Review/Revue Internationale de Statistique* 245–259. MR0963142

[12] Green, P.J., Silverman, B.W., 1993. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. CRC Press. MR1270012

[13] Härdle, W., Müller, M., Sperlich, S., Werwatz, A.W., 2004. *Nonparametric and Semiparametric Models*. Springer-VerlagBerlin, Heidelberg. MR2061786

[14] Hastie, T., Tibshirani, R., 1990. *Generalized Additive Models*. Wiley Online Library. MR1082147

[15] Ibacache-Pulgar, G., Paula, G.A., 2011. Local influence for Student-t partially linear models. *Computational Statistics & Data Analysis* **55**, 1462–1478. MR2741427

[16] Ibacache-Pulgar, G., Paula, G.A., Cysneiros, F., 2013. Semiparametric additive models under symmetric distributions. *Test* **22**, 103–121. MR3028246

[17] Lachos, V.H., Angolini, T., Abanto-Valle, C.A., 2011. On estimation and local influence analysis for measurement errors models under heavy-tailed distributions. *Statistical Papers* **52**, 567–590. MR2821058

[18] Lachos, V.H., Ghosh, P., Arellano-Valle, R.B., 2010. Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* **20**. SS-08-045. MR2640696

[19] Lee, S.Y., Xu, L., 2004. Influence analysis of nonlinear mixed-effects models. *Computational Statistics and Data Analysis* **45**, 321–341. MR2045635

[20] Lemus, M.N., Galarza, C.E., Matos, L.A., Lachos, V.H., 2018. PartCensReg: Partially Censored Regression Models Based on Heavy-Tailed Distributions. URL: https://CRAN.R-project.org/package=PartCensReg, doi: https://doi.org/10.13140/RG.2.2.19662.97600. r package version 1.38.

[21] Liu, C., Rubin, D.B., 1994. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika* **80**, 267–278. MR1326414

[22] Mark, R.S., Bacchetti, P., Jewell, N.P., 1994. Variances for maximum penalized likelihood estimates obtained via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **56**, 345–352.

[23] Massuia, B.M., Cabral, C.R.B., Matos, L.A., Lachos, V.H., 2015. Influence diagnostics for Student-t censored linear regression models. *Statistics* **49**, 1074–1094. MR3378032

[24] Matos, L.A., Bandyopadhyay, D., Castro, L.M., Lachos, V.H., 2015. Influence assessment in censored mixed-effects models using the multivariate Student's t distribution. *Journal of Multivariate Analysis* **141**, 104–117. MR3390061

[25] Matos, L.A., Lachos, V.H., Balakrishnan, N., Labra, F.V., 2013. Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistics & Data Analysis* **57**, 450–464. MR2981101

[26] Mroz, T.A., 1987. The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica: Journal of the Econometric Society* 765–799. MR2943175

[27] Osorio, F., Paula, G.A., Galea, M., 2007. Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis* **51**, 4354–4368. MR2364450

[28] Poon, W.Y., Poon, Y.S., 1999. Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**, 51–61. MR1664096

[29] R Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

[30] Relvas, C.E.M., Paula, G.A., 2016. Partially linear models with first-order autoregressive symmetric errors. *Statistical Papers* **57**, 795–825. MR3557373

[31] Vanegas, L.H., Paula, G.A., 2017. Log-symmetric regression models under the presence of non-informative left-or right-censored observations. *TEST* **26**, 405–428. MR3650533

[32] Wu, L., 2010. *Mixed Effects Models for Complex Data*. Chapman & Hall/CRC, Boca Raton, FL.

[33] Zhu, H., Lee, S., 2001. Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B* **63**, 111–126. MR1811994

[34] Zhu, H., Lee, S.Y., Wei, B.C., Zhou, J., 2001. Case-deletion measures for models with incomplete data. *Biometrika* **88**, 727–737. MR1859405

Marcela Nuñez Lemus
Departamento de Estatística
Universidade Estadual de Campinas
Campinas
Brazil
E-mail address: marcelita.nl@gmail.com

Victor H. Lachos
Department of Statistics
University of Connecticut
215 Glenbrook Rd. U-4120
Storrs, CT 06269-4120
USA
E-mail address: hlachos@uconn.edu

Christian E. Galarza
ESPOL Polytechnic University
Escuela Superior Politécnica del Litoral, ESPOL
Facultad de Ciencias Naturales y Matemáticas, FCNM
Campus Gustavo Galindo Km. 30.5 Vía Perimetral
P.O. Box 09-01-5863
Guayaquil
Ecuador
E-mail address: chedgala@espol.edu.ec

Larissa A. Matos
Departamento de Estatística
Universidade Estadual de Campinas
Rua Sérgio Buarque de Holanda, 651, Cidade Universitária
13083-859, Campinas/SP
Brazil
E-mail address: larissam@ime.unicamp.br