Influence diagnostics for censored linear regression models with skewed and heavy-tailed distributions

Thalita do Bem Mattos[a] Victor H. Lachos[a*] Aldo M. Garay[b]

[a]*Departamento de Estatística, Universidade Estadual de Campinas, Brazil*

[b]*Departamento de Estatística, Universidade Federal de Pernambuco, Brazil*

Abstract

The scale mixtures of skew-normal (SMSN) distributions (Lachos *et al.*, 2010) form an attractive class of asymmetrical heavy-tailed densities that includes the skew-normal, skew-t, skew-slash, skew - contaminated normal and the entire family of scale mixtures of normal (SMN) distributions as special cases. A robust censored linear model based on the scale mixtures of skew-normal (SMSN) distributions has been recently proposed by Mattos *et al.* (2015), where a stochastic approximation of the EM (SAEM) algorithm is presented for iteratively computing maximum likelihood estimates of the parameters. In this paper, to examine the performance of the proposed model, case-deletion and local influence techniques are developed to show its robust aspect against outlying and influential observations. This is done by analyzing the sensitivity of the SAEM estimates under some usual perturbation schemes in the model or data and by inspecting some proposed diagnostic graphs. The efficacy of the method is verified through the analysis of simulated datasets and modeling a real dataset from stellar astronomy previously analyzed under normal errors.

Keywords Case-deletion model; Censored regression model; Local influence; SAEM algorithm.

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, astrophysics, clinical testing, among many others. For example, in astrophysics, the study of the differences in the abundance of the light element beryllium (Be) in stars that do and do not host extrasolar planetary systems is usually conducted under the censored Tobit model (Santos *et al.*, 2002). In AIDS research, the viral load measures can be subject to some upper and lower 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, Wu, 2010).

*Corresponding author. Address for correspondence: Departamento de Estatística, Rua Sérgio Buarque de Holanda, 651, Cidade Universitária Zeferino Vaz, Campinas, São Paulo, Brazil. CEP 13083-859. E-mail: hlachos@ime.unicamp.br

In the framework of censored regression (CR) models, the random errors are routinely assumed to have a normal distribution for mathematical convenience. However, it is well known that several phenomena do not always fit under the assumptions of the normal model, yielding data with a distribution having simultaneously heavier tails and skewness. A good alternative is to consider observational errors with scale mixtures of skew normal (SMSN) distributions, so that the SMSN censored regression (SMSN-CR) model is defined. Recently, Mattos et al. (2015) developed a full likelihood approach for SMSN-CR models, including the implementation of the SAEM algorithm for maximum likelihood (ML) estimation with the likelihood function, predictions of unobservable values of the response and asymptotic standard errors as byproducts. In this paper, to perform diagnostics analysis in SMSN-CR models, we first discuss briefly the SAEM algorithm proposed by Mattos *et al.* (2015). After this, we develop and present the diagnostic measures for assessing global and local influence in SMSN-CR models

Since the classic normal model is very sensitive to outlying observations, the assessment of robustness aspects 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 Cook & Weisberg (1982) and the references therein. Nevertheless, research on the influence of small perturbations in the model/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. Following the pioneering work of Cook (1986), this area of research has received considerable attention in the statistical literature in linear regression models. However, for the SMSN-CR model the marginal log-likelihood function is complex for many applications, and a direct application of Cook's approach may be very difficult, because these measures involve the first and second partial derivatives of this function. The work of Zhu & Lee (2001) presents an approach to perform local influence analysis for general statistical models with missing data by working with a Q-displacement function, closely related to the conditional expectation of the complete-data log-likelihood at the E-step of the SAEM algorithm. This approach produces results very similar to those obtained from Cook's method. Moreover, the case-deletion can be studied by the Q-displacement function following the approach of Zhu *et al.* (2001) and Zhu *et al.* (2009). So, we develop here methods to obtain case-deletion measures and local influence measures by using the method of Zhu *et al.* (2001) (see also Zhu & Lee, 2001; Lee & Xu, 2004) in the context of regression models with censored data. This method or modifications of it have been applied successfully to perform influence analysis in several regression models, see for example Bolfarine *et al.* (2007), Ying-Zi *et al.* (2009), Zeller *et al.* (2010), Zeller *et al.* (2011), Lachos *et al.* (2011), Santana *et al.* (2011), Matos *et al.* (2013), among others. Using this general method and also applying the method of Lee & Xu (2004), in this paper we develop a local influence approach for SMSN-CR models and show that it leads to simple influence measures. We believe the results developed here form a necessary supplement to those presented by Mattos *et al.* (2015) for the analysis of SMSN-CR models.

The rest of the paper is organized as follows. Section 2 develops the SMSN-CR model specification and the SAEM algorithm for ML estimation developed by Mattos *et al.* (2015). In section 3 we develop influence diagnostic techniques, based on case deletion and local influence approaches. The method is illustrated in Section 4 using a motivating dataset from stellar astronomy, available in the R package **`astrodatR`**. Section 5 presents a simulation study evaluating the efficiency of our method in detecting outliers under various degrees of data perturbation and censoring. Finally, Section 6 presents some concluding remarks and possible avenues for future research.

2 The SMSN censored linear regression model

2.1 Preliminaries

A random variable Z has a skew-normal distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ, denoted by $Z \sim SN(\mu,\sigma^2,\lambda)$, if its probability density function (pdf) is given by:

$$f_{SN}(z|\mu,\sigma^2,\lambda) = 2\phi(z|\mu,\sigma^2)\Phi\left(\frac{\lambda(z-\mu)}{\sigma}\right), \quad z \in \mathbb{R}, \tag{1}$$

where $\phi\left(\,\cdot\,|\mu,\sigma^2\right)$ denotes the density of the univariate normal distribution with mean μ and variance $\sigma^2 > 0$ and $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard univariate normal distribution. A random variable Y has a SMSN distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ, denoted by SMSN$(\mu,\sigma^2,\lambda;H)$, if it has the following stochastic representation:

$$Y = \mu + \kappa(U)^{1/2}Z, \;\; U \perp Z, \tag{2}$$

where $Z \sim SN(0,\sigma^2,\lambda)$, $\kappa(\cdot)$ is a positive function, U is a positive random variable with cdf $H(\,\cdot\,|\boldsymbol{\nu})$ indexed by a scalar or vector parameter $\boldsymbol{\nu}$, and $U \perp Z$ indicates that the random variables U and Z are independent.

The random variable U is known as *the scale factor* and its cdf $H(\,\cdot\,|\boldsymbol{\nu})$ is called the *mixing distribution function*. Note that when $\lambda = 0$, the SMSN family reduces to the symmetric class of SMN distributions. A random variable $Y \sim$ SMSN$(\mu,\sigma^2,\lambda;H)$, has a stochastic representation given by

$$Y = \mu + \Delta T + \kappa(U)^{1/2}\tau^{1/2}T_1, \tag{3}$$

where $\Delta = \sigma\delta$, $\tau = (1-\delta^2)\sigma^2$, $\delta = \dfrac{\lambda}{\sqrt{1+\lambda^2}}$, $T = \kappa(U)^{1/2}|T_0|$, $T_0 \perp T_1$ and $|\cdot|$ denotes absolute value. The representation given in (3) is very appropriate to derive some mathematical properties and can be used to simulate pseudo-realizations of Y and also to implement the SAEM algorithm (Mattos *et al.*, 2015).

Using the representation given in Equation (2), we observe that the marginal density of Y is given by:

$$f_{SMSN}(y|\mu,\sigma^2,\lambda;H) = 2\int_0^{\infty} \phi(y|\mu,\kappa(u)\sigma^2)\Phi\left(\frac{\lambda(y-\mu)}{\sigma\kappa(u)^{1/2}}\right) dH(u). \tag{4}$$

Another important result is the cdf of a SMSN random variable, $Y \sim$ SMSN$(\mu,\sigma^2,\lambda;H)$, which is given by:

$$F_{SMSN}(y \mid \mu,\sigma^2,\lambda\,;\,H) = 2\int_0^{\infty} \Phi_2\Big(\boldsymbol{y}(u)^*|\boldsymbol{\mu}^*,\boldsymbol{\Sigma}\Big)dH(u), \tag{5}$$

where

$$\boldsymbol{y}(u)^* = (\kappa(u)^{-1/2}y,0)^\top, \;\; \boldsymbol{\mu}^* = (\mu,0)^\top, \;\; \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & -\delta\sigma \\ -\delta\sigma & 1 \end{pmatrix} \tag{6}$$

and $\Phi_m(\cdot \mid \boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$ denotes the cdf of the $m-$variate normal distribution with mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$.

Another important class of distributions, which will be useful in our development, is the truncated SMSN distributions.

Let the random variable $W \sim \text{SMSN}(\mu,\sigma^2,\lambda;H)$ with $\mathbb{P}(a < W < b) > 0$, where $a < b$. A random variable Y has a truncated SMSN distribution in the interval $\lfloor a,b \rfloor$, denoted by $Y \sim \text{TSMSN}(\mu,\sigma^2,\lambda;H,\lfloor a,b \rfloor)$, if it has the same distribution as $W|W \in \lfloor a,b \rfloor$, where $\lfloor a,b \rfloor$ means that each extreme of the interval can either be open or closed.

Thus, the pdf of the random variable $Y \sim \text{TSMSN}(\mu,\sigma^2,\lambda;H,\lfloor a,b \rfloor)$ is given by:

$$f_{TSMSN}(y \mid \mu,\sigma^2,\lambda\,;\,H,\lfloor a,b \rfloor) = \frac{f_{SMSN}(y \mid \mu,\sigma^2,\lambda\,;\,H)}{F_{SMSN}(b \mid \mu,\sigma^2,\lambda\,;\,H) - F_{SMSN}(a \mid \mu,\sigma^2,\lambda\,;\,H)} \mathbb{1}_{\lfloor a,b \rfloor}(y),$$

where $\mathbb{1}_{\mathbb{A}}(\cdot)$ denotes the indicator function of the set $\mathbb{A}$, i.e., $\mathbb{1}_{\mathbb{A}}(y) = 1$ if $y \in \mathbb{A}$ and $\mathbb{1}_{\mathbb{A}}(y) = 0$ otherwise, and $f_{SMSN}(\cdot \mid \mu,\sigma^2,\lambda;H)$ and $F_{SMSN}(\cdot \mid \mu,\sigma^2,\lambda;H)$ represent the pdf and cdf of the SMSN distribution, respectively.

Although we can deal with any $\kappa(.)$ function, hereafter we restrict our attention to the case where $\kappa(u) = 1/u$, since it leads to good mathematical properties (see, for instance, Basso *et al.* (2010) and Labra *et al.* (2012)).

2.2 Model specification

The SMSN-CR model introduced by Mattos *et al.* (2015) is defined by:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1,\dots,n, \tag{7}$$

where $\varepsilon_i \overset{\text{iid}}{\sim} SMSN\left(-\sqrt{\frac{2}{\pi}}k_1\Delta,\sigma^2,\lambda;H\right)$, $\boldsymbol{\beta} = (\beta_1,\dots,\beta_p)^\top$ is a vector of regression parameters, Y_i is a response variable, $\mathbf{x}_i = (x_{i1},\dots,x_{ip})^\top$ is a vector of explanatory variables for the $i-$th subject and *iid* means that the errors $\boldsymbol{\varepsilon} = (\varepsilon_1,\varepsilon_2,\dots,\varepsilon_n)$ are independent and identically distributed.

It is important to note that the value of the location parameter of ε_i, $-\sqrt{\frac{2}{\pi}}k_1\Delta$ with $k_1 = E[U^{-1}]$, was chosen in order to obtain $E[\varepsilon_i] = 0$, as in the normal model. Thus, when the moments exist, we have:

$$Y_i \sim SMSN\left(\mathbf{x}_i^\top \boldsymbol{\beta} - \sqrt{\frac{2}{\pi}}k_1\Delta,\sigma^2,\lambda;H\right),$$

with $E[Y_i] = \mathbf{x}_i^\top \boldsymbol{\beta}$. Moreover, we are interested in the situation in which the response variable is not fully observed for all subjects. Thus, for the $i-$th subject and assuming left-censoring, Y_i is a latent variable and the observe data (V_i,ρ_i) is of the form:

$$V_i = \begin{cases} c_i & \text{if} \quad \rho_i = 1 \;\; (\text{i.e. } Y_i \le c_i); \\ Y_i & \text{if} \quad \rho_i = 0 \;\; (\text{i.e. } Y_i > c_i), \end{cases} \tag{8}$$

for some known threshold point c_i, $i = 1,\dots,n$.

In the SMSN-CR model, defined by combining Equations (7)–(8), the log-likelihood function of parameters $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top,\sigma^2,\lambda\right)^\top$, given the observed data $(\boldsymbol{\nu},\boldsymbol{\rho})$, is defined by:

$$\ell(\boldsymbol{\theta}|\boldsymbol{\nu},\boldsymbol{\rho}) = \log\left\{\prod_{i=1}^{n}\left[F_{SMSN}\left(\frac{v_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}|\boldsymbol{\theta};H\right)\right]^{\rho_i}\left[f_{SMSN}(v_i|\boldsymbol{\theta};H)\right]^{1-\rho_i}\right\}, \tag{9}$$

where $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ is the observed sample of $\mathbf{V} = (V_1, V_2, \ldots, V_n)$ and $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_n)$. Thus, $\rho_i = 1(\text{or } = 0)$ means that the i-th observation is censored (or not censored), i.e., $Y_i \leq c_i$ (or $Y_i \leq c_i$), respectively. For simplicity, we will assume the data are left censored, and develop the SAEM algorithm for ML estimation. Extensions to right and interval censored data are immediate.

We assume that the degrees of freedom parameter $(\boldsymbol{\nu})$ of the SMSN distributions is fixed. To choose the most appropriate value of this parameter, we use the log-likelihood profile (see also Lange *et al.*, 1989; Meza *et al.*, 2012). This assumption is based on the work of Lucas (1997), in which he showed that the protection against outliers is preserved only if the degrees of freedom parameter is fixed.

2.3 ML estimation via the SAEM algorithm

In this section we discuss briefly the ML estimation of the parameters in the SMSN-CR models via the SAEM algorithm. Here, we do not focus on the ML estimation. Interested readers can refer to Mattos *et al.* (2015) for further details.

Let $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^\top, \Delta, \tau)^\top$ be the vector of parameters in question, which has a one-to-one correspondence with the original vector of parameters $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma^2, \lambda\right)^\top$. We can obtain σ^2 and λ from Δ and τ considering:

$$\sigma^2 = \tau + \Delta^2 \quad \text{and} \quad \lambda = \Delta/\sqrt{\tau}. \tag{10}$$

From Basso *et al.* (2010), a useful straightforward result is that the conditional distribution of T_i given y_i and u_i is $TN\left(\mu_{Ti} - \sqrt{\frac{2}{\pi}}k_1, u_i^{-1}M_T^2; \lfloor -\sqrt{\frac{2}{\pi}}k_1, \infty \rfloor\right)$, with

$$\mu_{T_i} = \frac{\Delta}{\Delta^2 + \tau}\left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right) \quad \text{and} \quad M_T^2 = \frac{\tau}{\Delta^2 + \tau}. \tag{11}$$

As in the classic EM algorithm, in order to implement the SAEM algorithm, we consider a data augmentation scheme that consists of assuming that the latent variables (*missing data*) in the model, given by the vector of censored responses $\mathbf{Y} = (y_1, y_2, \ldots, y_n)^\top$, the vector $\mathbf{t} = (t_1, t_2, \ldots, t_n)^\top$ and $\mathbf{u} = (u_1, u_2, \ldots, u_n)^\top$, were in fact observed. Thus, considering the observed data $(\mathbf{V}, \boldsymbol{\rho})$ and the latent variables $(\mathbf{Y}, \mathbf{t}, \mathbf{u})$, we define the complete dataset by $\mathbf{Y}_{comp} = (\mathbf{V}^\top, \boldsymbol{\rho}^\top, \mathbf{Y}^\top, \mathbf{t}^\top, \mathbf{u}^\top)^\top$. Then, it is easy to derive the complete data log-likelihood, defined by $\ell_{comp}(\boldsymbol{\theta}^* | \mathbf{Y}_{comp})$, which is given by:

$$\ell_{comp}(\boldsymbol{\theta}^* | \mathbf{Y}_{comp}) \propto cte - \frac{n}{2}\log\tau - \frac{1}{2\tau}\sum_{i=1}^{n} u_i (y_i - \mathbf{x}_i^\top \boldsymbol{\beta} - \Delta t_i)^2 + \sum_{i=1}^{n} \log h(u_i | \boldsymbol{\nu}), \tag{12}$$

where cte is a constant that is independent of $\boldsymbol{\theta}^*$ and $h(\cdot|\boldsymbol{\nu})$ is the pdf of the random variable U. In what follows the superscript (j) indicates the estimate of the related parameter at stage j of the algorithm.

Thus, our SAEM algorithm for the SMN-CR models can be summarized in the following steps:

- **E-step:** Given the current estimate $\boldsymbol{\theta}^{*(j)} = (\beta^{(j)\top}, \Delta^{(j)}, \tau^{(j)})^\top$ at the $j-$th iteration, we obtain the conditional expectation of $\ell_{comp}(\boldsymbol{\theta}^*|\mathbf{Y}_{comp})$, denoted the Q-function, which is given by

$$\begin{aligned} Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{*(j)}) &= E\left[\ell_{comp}(\boldsymbol{\theta}^*|\mathbf{Y}_{comp})|\mathbf{V},\boldsymbol{\rho},\boldsymbol{\theta}^{*(j)}\right] \\ &= cte - \frac{n}{2}\log(\tau) - \frac{1}{2\tau}\sum_{i=1}^{n}\left[\mathcal{E}_{02i}(\boldsymbol{\theta}^{*(j)}) - 2\mathcal{E}_{01i}(\boldsymbol{\theta}^{*(j)})\mathbf{x}_i^\top\boldsymbol{\beta} + \mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)})(\mathbf{x}_i^\top\boldsymbol{\beta})^2\right. \\ &- \left. 2\Delta\mathcal{E}_{11i}(\boldsymbol{\theta}^{*(j)}) + 2\Delta\mathcal{E}_{10i}(\boldsymbol{\theta}^{*(j)})\mathbf{x}_i^\top\boldsymbol{\beta} + \Delta^2\mathcal{E}_{20i}(\boldsymbol{\theta}^{*(j)})\right]. \end{aligned} \tag{13}$$

Observe that the expression of the Q-function is completely determined by knowledge of the following expectations:

$$\mathcal{E}_{rsi}(\boldsymbol{\theta}^{*(j)}) = \mathrm{E}[U_i T_i^r Y_i^s|V_i,\rho_i,\boldsymbol{\theta}^{*(j)}] \quad \text{for } r,s=0,1,2, \text{ and } i=1,2,\ldots,n.$$

As presented by Basso *et al.* (2010), considering known properties of conditional expectation and Equation (11), we obtain

$$\mathcal{E}_{10i}(\boldsymbol{\theta}^{*(j)}) = \mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)})\mu_{T_i}^{(j)} + M_T^{(j)}\psi_i^{(j)}, \tag{14}$$

$$\mathcal{E}_{20i}(\boldsymbol{\theta}^{*(j)}) = \mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)})\mu_{T_i}^{2(j)} + M_T^{2(j)} + M_T^{(j)}\mu_{T_i}^{(j)}\psi_i^{(j)}, \tag{15}$$

where

$$\psi_i^{(j)} = \mathrm{E}\left[U_i W_\Phi\left(\frac{U_i\mu_{T_i}^{(j)}}{M_T^{(j)}}\right)|V_i,\rho_i,\boldsymbol{\theta}^{*(j)}\right] \text{ and } W_\Phi(a) = \frac{\phi(a)}{\Phi(a)} \text{ for } a\in\mathbb{R}.$$

Thus, at each step, to compute $\mathcal{E}_{rsi}(\boldsymbol{\theta}^{*(j)})$ we need to obtain the conditional expectations $\mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)})$ and $\psi_i^{(j)}$ for the different SMSN distributions considering two different situations:

a) *For an uncensored observation "i":*
In this case, we have that $\rho_i = 0$, thus $V_i = Y_i \sim SMSN\left(\mathbf{x}_i^\top\boldsymbol{\beta} - \sqrt{\frac{2}{\pi}}k_1\Delta, \tau+\Delta^2, \lambda; H\right)$ and, therefore,

$$\mathcal{E}_{rsi}(\boldsymbol{\theta}^{*(j)}) = y_i^s\mathcal{E}_{r0i}(\boldsymbol{\theta}^{*(j)}), \tag{16}$$

where $\mathcal{E}_{r0i}(\boldsymbol{\theta}^{*(j)})$ can be obtained using equations (14)-(15).

b) *For a censored observation* "*i*":
In this case, we have that $\rho_i = 1$, i.e. $Y_i \le c_i$, therefore

$$\mathcal{E}_{rsi}\left(\boldsymbol{\theta}^{*(j)}\right) = \mathrm{E}[U_i T_i^r Y_i^s|V_i, Y_i\le c_i,\boldsymbol{\theta}^{*(j)}], \text{ with } r,s=0,1,2. \tag{17}$$

As this conditional expectation does not have closed form, we need to introduce two intermediate steps to replace the E-step by a stochastic approximation using simulated data. Thus, the $j-$th iteration consists of the following steps:

* **S-step (Sampling)**
Let $\mathbf{Y}^{(c)} = \left(Y_1^{(c)}, Y_2^{(c)}, \ldots, Y_{n^c}^{(c)}\right)$ the vector of n^c censored cases, where $Y_i^{(c)}$ is

generated from TSMSN $\left(\mathbf{x}_i^\top \boldsymbol{\beta} - \sqrt{\frac{2}{\pi}} k_1 \Delta, \tau + \Delta^2, \lambda; H, \lfloor -\infty, c_i \rfloor \right)$ for $i = 1, \ldots, n^c$.

Thus, the new vector of observations $\mathbf{Y}^{(l,j)} = (Y_{i1}^{(l,j)}, \ldots, Y_{in^c}^{(l,j)}, Y_{n_i^c+1}, \ldots, Y_n)$ is a sample generated for the n^c censored cases and the observed values (uncensored cases), for $l = 1, \ldots, M$. Mattos *et al.* (2015) describes the details of the methods used to generate from the random variable $\mathbf{Y}^{(c)}$.

∗ **AE-step (Approximation Expectation)**
Since we have the sequence $\mathbf{Y}^{(l,j)}$, at the j-th iteration, considering Equations (14)-(15) and the results given in Basso *et al.* (2010), we replace the conditional expectations in (16) by the following stochastic approximations:

$$\mathcal{E}_{rsi}\left(\boldsymbol{\theta}^{*(j)}\right) = \mathcal{E}_{rsi}(\boldsymbol{\theta}^{*(j-1)}) + \gamma_j \left[\frac{1}{M} \sum_{l=1}^{M} \mathrm{E}[U_i T_i^r Y_i^{s(l,j)} | V_i, \rho_i, \boldsymbol{\theta}^{*(j)}] - \mathcal{E}_{rsi}(\boldsymbol{\theta}^{*(j-1)}) \right],$$

for $r, s = 0, 1, 2$.
As presented by Kuhn & Lavielle (2004), γ_j is a decreasing sequence of positive numbers such that

$$\sum_{j=1}^{\infty} \gamma_j = \infty \quad \text{and} \quad \sum_{j=1}^{\infty} \gamma_j^2 < \infty.$$

An advantage of the SAEM algorithm is that even though it performs a Monte Carlo E-step, it requires a small and fixed Monte Carlo sample size, making it much faster than MCEM.

- **CM-step:** Maximize $Q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{*(j)})$ with respect to $\boldsymbol{\theta}^*$ obtaining $\boldsymbol{\theta}^{*(j+1)}$, which leads to the following expressions:

$$\begin{aligned}
\boldsymbol{\beta}^{(j+1)} &= \left(\sum_{i=1}^{n} \mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)})(\mathbf{x}_i \mathbf{x}_i^\top) \right)^{-1} \left[\sum_{i=1}^{n} \mathbf{x}_i \mathcal{E}_{01i}(\boldsymbol{\theta}^{*(j)}) - \Delta \sum_{i=1}^{n} \mathbf{x}_i \mathcal{E}_{10i}(\boldsymbol{\theta}^{*(j)}) \right]; \\
\Delta^{(j+1)} &= \frac{\sum_{i=1}^{n} \mathcal{E}_{11i}(\boldsymbol{\theta}^{*(j)}) - \sum_{i=1}^{n} \mathcal{E}_{10i}(\boldsymbol{\theta}^{*(j)})(\mathbf{x}_i^\top \boldsymbol{\beta}^{(j+1)})}{\sum_{i=1}^{n} \mathcal{E}_{20i}(\boldsymbol{\theta}^{*(j)})}; \\
\tau^{(j+1)} &= \frac{1}{n} \left(\sum_{i=1}^{n} \left[\mathcal{E}_{02i}(\boldsymbol{\theta}^{*(j)}) - 2\mathcal{E}_{01i}(\boldsymbol{\theta}^{*(j)})(\mathbf{x}_i^\top \boldsymbol{\beta}^{(j+1)}) + \mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)})(\mathbf{x}_i^\top \boldsymbol{\beta}^{(j+1)})^2 \right.\right. ; \\
&\quad \left.\left. -\, 2\Delta^{(j+1)} \mathcal{E}_{11i}(\boldsymbol{\theta}^{*(j)}) + 2\Delta^{(j+1)} \mathcal{E}_{10i}(\boldsymbol{\theta}^{*(j)})(\mathbf{x}_i^\top \boldsymbol{\beta}^{(j+1)}) + (\Delta^{(j+1)})^2 \mathcal{E}_{20i}(\boldsymbol{\theta}^{*(j)}) \right] \right).
\end{aligned}$$

Note that using Equation (10), we have that:

$$\sigma^{2(j+1)} = \tau^{(j+1)} + \Delta^{2(j+1)} \quad \text{and} \quad \lambda^{(j+1)} = \frac{\Delta^{(j+1)}}{\sqrt{\tau^{(j+1)}}}.$$

Thus, considering the original vector of parameters $\boldsymbol{\theta}^{(j+1)} = \left(\boldsymbol{\beta}^{(j+1)\top}, \sigma^{2(j+1)}, \lambda^{(j+1)} \right)^\top$, this process is iterated until some distance, involving two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta} | \mathbf{y}_{obs})$, like

$$||\ell(\boldsymbol{\theta}^{(j+1)} | \mathbf{V}, \boldsymbol{\rho}) - \ell(\boldsymbol{\theta}^{(j)} | \mathbf{V}, \boldsymbol{\rho})|| \quad \text{or} \quad ||\ell(\boldsymbol{\theta}^{(j+1)} | \mathbf{V}, \boldsymbol{\rho}) / \ell(\boldsymbol{\theta}^{(j)} | \mathbf{V}, \boldsymbol{\rho}) - 1||,$$

is small enough.

3 Diagnostic Analysis

Influence diagnostic techniques consist of evaluating the sensitivity of the parameter estimates of a particular model when perturbation occurs either in the dataset or in the model's underlying assumptions. There are two main approaches to detect influential observations. The first one is the case-deletion technique (Cook, 1977), in which the effect or influence of a given observation is measured by comparison of parameter estimates before and after its deletion. This is done by analyzing one or more fitted models after the exclusion and then assessing the result by some metrics such as the likelihood distance or Cook's distance. The second method is the local influence approach (Cook, 1986), which evaluates the changes in the results of the analysis as a consequence of a minor perturbation of the subject, not its total deletion. In the next subsections, we introduce the case-deletion measures and the local influence measures to the censored data on the basis of the Q-function previously determined in the E-step of the SAEM algorithm. We first consider the case-deletion measures, then the local influence and finally the perturbation schemes used.

3.1 Case-deletion Measures

Case-deletion is a common approach to study the effect of dropping the $i-$th case from the dataset. From now on, the subscript "$[i]$" will denote the original dataset with the $i-$th case deleted. For example, $\mathbf{Y}_{comp[i]}$ corresponds to the complete data with the $i-$th observation deleted. Let $\widehat{\boldsymbol{\theta}}_{[i]} = \left(\widehat{\boldsymbol{\beta}}^{\top}_{[i]}, \widehat{\sigma^2}_{[i]}, \widehat{\lambda}_{[i]}\right)^{\top}$ be the maximizer of the function $Q_{[i]}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right) = \mathrm{E}\left[\ell_{comp}\left(\boldsymbol{\theta}|\mathbf{Y}_{comp[i]}\right)|\mathbf{V}_{[i]}, \boldsymbol{\rho}_{[i]}, \widehat{\boldsymbol{\theta}}\right]$, where $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\beta}}^{\top}, \widehat{\sigma^2}, \widehat{\lambda}\right)^{\top}$ is the ML estimates of $\boldsymbol{\theta}$. To assess the influence of the $i-$th case on $\widehat{\boldsymbol{\theta}}$, we compare the difference between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$. Note that $Q\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)$ achieves its global of a maximum at $\widehat{\boldsymbol{\theta}}$, if deletion of a case seriously influences the estimates, so more attention should be paid to that case. In other words, if $\widehat{\boldsymbol{\theta}}_{[i]}$ is fairly far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the $i-$th case could be considered influential. Since $\widehat{\boldsymbol{\theta}}_{[i]}$ is needed for every case, the total computational burden involved can be quite heavy, so the following one-step approximation $\widetilde{\boldsymbol{\theta}}_{[i]}$ is used to reduce the burden (Cook & Weisberg, 1982):

$$\widetilde{\boldsymbol{\theta}}_{[i]} = \widehat{\boldsymbol{\theta}} + \left\{-\ddot{Q}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)\right\}^{-1}\dot{Q}_{[i]}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right), \quad \text{for} \quad i = 1, \ldots, n, \tag{18}$$

where $\ddot{Q}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) = \left\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}Q\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)\right\}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\dot{Q}_{[i]}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) = \left\{\frac{\partial}{\partial\boldsymbol{\theta}}Q_{[i]}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)\right\}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ represent the Hessian matrix and the individual score vector, respectively.

Thus, $\dot{Q}_{[i]}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) = \left(\dot{Q}_{[i]\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right), \dot{Q}_{[i]\sigma^2}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right), \dot{Q}_{[i]\lambda}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)\right)$, with its elements given by:

$$\begin{aligned}
\dot{Q}_{[i]\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) &= \left\{\frac{\partial}{\partial\boldsymbol{\beta}}Q_{[i]}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)\right\}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \frac{\widehat{\lambda}^2+1}{\widehat{\sigma}^2}\widehat{E}_{1[i]},\\
\dot{Q}_{[i]\sigma^2}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) &= \left\{\frac{\partial}{\partial\sigma^2}Q_{[i]}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)\right\}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\frac{1}{2\widehat{\sigma}^2}\left[(n-1)-\frac{\widehat{\lambda}^2+1}{\widehat{\sigma}^2}\widehat{E}_{2[i]}+\frac{\widehat{\lambda}\sqrt{\widehat{\lambda}^2+1}}{\widehat{\sigma}}\widehat{E}_{3[i]}\right],\\
\dot{Q}_{[i]\lambda}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) &= \left\{\frac{\partial}{\partial\lambda}Q_{[i]}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)\right\}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \frac{(n-1)\widehat{\lambda}}{\widehat{\lambda}^2+1}-\frac{\widehat{\lambda}}{\widehat{\sigma}^2}\widehat{E}_{2[i]}+\frac{2\widehat{\lambda}^2+1}{\widehat{\sigma}\left(\widehat{\lambda}^2+1\right)^{1/2}}\widehat{E}_{3[i]}-\widehat{\lambda}\sum_{j\neq i}\mathcal{E}_{20j}(\widehat{\boldsymbol{\theta}}),
\end{aligned}$$

where

$$\begin{aligned}
\widehat{E}_{1[i]} &= \sum_{j\neq i}\left[\mathbf{x}_j\mathcal{E}_{01j}(\widehat{\boldsymbol{\theta}})-\mathcal{E}_{00j}(\widehat{\boldsymbol{\theta}})\mathbf{x}_j\mathbf{x}_j^{\top}\widehat{\boldsymbol{\beta}}-\frac{\widehat{\sigma}\widehat{\lambda}}{\sqrt{\widehat{\lambda}^2+1}}\mathbf{x}_j\mathcal{E}_{10j}(\widehat{\boldsymbol{\theta}})\right], \quad (19)\\
\widehat{E}_{2[i]} &= \sum_{j\neq i}\left[\mathcal{E}_{02j}(\widehat{\boldsymbol{\theta}})-2\mathcal{E}_{01j}(\widehat{\boldsymbol{\theta}})\mathbf{x}_j^{\top}\widehat{\boldsymbol{\beta}}+\mathcal{E}_{00j}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_j^{\top}\widehat{\boldsymbol{\beta}})^2\right] \quad (20)\\
\text{and}\ \widehat{E}_{3[i]} &= \sum_{j\neq i}\left[\mathcal{E}_{11j}(\widehat{\boldsymbol{\theta}})-\mathcal{E}_{10j}(\widehat{\boldsymbol{\theta}})\mathbf{x}_j^{\top}\widehat{\boldsymbol{\beta}}\right]. \quad (21)
\end{aligned}$$

Case-deletion measures can be developed to assess influential observations, such as the generalized Cook's distance and the likelihood distance (Zhu *et al.*, 2001). To assess the influence of the $i-$th case on the EM estimate $\widehat{\boldsymbol{\theta}}$, we need to compare $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$. If $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$, in some sense, then the $i-$th case is regarded as influential. Based on the metric for measuring the distance between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$ proposed by Zhu *et al.* (2001), we consider here the following *generalized Cook's distance*:

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

Upon substituting (18) into (22), we obtain the following approximation of the *generalized Cook's distance*:

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

Another measure of the influence of the $i-$th case is the following Q-distance function, similar to the *likelihood distance* LD_i (Cook & Weisberg, 1982), defined as:

$$QD_i = 2\left\{Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)-Q\left(\widehat{\boldsymbol{\theta}}_{[i]}|\widehat{\boldsymbol{\theta}}\right)\right\}. \quad (23)$$

We can compute an approximation of the likelihood displacement QD_i by substituting (18) into (23), resulting in the following approximation QD_i^1 of QD_i:

$$QD_i^1 = 2\left\{Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)-Q\left(\widetilde{\boldsymbol{\theta}}_{[i]}|\widehat{\boldsymbol{\theta}}\right)\right\}.$$

The approximated measures QD^1 and GD_i^1 have been satisfactorily applied in the context of censored regression models by Matos *et al.* (2013) and Massuia *et al.* (2015).

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

In order to obtain the measures for case-deletion diagnostics and local influence considering a particular perturbation scheme, it is necessary to compute $\ddot{Q}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)$, where $\boldsymbol{\theta}=\left(\boldsymbol{\beta}^{\top},\sigma^2,\lambda\right)^{\top}$ is the original parameter vector.

However, from Zeller *et al.* (2011), considering the parameterizations $\boldsymbol{\theta}^*=(\boldsymbol{\theta}_1^*,\boldsymbol{\theta}_2^*)$ where $\boldsymbol{\theta}_1^*=\left(\boldsymbol{\beta}^{\top},\Delta\right)^{\top}$ and $\boldsymbol{\theta}_2^*=\tau$, we have that the Hessian matrix of $\boldsymbol{\theta}^*$ is block-diagonal of the form

$$\ddot{Q}\left(\widehat{\boldsymbol{\theta}}^*|\widehat{\boldsymbol{\theta}}^*\right)=\text{block diag}\left\{\ddot{Q}_{\boldsymbol{\theta}_1^*}\left(\widehat{\boldsymbol{\theta}}^*|\widehat{\boldsymbol{\theta}}^*\right),\ddot{Q}_{\boldsymbol{\theta}_2^*}\left(\widehat{\boldsymbol{\theta}}^*|\widehat{\boldsymbol{\theta}}^*\right)\right\},$$

where

$$\ddot{Q}_{\boldsymbol{\theta}_1^*}\left(\widehat{\boldsymbol{\theta}}^*|\widehat{\boldsymbol{\theta}}^*\right)=\left\{\frac{\partial^2}{\partial\boldsymbol{\theta}_1^*\partial\boldsymbol{\theta}_1^{*\top}}Q\left(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\theta}}^*\right)\right\}\Big|_{\boldsymbol{\theta}^*=\widehat{\boldsymbol{\theta}}^*}=-\frac{1}{\widehat{\tau}}\begin{pmatrix}\sum_{i=1}^n\mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}}^*)(\mathbf{x}_i\mathbf{x}_i^{\top}) & \sum_{i=1}^n\mathbf{x}_i\mathcal{E}_{01i}(\widehat{\boldsymbol{\theta}}^*)\\ \sum_{i=1}^n\mathbf{x}_i\mathcal{E}_{01i}(\widehat{\boldsymbol{\theta}}^*) & \sum_{i=1}^n\mathcal{E}_{20i}(\widehat{\boldsymbol{\theta}}^*)\end{pmatrix}$$

and

$$\begin{aligned}\ddot{Q}_{\boldsymbol{\theta}_2^*}\left(\widehat{\boldsymbol{\theta}}^*|\widehat{\boldsymbol{\theta}}^*\right) &= \left\{\frac{\partial^2}{\partial\boldsymbol{\theta}_2^*\partial\boldsymbol{\theta}_2^{*\top}}Q\left(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\theta}}^*\right)\right\}\Big|_{\boldsymbol{\theta}^*=\widehat{\boldsymbol{\theta}}^*}\\ &= \frac{n}{2\widehat{\tau}^2}-\frac{1}{\widehat{\tau}^3}\left(\sum_{i=1}^n\left[\mathcal{E}_{02i}(\widehat{\boldsymbol{\theta}}^*)-2\mathcal{E}_{01i}(\widehat{\boldsymbol{\theta}}^*)(\mathbf{x}_i^{\top}\widehat{\boldsymbol{\beta}})+\mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}}^*)(\mathbf{x}_i^{\top}\widehat{\boldsymbol{\beta}})^2-2\widehat{\Delta}\mathcal{E}_{11i}(\widehat{\boldsymbol{\theta}}^*)\right.\right.\\ &+ \left.\left.2\widehat{\Delta}\mathcal{E}_{10i}(\widehat{\boldsymbol{\theta}}^*)(\mathbf{x}_i^{\top}\widehat{\boldsymbol{\beta}})+\widehat{\Delta}^2\mathcal{E}_{20i}(\widehat{\boldsymbol{\theta}}^*)\right]\right).\end{aligned}$$

Now, returning to our original parameterization, we find the Hessian matrix for the original parameter vector $\boldsymbol{\theta}$,

$$\ddot{Q}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)=\mathrm{J}\left(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\theta}}\right)\ddot{Q}\left(\widehat{\boldsymbol{\theta}}^*|\widehat{\boldsymbol{\theta}}^*\right)\mathrm{J}\left(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\theta}}\right)^{\top}, \tag{24}$$

where $\mathrm{J}\left(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\theta}}\right)$ is the Jacobian matrix of order $(p+2)\times(p+2)$, defined by:

$$\mathrm{J}\left(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\theta}}\right)=\frac{\partial\boldsymbol{\theta}^*}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}=\begin{pmatrix}\mathbf{I}_p & \mathbf{0}_p & \mathbf{0}_p\\ \mathbf{0}_p^{\top} & \frac{\widehat{\lambda}}{2\widehat{\sigma}\sqrt{\widehat{\lambda}^2+1}} & \frac{1}{\widehat{\lambda}^2+1}\\ \mathbf{0}_p^{\top} & \frac{\widehat{\sigma}}{(\widehat{\lambda}^2+1)^{3/2}} & -\frac{2\widehat{\lambda}\widehat{\sigma}^2}{(\widehat{\lambda}^2+1)^2}\end{pmatrix},$$

where $\mathbf{I}_p$ represents the identity matrix of order $p\times p$ and $\mathbf{0}_p$ is a zero $p\times 1$ vector.

3.2 Local Influence

In this section, we derive the normal curvature of the local influence on the basis of the Q-function previously determined for some common perturbation schemes, either in the model or in the data. Thus, consider a perturbation vector $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_g)^{\top}$ varying in an open region $\boldsymbol{\Omega}\subset\mathbb{R}^g$. Let

$\ell_{comp}(\boldsymbol{\theta}|\mathbf{Y}_{comp},\boldsymbol{\omega})$ be the complete-data log-likelihood of the perturbed model. We assume there is a $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ such that $\ell_{comp}(\boldsymbol{\theta}|\mathbf{Y}_{comp},\boldsymbol{\omega}_0) = \ell_{comp}(\boldsymbol{\theta}|\mathbf{Y}_{comp})$ for all $\boldsymbol{\theta}$. Let us define

$$\begin{aligned} Q_{\boldsymbol{\omega}}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right) &= \mathrm{E}\left[\ell_{comp}(\boldsymbol{\theta}|\mathbf{Y}_{comp},\boldsymbol{\omega})\,|\mathbf{V},\boldsymbol{\rho},\widehat{\boldsymbol{\theta}}\right] \quad \text{and} \\ \widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}) &= \arg\max_{\boldsymbol{\theta}}\left\{Q_{\boldsymbol{\omega}}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)\right\} = \left(\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})^\top, \widehat{\sigma^2}(\boldsymbol{\omega}), \widehat{\lambda}(\boldsymbol{\omega})\right)^\top. \end{aligned}$$

The influence graph is then defined as $\boldsymbol{\alpha}(\boldsymbol{\omega}) = \left(\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega})\right)^\top$, where $f_Q(\boldsymbol{\omega})$ is the Q-displacement Function, defined as follows:

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

Following the approach of Cook (1986) and Zhu & Lee (2001), the normal curvature $C_{f_Q,\mathbf{d}}$ of $\boldsymbol{\alpha}(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\mathbf{d}$ can be used to summarize the local behavior of the Q-displacement function. Let

$$\boldsymbol{\nabla}_{\boldsymbol{\theta},\boldsymbol{\omega}} = \left\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}Q_{\boldsymbol{\omega}}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)\right\}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \text{ and } \ddot{Q}_{\boldsymbol{\omega}_0} = \left\{\frac{\partial^2}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^\top}Q\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}}\right)\right\}|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

Then, it can be shown that

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top\ddot{Q}_{\boldsymbol{\omega}_0}\mathbf{d} = 2\mathbf{d}^\top\boldsymbol{\nabla}^\top_{\boldsymbol{\theta},\boldsymbol{\omega}_0}\left\{-\ddot{Q}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)\right\}^{-1}\boldsymbol{\nabla}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}\mathbf{d},$$

where $\ddot{Q}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)$ is as defined in (24).

Following the same procedure adopted by Cook (1986), the information provided by the symmetric matrix $-\ddot{Q}_{\boldsymbol{\omega}_0}$ is quite useful for detecting influential observations. First, we consider the spectral decomposition

$$-2\ddot{Q}_{\boldsymbol{\omega}_0} = \sum_{k=1}^{g}\zeta_k\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^\top,$$

where $\{(\zeta_k,\boldsymbol{\varepsilon}_k), k=1,\ldots,g\}$ are eigenvalue–eigenvector pairs of $-2\ddot{Q}_{\boldsymbol{\omega}_0}$ with $\zeta_1 \geq \ldots \geq \zeta_r > \zeta_{r+1} = \ldots = 0$ and orthonormal eigenvectors $\boldsymbol{\varepsilon}_k$, for $k=1,\ldots,g$. Zhu & Lee (2001) proposed to inspect all eigenvectors corresponding to nonzero eigenvalues to capture more information, according to the following method:

$$\tilde{\zeta}_k = \frac{\zeta_k}{\zeta_1+\ldots+\zeta_r},\ \boldsymbol{\varepsilon}_k^2 = (\varepsilon_{k1}^2,\ldots,\varepsilon_{kg}^2)^\top \text{ and } M(0) = \sum_{k=1}^{r}\tilde{\zeta}_k\boldsymbol{\varepsilon}_k^2.$$

Let $M(0)_l = \sum_{k=1}^{r}\tilde{\zeta}_k\varepsilon_{kl}^2$ be the $l-$th component of $M(0)$. The assessment of influential cases is based on visual inspection of $M(0)_l, l=1,\ldots,g$ plotted against the index l. The $l-$th case may be regarded as influential if $M(0)_l$ is larger than a specified benchmark.

There is some inconvenience when using the normal curvature to decide about the influence of the observations, since $C_{f_Q,\mathbf{d}}$ may assume any value and it is not invariant under a uniform change of scale. Based on the work of Poon & Poon (1999), Zhu & Lee (2001) considered using the following conformal normal curvature:

$$B_{f_Q,\mathbf{d}} = \frac{C_{f_Q,\mathbf{d}}}{\mathrm{tr}\left[-2\ddot{Q}_{\boldsymbol{\omega}_0}\right]},$$

whose computation is quite simple and also has the property that $0 \leq B_{f_Q,\mathbf{d}} \leq 1$. Let $\mathbf{d}_l$ be a basic perturbation vector with l-th entry equal to 1 and all other entries equal to 0. Zhu & Lee (2001) showed that $M(0)_l = B_{f_Q,\mathbf{d}_l}$ for all l. We can therefore obtain $M(0)_l$ via $B_{f_Q,\mathbf{d}_l}$.

So far, there is no general rule to judge how large the influence of a given case is. Let $\overline{M(0)}$ and $SM(0)$ denote, respectively, the mean and the standard error of $\{M(0)_l;\ l = 1,\ldots,g\}$. Using the fact that the vectors $\boldsymbol{\varepsilon}_k$ are orthonormal, it is easy to prove that $\overline{M(0)} = 1/g$. Poon & Poon (1999) proposed to use $2\overline{M(0)}$ as a benchmark for $M(0)$. However, one may use different functions of $M(0)$. For instance, Zhu & Lee (2001) proposed using $\overline{M(0)} + 2SM(0)$ as a benchmark to take into account the variance of $\{M(0)_l;\ l = 1,\ldots,g\}$. According to Lee & Xu (2004), the exact choice of the function of $\overline{M(0)}$ as the benchmark is subjective. For example, they proposed using $\overline{M(0)} + c^* SM(0)$, where c^* is a selected constant, and depending on the application, c^* may be taken to be any value.

3.3 Perturbation schemes

We will evaluate the matrix $\boldsymbol{\nabla}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ under the following perturbation schemes for the SMSN-CR model: *case-weight perturbation* to detect observations with outstanding contribution of the log-likelihood function and that can exercise high influence on the maximum likelihood estimates; *response perturbation* of the response values, which can indicate observations with large influence on their own predicted values; *scale perturbation* of σ^2, which can reveal individuals that are most influential, in the sense of the likelihood displacement on the scale structure; and finally *explanatory variables perturbation*.

For each perturbation scheme, we have the partitioned form:

$$\boldsymbol{\nabla}_{\boldsymbol{\theta},\boldsymbol{\omega}_0} = \left(\boldsymbol{\nabla}^\top_{\beta,\boldsymbol{\omega}_0}, \boldsymbol{\nabla}^\top_{\sigma^2,\boldsymbol{\omega}_0}, \boldsymbol{\nabla}^\top_{\lambda,\boldsymbol{\omega}_0}\right)^\top,$$

where

$$\begin{aligned}
\boldsymbol{\nabla}_{\beta,\boldsymbol{\omega}_0} &= \left\{\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top} Q_{\boldsymbol{\omega}}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)\right\}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)} \in \mathbb{R}^{p\times g},\\
\boldsymbol{\nabla}_{\sigma^2,\boldsymbol{\omega}_0} &= \left\{\frac{\partial^2}{\partial\sigma^2\partial\boldsymbol{\omega}^\top} Q_{\boldsymbol{\omega}}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)\right\}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)} \in \mathbb{R}^{1\times g},\\
\boldsymbol{\nabla}_{\lambda,\boldsymbol{\omega}_0} &= \left\{\frac{\partial^2}{\partial\lambda\partial\boldsymbol{\omega}^\top} Q_{\boldsymbol{\omega}}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right)\right\}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)} \in \mathbb{R}^{1\times g}.
\end{aligned}$$

3.3.1 Case-weight perturbation

First, we consider an arbitrary attribution of weights to the expected value of the complete-data log-likelihood function (perturbed Q-function), which can capture departures in general directions, represented by writing:

$$Q_{\boldsymbol{\omega}}\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right) = \mathrm{E}\left[\ell_{comp}\left(\boldsymbol{\theta}|\mathbf{Y}_{comp},\boldsymbol{\omega}\right)|\mathbf{V},\boldsymbol{\rho},\widehat{\boldsymbol{\theta}}\right] = \sum_{i=1}^{n}\omega_i \mathrm{E}\left[\ell_{comp}\left(\boldsymbol{\theta}|Y_{comp_i}\right)|V_i,\rho_i,\widehat{\boldsymbol{\theta}}\right] = \sum_{i=1}^{n}\omega_i Q_i\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}\right).$$

Here, $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_n)^\top$ is an $n\times 1$ vector and $\boldsymbol{\omega}_0 = (1,\ldots,1)^\top$. Note that for $\omega_i = 0$ and $\omega_j = 1, j \neq i$, the $i-$th observation is dropped from the log-likelihood function for complete data. For this perturbation scheme, we find:

$$\begin{aligned}
\nabla_{\beta,\boldsymbol{\omega}_0} &= \frac{1}{\widehat{\tau}}\left[\mathbf{X}^\top \mathrm{Diag}\left\{\mathcal{E}_{01}(\widehat{\boldsymbol{\theta}})\right\} - \widehat{\mathbf{A}} - \widehat{\Delta}\mathbf{X}^\top \mathrm{Diag}\left\{\mathcal{E}_{10}(\widehat{\boldsymbol{\theta}})\right\}\right];\\
\nabla_{\sigma^2,\boldsymbol{\omega}_0} &= -\frac{1}{2\widehat{\sigma}^2}\left[\mathbf{1}_n^\top - \frac{1}{\widehat{\tau}}\widehat{\mathbf{B}}^\top + \frac{\widehat{\Delta}}{\widehat{\tau}}\widehat{\mathbf{C}}^\top\right];\\
\nabla_{\lambda,\boldsymbol{\omega}_0} &= \frac{\widehat{\lambda}}{\widehat{\lambda}^2+1}\mathbf{1}_n^\top - \frac{\widehat{\lambda}}{\widehat{\sigma}^2}\mathbf{B}^\top + \frac{2\widehat{\lambda}^2+1}{\widehat{\sigma}\sqrt{\widehat{\lambda}^2+1}}\mathbf{C}^\top - \widehat{\lambda}\mathcal{E}_{20}^\top(\widehat{\boldsymbol{\theta}}),
\end{aligned}$$

where $\widehat{\mathbf{A}}$ is a matrix with n columns equal to $\mathbf{X}^\top \mathrm{Diag}\left\{\mathcal{E}_{00}(\widehat{\boldsymbol{\theta}})\right\}\mathbf{X}\widehat{\boldsymbol{\beta}}$, $\mathcal{E}_{rs}(\widehat{\boldsymbol{\theta}}) = \left(\mathcal{E}_{rs1}(\widehat{\boldsymbol{\theta}}),\ldots,\mathcal{E}_{rsn}(\widehat{\boldsymbol{\theta}})\right)^\top$, $r,s=0,1,2$, $\mathbf{X}$ is a design matrix with rows $\mathbf{x}_i^\top$. $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{C}}$ are $n-$dimensional vectors with coordinates $\widehat{B}_i = \mathcal{E}_{02i}(\widehat{\boldsymbol{\theta}}) - 2\mathcal{E}_{01i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}} + \mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}})^2$ and $\widehat{C}_i = \mathcal{E}_{11i}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{10i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}$, respectively.

3.3.2 Scale perturbation

To study the effects of departures from the assumption regarding the scale parameter σ^2, we consider the perturbation $\sigma^2(\omega_i) = \omega_i^{-1}\sigma^2$, for $i=1,\ldots,n$. Under this perturbation scheme, the non-perturbed model is obtained when $\boldsymbol{\omega}_0 = \mathbf{1}_n$. Moreover, the perturbed Q-function is as in (13), with $\sigma^2(\omega_i)$ and $\widehat{\boldsymbol{\theta}}$ replacing σ^2 and $\boldsymbol{\theta}^{(k)}$, respectively. The matrix $\nabla_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ has the following elements:

$$\begin{aligned}
\nabla_{\beta,\boldsymbol{\omega}_0} &= \frac{1}{\widehat{\tau}}\left[\mathbf{X}^\top \mathrm{Diag}\left\{\mathcal{E}_{01}(\widehat{\boldsymbol{\theta}})\right\} - \widehat{\mathbf{A}} - \frac{\widehat{\Delta}}{2}\mathbf{X}^\top \mathrm{Diag}\left\{\mathcal{E}_{10}(\widehat{\boldsymbol{\theta}})\right\}\right];\\
\nabla_{\sigma^2,\boldsymbol{\omega}_0} &= \frac{1}{2\widehat{\sigma}^2}\left[\frac{1}{\widehat{\tau}}\widehat{\mathbf{B}}^\top - \frac{\widehat{\Delta}}{2\widehat{\tau}}\widehat{\mathbf{C}}^\top\right];\\
\nabla_{\lambda,\boldsymbol{\omega}_0} &= -\frac{\widehat{\lambda}}{\widehat{\sigma}^2}\widehat{\mathbf{B}}^\top + \frac{2\widehat{\lambda}^2+1}{2\widehat{\sigma}\sqrt{\widehat{\lambda}^2+1}}\widehat{\mathbf{C}}^\top.
\end{aligned}$$

3.3.3 Response perturbation

A perturbation of the response variables V_i, $i=1,\ldots,n$, can be introduced by replacing V_i by $V_i(\omega_i) = V_i + \omega_i\mathbf{S}_v$, where $\mathbf{S}_v$ is a scale factor that can represent the standard deviation of the censored response. Now substituting $V_i(\omega_i)$ into (8), we can write the perturbed model as:

$$V_i(\omega_i) = \begin{cases} c_i(\omega_i) & \text{if} \quad \rho_i = 1;\\ Y_i(\omega_i) & \text{if} \quad \rho_i = 0,\end{cases}$$

where $c_i(\omega_i) = c_i - \omega_i\mathbf{S}_v$ and $Y_i(\omega_i) = Y_i - \omega_i\mathbf{S}_v$. Hence, the perturbed Q-function follows (13), with $\mathcal{E}_{rsi}(\widehat{\boldsymbol{\theta}}^{(j)}) = \mathrm{E}[U_iT_i^rY_i^s|V_i,\rho_i,\widehat{\boldsymbol{\theta}}^{(j)}]$ replaced by $\mathcal{E}_{rsi}(\widehat{\boldsymbol{\theta}}^{(j)},\omega_i) = \mathrm{E}[U_iT_i^rY_i^s(\omega_i)|V_i(\omega_i),\rho_i,\widehat{\boldsymbol{\theta}}^{(j)}]$. Under this perturbation scheme, the vector $\boldsymbol{\omega}_0$, representing no perturbation, is given by $\boldsymbol{\omega}_0 = \mathbf{0}$

and $\boldsymbol{\nabla}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ has the following elements:

$$
\begin{aligned}
\boldsymbol{\nabla}_{\beta,\boldsymbol{\omega}_0} &= -\frac{\mathbf{S}_v}{\widehat{\tau}}\mathbf{X}^\top \mathrm{Diag}\left\{\mathcal{E}_{00}(\widehat{\boldsymbol{\theta}})\right\};\\
\boldsymbol{\nabla}_{\sigma^2,\boldsymbol{\omega}_0} &= -\frac{\mathbf{S}_v}{\widehat{\sigma}^2\widehat{\tau}}\left[\mathcal{E}_{01}^\top(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathrm{Diag}\left\{\mathcal{E}_{00}(\widehat{\boldsymbol{\theta}})\right\} - \frac{\widehat{\Delta}}{2}\mathcal{E}_{10}^\top(\widehat{\boldsymbol{\theta}})\right];\\
\boldsymbol{\nabla}_{\lambda,\boldsymbol{\omega}_0} &= \frac{2\widehat{\lambda}}{\widehat{\sigma}^2}\mathbf{S}_v\left[\mathcal{E}_{01}^\top(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathrm{Diag}\left\{\mathcal{E}_{00}(\widehat{\boldsymbol{\theta}})\right\}\right] - \frac{2\widehat{\lambda}^2+1}{\widehat{\sigma}(\widehat{\lambda}^2+1)^{1/2}}\mathbf{S}_v\mathcal{E}_{10}^\top(\widehat{\boldsymbol{\theta}}).
\end{aligned}
$$

3.3.4 Explanatory variables perturbation

Here, we consider the influence that perturbation of the explanatory variables can produce on the parameter estimates. In this case, we are interested in perturbing a specific explanatory variable, thus we consider the perturbation $\mathbf{x}_{i\omega}^\top = \mathbf{x}_i^\top + \omega_i\mathbf{S}_t\mathbb{1}_t^\top$, $\mathbf{S}_t$ is a scale factor that can represent the standard deviation of the t-th explanatory variable and $\mathbb{1}_t^\top = (0,\dots,1,\dots,0)$ is a $1\times p$ vector with 1 in the t-th column, $t = 1,\dots,p$. Hence, this case covers situations where x is measured with error. The perturbed Q-function is as in (13), switching $\mathbf{x}_{i\omega}^\top$ with $\mathbf{x}_i^\top$ and the no perturbation case is obtained by taking $\boldsymbol{\omega}_0 = \mathbf{0}$. Under this perturbation scheme, $\boldsymbol{\nabla}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ has the following elements:

$$
\begin{aligned}
\boldsymbol{\nabla}_{\beta,\boldsymbol{\omega}_0} &= \frac{\mathbf{S}_t}{\widehat{\tau}}\mathbb{1}_t\left[\mathcal{E}_{01}^\top(\widehat{\boldsymbol{\theta}}) - 2\widehat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathrm{Diag}\left\{\mathcal{E}_{00}(\widehat{\boldsymbol{\theta}})\right\} - \widehat{\Delta}\mathcal{E}_{10}^\top(\widehat{\boldsymbol{\theta}})\right];\\
\boldsymbol{\nabla}_{\sigma^2,\boldsymbol{\omega}_0} &= -\frac{\mathbf{S}_t}{\widehat{\sigma}^2\widehat{\tau}}\mathbb{1}_t^\top\widehat{\boldsymbol{\beta}}\left[\mathcal{E}_{01}^\top(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathrm{Diag}\left\{\mathcal{E}_{00}(\widehat{\boldsymbol{\theta}})\right\} - \frac{\widehat{\Delta}}{2}\mathcal{E}_{10}^\top(\widehat{\boldsymbol{\theta}})\right];\\
\boldsymbol{\nabla}_{\lambda,\boldsymbol{\omega}_0} &= \frac{\mathbf{S}_t}{\widehat{\sigma}}\mathbb{1}_t^\top\widehat{\boldsymbol{\beta}}\left[\frac{2\widehat{\lambda}}{\widehat{\sigma}}\left(\mathcal{E}_{01}^\top(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathrm{Diag}\left\{\mathcal{E}_{00}(\widehat{\boldsymbol{\theta}})\right\}\right) - \frac{2\widehat{\lambda}^2+1}{(\widehat{\lambda}^2+1)^{1/2}}\mathcal{E}_{10}^\top(\widehat{\boldsymbol{\theta}})\right].
\end{aligned}
$$

4 Application: Stellar abundances dataset

In this section, we use a censored dataset from stellar astronomy, previously analyzed by Santos *et al.* (2002), where the authors seek differences in the abundance of the light element beryllium (Be) in stars that do and do not host extrasolar planetary systems. These data are available in the **R** package **astrodatR**, under the name *Stellar abundances*. The dataset consists of 68 solar-type stars where the response variable is logN(Be), representing a measure of beryllium abundance with respect to the Sun's abundance.

According to Feigelson & Babu (2012) in a supervised astronomical survey where a particular property of a previously defined sample of objects is sought, some objects in the sample may be too faint to detect. Thus, the dataset contains the full sample of interest, but some objects have upper limits and others have detections. In this dataset we have 12 left-censored data points (see Figure 1, panel b). The predictor variable (x) is effective stellar surface temperature (in Kelvin, $\mathrm{T}_{eff}/1000$). To verify the existence of skewness in the data, Figure 1 (panel a) presents the histogram of the response variable and shows an apparent non-normal pattern. This figure also presents the normal quantile-quantile (Q-Q) plot for the residuals (panel c) obtained by fitting a Gaussian regression model using the **R** package **stats**. The Q-Q plot exhibits an asymmetrical heavy-tailed behavior, suggesting that the normality assumption for the errors might be inappropriate. In addition, Figure 1 (panel b) also indicates it is plausible to use a linear regression for the dataset.

Thus, in this section, we analyzed the *Stellar abundances* dataset with the aim of providing additional inferences by using SMSN distributions in the context of censored models.

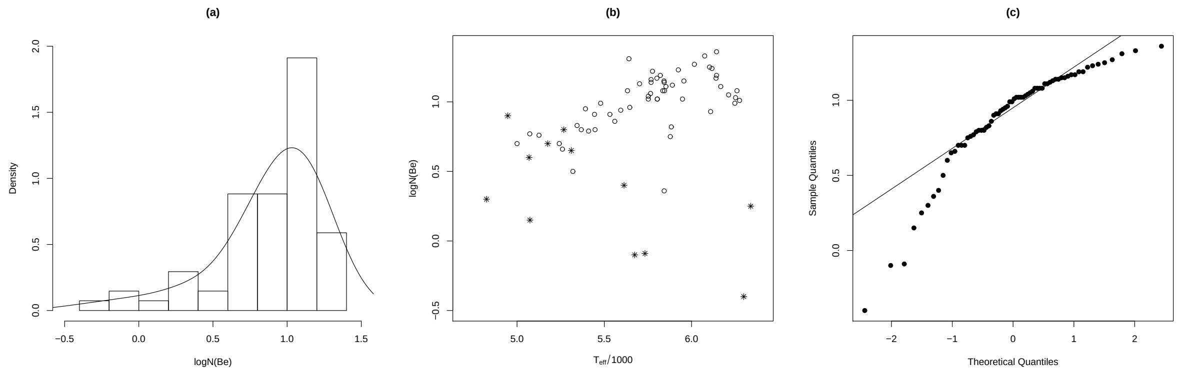

Figure 1: Stellar abundances dataset. (a) Histogram of the logN(Be). (b) Scatter-plot of the dataset. $(*)$ represents the censored observations. (c) Normal Q-Q plot for model residuals obtained by using the **R** package **stats**.

4.1 ML estimates using SAEM algorithm

We fitted a regression model with an intercept parameter β_0 and applied the SAEM algorithm for censored data, as described in Sub-section 2.3 (see also Mattos *et al.*, 2015). We focus on the SN-CR, ST-CR, SCN-CR and SSL-CR distributions from the SMSN-CR class.

Table 1 contains the ML estimates for the parameters of the four models, together with the values of their corresponding standard errors (SE). The log-likelihood values (see column $\ell(\widehat{\theta})$) indicate that the heavy-tailed SMSN distributions have a significantly better fit than the SN model. This finding is also confirmed by inspecting some model selection criteria, say the Akaike information criterion (AIC) (Akaike, 1974), the Bayesian information criterion (BIC) (Schwarz, 1978) and the efficient determination criterion (EDC) (Bai *et al.*, 1989). Moreover, the SE values of the ST-CR, SCN-CR and SSL-CR models are smaller than that of the SN-CR model. These results indicate that the use of the SMSN-CR models with heavy tails produces more accurate estimates.

In order to identify atypical observations and/or model misspecification, we use the martingale-type residuals, r_{MT_i}, proposed by Barros *et al.* (2010) (see also Garay *et al.*, 2015; Mattos *et al.*, 2015) for censored models. These residuals are defined by:

$$r_{MT_i} = \mathrm{sign}(r_{M_i})\sqrt{-2[r_{M_i} + \delta_i \log(\delta_i - r_{M_i})]}, \quad i = 1, \ldots, n,$$

where $r_{M_i} = \delta_i + \log(S(y_i; \widehat{\boldsymbol{\theta}}))$ is the martingale residual, with $\delta_i = 0, 1$ indicating whether the observation is censored or not, respectively, $S(y_i, \widehat{\boldsymbol{\theta}})$ is the SAEM estimate of the survival function of y.

The normal probability plot of the MT residuals with generated envelopes is presented in Figure 2. We observe that the ST-CR, SCN-CR and SSL-CR models fit the data better than the SN-CR model, since, in that case, there are fewer observations which lie outside the envelopes.

In Figure 3, we present the Mahalanobis distance, given by $d_i^2 = \frac{(y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}})^2}{\widehat{\sigma}^2}$ vs, the estimated weights $u_i = \mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}})$, for $i = 1, \ldots, 68$, considering the ST-CR, SSL-CR and SCN-CR models. We

Table 1: Stellar abundances dataset. Parameter estimates of the SMSN-CR models. SE values in parentheses.

Model	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\sigma}^2$	$\widehat{\lambda}$	ν	γ	$\ell(\widehat{\theta})$	AIC	BIC	EDC
SN-CR	-2.0399 (0.8974)	0.4944 (0.1549)	0.2942 (0.0426)	-7.7400 (3.5972)	-	-	-18.2141	44.42811	53.30614	43.02508
ST-CR	-2.2350 (0.4690)	0.5441 (0.0815)	0.0672 (0.0167)	-6.4338 (2.1758)	3	-	-2.1267	12.2535	21.1315	10.8505
SCN-CR	-2.2452 (0.5457)	0.5357 (0.0936)	0.0438 (0.0082)	-6.4700 (1.9214)	0.5	0.1	-3.7231	15.4462	24.3243	14.0432
SSL-CR	-2.2294 (0.4322)	0.5452 (0.0750)	0.0401 (0.0101)	-6.8774 (2.3953)	1.20	-	-2.7259	13.4518	22.3299	12.0488

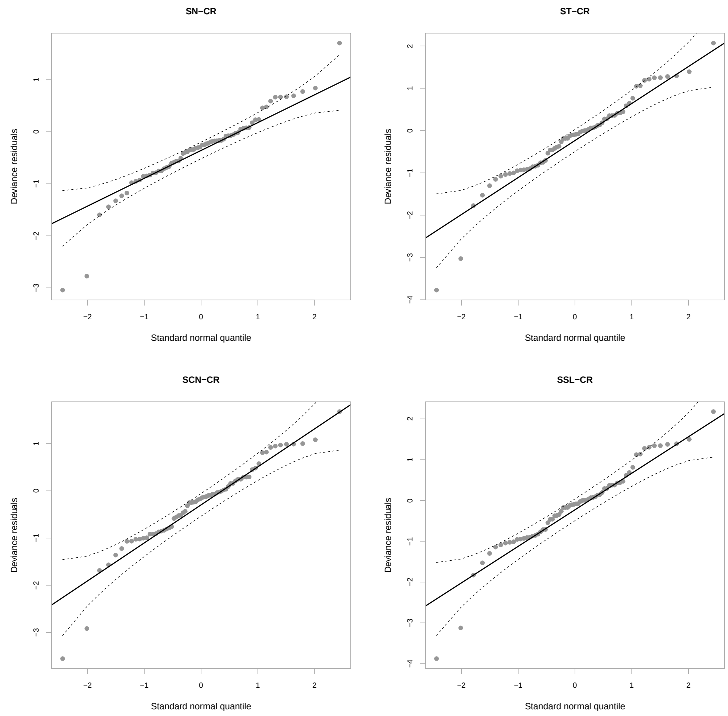

Figure 2: Stellar abundances dataset. Envelopes of the MT residuals for the SMSN-CR models

observe that when we use distributions with heavier tails than the SN one, the SAEM algorithm allows accommodating atypical observations by attributing small weights to them in the estimation procedure. The estimated weights for the SN-CR distribution are indicated as a continuous line. These results agree with similar considerations, presented for instance in Labra *et al.* (2012), where a nonlinear regression model under SMSN distributions is studied.

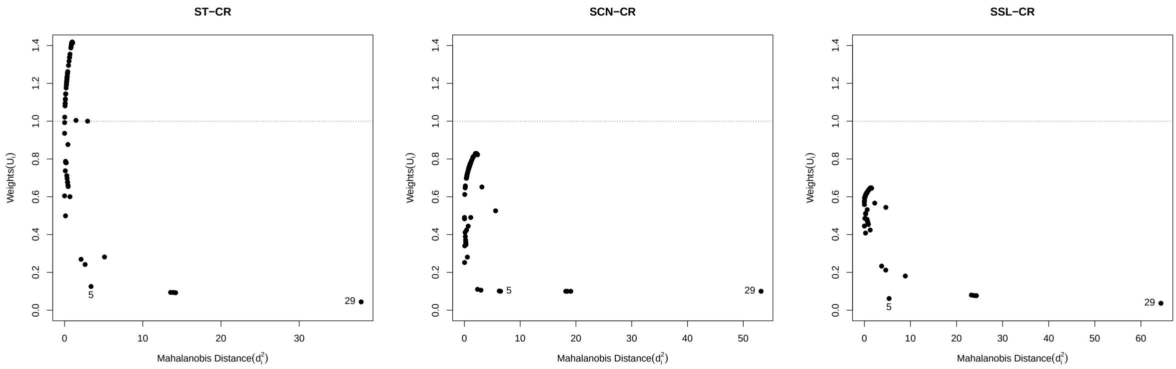

Figure 3: Stellar abundances dataset. Estimated u_i for the ST-CR, SCN-CR and SSL-CR models

4.2 Diagnostic Analysis

In this section, we compute case-deletion measures and analysis of local influence for the *Stellar abundances* dataset by using the SMSN-CR models.

4.2.1 Global Influence

In order to evaluate the effect on the ML estimates of the regression coefficients of the SMSN-CR models, when some observation is eliminated, we analyze the GD_i^1 and QD_i^1 plot in Figures 4-5. As can be seen, case # 5 is the most influential in the estimation of the parameters, for the four SMSN-CR models considered.

4.2.2 Local Influence

Next, we conduct a local influence study, with interest focused on $\boldsymbol{\theta}$, using the benchmark $M(0)$ from the conformal curvature $\mathbf{B}_{f_{Q,\mathbf{d}}}$, as described in Section 3.3 by considering the four different perturbation schemes. Thus, here we present a local influence analysis, using $c^* = 4$ to compute the benchmark $M(0)$.

From Figure 6, we observe that under the case-weight perturbation, case # 5 is identified as influential for the four SMSN-CR models considered. Under the scale perturbation, Figure 7, case # 29 appears influential in the ML estimates of $\boldsymbol{\theta}$ for the SN-CR, ST-CR and SCN-CR models. In addition, from Figure 8, this same observation # 29 is considered as influential for the SN-CR model, under the response variable perturbation. Finally, from Figure 9, we have that under explanatory variable perturbation no observations are considered potentially influential.

4.2.3 Impact of the detected influential observations

Table 2 shows that based in the global and local influence diagnostics, cases # 5 and # 29 are detected as potentially influential observations under different perturbation schemes. In order to assess the impact of these possible influential observations on the ML estimates, we refitted the model dropping each one of these cases. Thus, in Table 3 we present the relative changes (RC) of these estimates, $\mathrm{RC}(\widehat{\boldsymbol{\theta}}) = \left|\frac{(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{[j]})}{\widehat{\boldsymbol{\theta}}}\right|$, where $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2, \lambda)$ and $\widehat{\boldsymbol{\theta}}_{[j]}$ denotes the SAEM estimate of $\boldsymbol{\theta}$ with the $j-$th observation removed. We observe that observation # 5, detected as influential

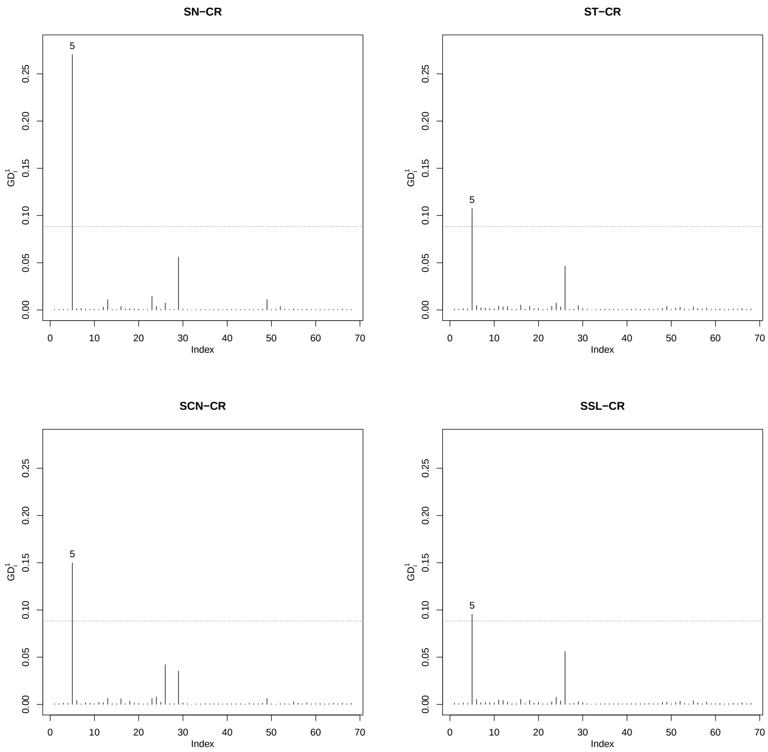

Figure 4: Stellar abundances dataset. Approximate generalized Cook's distance GD_i^1 for SMSN-CR models.

under global and local influence diagnostics, causes a significant change in the parameters σ^2 and λ, and observation # 29 causes significant changes (in particular) to the parameter σ^2.

Table 2: Stellar abundances dataset. Influential observations for SMSN-CR models

	Models			
Influence Measures	SN	ST	SSL	SCN
GD_i^1	# 5	# 5	# 5	# 5
QD_i^1	# 5	# 5	# 5	# 5
Case-weight perturbation	# 5	# 5	# 5	# 5
Scale perturbation	# 29	# 29	-	# 29
Response perturbation	# 29	-	-	-
Explanatory variable perturbation	-	-	-	-

5 Simulation study

In order to examine the performance of the proposed methods, we present a simulation study to show the capacity of the method to detect atypical data. We consider the SMSN-CR model, de-

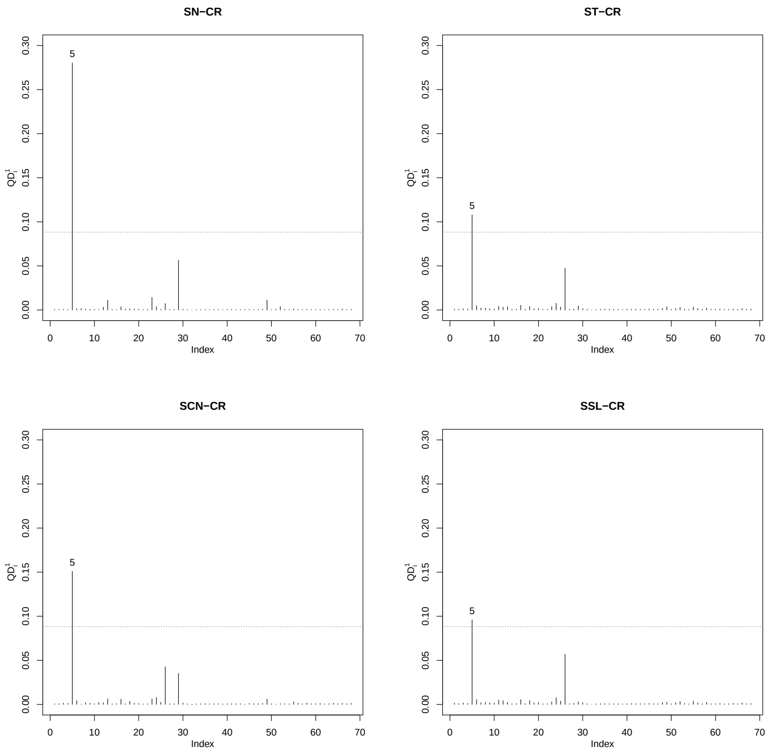

Figure 5: Stellar abundances dataset. Approximate likelihood displacement QD_i^1 for SMSN-CR models.

Table 3: Relative change for the dataset.

	SN-CR				ST-CR			
Dropping	$RC(\widehat{\beta}_0)$	$RC(\widehat{\beta}_1)$	$RC(\widehat{\sigma}^2)$	$RC(\widehat{\lambda})$	$RC(\widehat{\beta}_0)$	$RC(\widehat{\beta}_1)$	$RC(\widehat{\sigma}^2)$	$RC(\widehat{\lambda})$
[# 5]	0.0697	0.0551	0.2451	1.9120	0.0345	0.0208	0.0371	1.1948
[# 29]	0.0678	0.0669	0.2689	0.1575	0.0037	0.0034	0.1465	0.0763
[# 5, # 29]	0.1366	0.1199	0.4439	1.5015	0.0242	0.0214	0.1451	1.0246
	SCN-CR				**SSL-CR**			
Dropping	$RC(\widehat{\beta}_0)$	$RC(\widehat{\beta}_1)$	$RC(\widehat{\sigma}^2)$	$RC(\widehat{\lambda})$	$RC(\widehat{\beta}_0)$	$RC(\widehat{\beta}_1)$	$RC(\widehat{\sigma}^2)$	$RC(\widehat{\lambda})$
[# 5]	0.0347	0.0199	0.0117	1.6863	0.0389	0.0197	0.1673	1.0833
[# 29]	0.0064	0.0197	0.2729	0.1557	0.0045	0.0013	0.1161	0.0652
[# 5, # 29]	0.0303	0.0319	0.2539	1.5163	0.0261	0.0199	0.0680	0.8788

fined by combining equations (7)-(8), where $\boldsymbol{\beta}^\top = (\beta_0, \beta_1) = (3, -1)$, $\sigma^2 = 2$, $\lambda = 4$ and $\mathbf{x}_i^\top = (1, x_i)$. The values x_i, $i = 1, \ldots, 400$, were generated independently from a uniform distribution in the interval $(2, 5)$ and those values were kept constant throughout the experiment. The degree of freedom $(\boldsymbol{\nu})$ for the different cases of SMSN-CR models was fixed: $\nu = 3$ for the ST-CR and SSL-CR models and $(\nu, \gamma) = (0.1, 0.1)$ for the SCN-CR distribution. We generated 500 samples of size $n = 400$ from the SMSN-CR model, considering five censoring proportions,

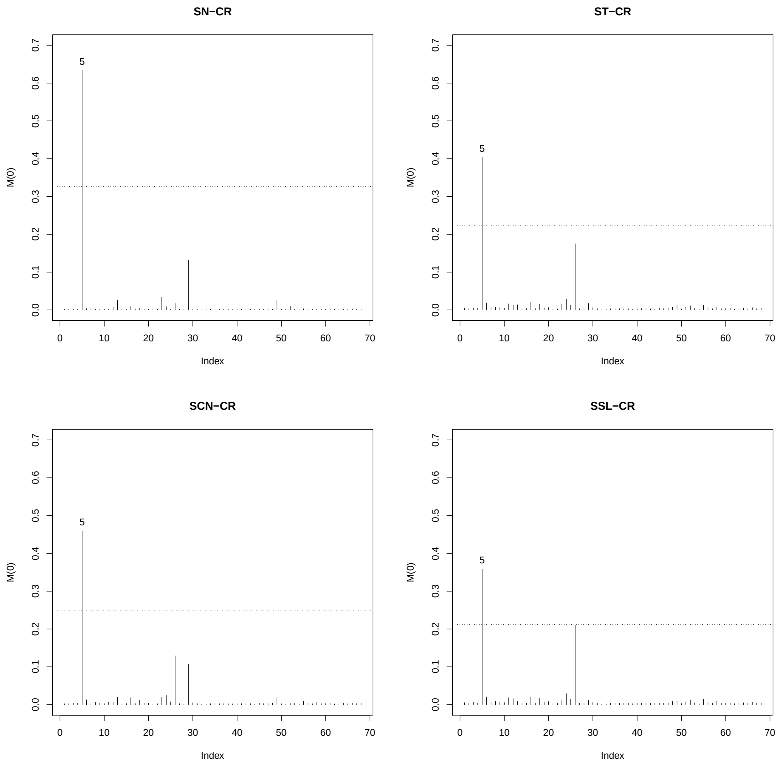

Figure 6: Stellar abundances dataset. Index plots of M(0) under the case-weight perturbation for SMSN-CR models.

$p = \{0\%,\ 10\%,\ 20\%,\ 30\%,\ \text{and}\ 40\%\}$. To guarantee the presence of a perturbed observation, we chose case # 200 and we replaced its parameters $\boldsymbol{\beta}$ by $\{2.5\boldsymbol{\beta},\ 5\boldsymbol{\beta},\ 7.5\boldsymbol{\beta},\ 10\boldsymbol{\beta}\}$. Considering the criteria $M(0)_i > \overline{M(0)} + 3SM(0)$ and GD_i^1 for $i = 1,\ldots,400$, to decide which point is influential or not. Table 4 shows, in percentage, the number of times that the measure correctly identified observation # 200 as the most influential. As expected, the percentage of correctly detecting atypical observations increases for increasing perturbation rates and we observe that in general, there is high sensitivity of the estimates in the presence of atypical data when the SN-CR model is considered.

6 Conclusion

In this work we presented a diagnostic analysis of linear regression models with censored responses and observational errors following a distribution belonging the class of SMSN distributions. This model was recently introduced by Mattos et al. (2015), where a full likelihood approach based on the SAEM algorithm was presented. The diagnostic analysis was based on the case-deletion and local influence techniques suggested by Zhu *et al.* (2001) and Zhu & Lee (2001), respectively, which are the counterparts for missing data models of the well-known ones proposed by Cook (1977) and Cook (1986). The structure of the complete-data likelihood function, obtained considering as if the missing data were in fact observed, is an essential element of the theory. Its simple form allows obtaining a tractable expression for the Q-function, which is essentially what is needed to provide an approximation of the maximum likelihood estimate of the parameters when an observation is excluded (for the case-deletion method). The same is true for the local influence method, in the

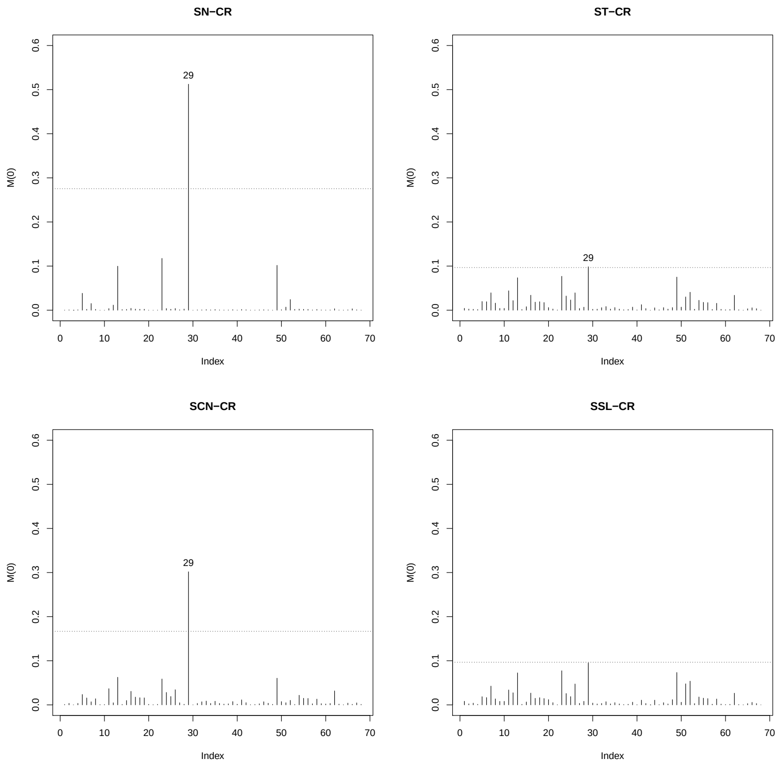

Figure 7: Stellar abundances dataset. Index plots of M(0) under scale perturbation for SMSN-CR models.

case of the normal curvature expressions. Using the developed method, we analyzed a real dataset and carried out extensive simulation studies. We observe, through influence diagnostic procedures, that when we used distributions with heavier tails than the SN-CR model, some aspects of robustness of the SAEM estimators under heavy-tail SMSN distributions were noted. The method is implemented using the R software (codes available upon request from the corresponding author), providing practitioners with a convenient tool for further applications in their domain.

The proposed methods can be extended to multivariate settings, such as the recent proposals of Matos *et al.* (2015) for censored mixed-effects models using the multivariate Student-t distribution. We intend to pursue this in future research. Due to the popularity of Markov chain Monte Carlo techniques, another potential work is to pursue a fully Bayesian treatment in this context for producing posterior inference. The method can also be extended to mixtures of regressions with skewed and heavy-tailed censored responses based on recent approaches by Caudill (2012) and Karlsson & Laitila (2014).

Acknowledgements

The research of Víctor H. Lachos was supported by Grant 305054/2011-2 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and by Grant 2014/02938-9 and 2013/21468-0 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil). Thalita B. Mattos was supported by CAPES.

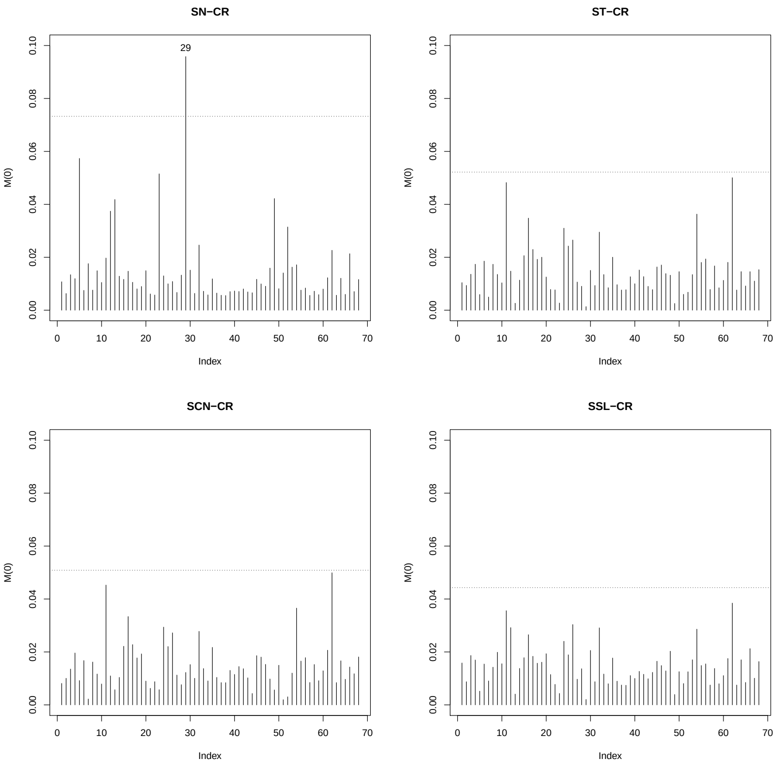

Figure 8: Stellar abundances dataset. Index plots of M(0) under response perturbation for SMSN-CR models.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Cont.*, **19**, 379–397.

Bai, Z. D., Krishnaiah, P. R. & Zhao, L. C. (1989). On rates of convergence of efficient detection criteria in signal processing with white noise. *IEEE Trans. Info. Theory*, **35**, 380–388.

Barros, M., Galea, M., González, M. & Leiva, V. (2010). Influence diagnostics in the Tobit censored response model. *Statistical Methods & Applications*, **19**, 716–723.

Basso, R. M., Lachos, V. H., Cabral, C. R. B. & Ghosh, P. (2010). Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics and Data Analysis*, **54**, 2926–2941.

Bolfarine, H., Montenegro, L. C. & Lachos, V. H. (2007). Influence diagnostics for skew-normal linear mixed models. *Sankhyā*, **69**, 648–670.

Caudill, S. B. (2012). A partially adaptive estimator for the censored regression model based on a mixture of normal distributions. *Statistical Methods & Applications*, **21**, 121–137.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15–18.

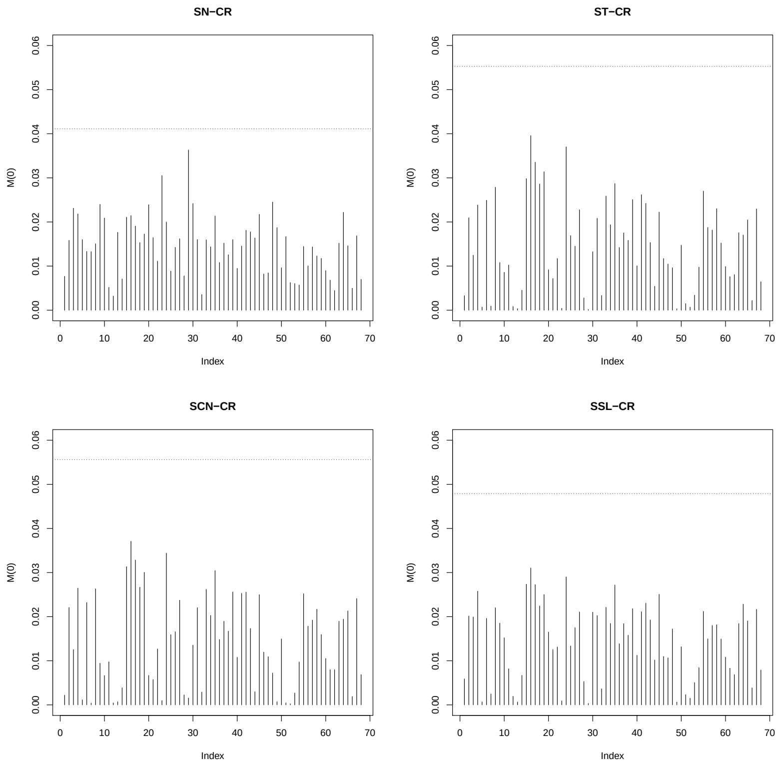

Figure 9: Stellar abundances dataset. Index plots of M(0) under explanatory variable perturbation for SMSN-CR models.

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

Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall/CRC.

Feigelson, E. D. & Babu, G. J. (2012). *Modern statistical methods for astronomy: with R applications*. Cambridge University Press.

Garay, A. M., Lachos, V. H., Bolfarine, H. & Cabral, C. R. B. (2015). Linear censored regression models with scale mixtures of normal distributions. *Statistical Papers*. DOI 10.1007/s00362-015-0696-9.

Karlsson, M. & Laitila, T. (2014). Finite mixture modeling of censored regression models. *Statistical Papers*, **55**, 627–642.

Kuhn, E. & Lavielle, M. (2004). Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: Probability and Statistics*, **8**, 115–131.

Labra, F. V., Garay, A. M., Lachos, V. H. & Ortega, E. M. M. (2012). Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of Statistical Planning and Inference*, **142**, 2149–2165.

Lachos, V. H., Ghosh, P. & Arellano-Valle, R. B. (2010). Likelihood based inference for skew–normal independent linear mixed models. *Statistica Sinica*, **20**, 303–322.

Table 4: Simulation study. % of correctly identified influential observation using GD_i^1 and under case-weight perturbation.

Influence Measures	Perturbation	SN-CR					ST-CR				
		Cens. Level					Cens. Level				
		0%	10%	20%	30%	40%	0%	10%	20%	30%	40%
GD_i	2.5β	19.8	15.8	15.8	14.8	14.0	4.6	5.6	4.0	3.4	0.8
	5β	64.2	63.2	60.8	64.8	60.6	39.2	35.4	36.4	30.8	20.8
	7.5β	89.8	86.0	86.2	87.8	85.6	70.6	68.6	70.2	62.0	62.2
	10β	96.4	98.6	96.6	98.4	96.6	84.4	83.6	80.8	87.2	82.4
Case-Weight perturbation	2.5β	15.0	14.4	16.0	17.0	16.0	14.0	18.6	19.6	19.6	20.4
	5β	62.0	62.6	61.2	65.6	63.8	50.0	56.0	56.4	59.6	56.2
	7.5β	88.4	85.8	87.4	88.0	87.6	77.8	78.8	79.0	80.4	83.2
	10β	96.2	98.6	96.8	98.4	97.0	88.8	88.4	88.0	92.6	90.2
		SCN-CR					**SSL-CR**				
Influence Measures	Perturbation	Cens. Level					Cens. Level				
		0%	10%	20%	30%	40%	0%	10%	20%	30%	40%
GD_i	2.5β	15.4	11.0	11.0	9.0	4.8	14.2	12.0	13.0	7.2	7.2
	5β	36.2	34.4	37.2	40.2	38.0	53.6	50.0	53.4	46.4	46.8
	7.5β	54.6	44.4	39.6	31.8	42.2	79.8	78.6	80.2	77.6	73.2
	10β	80.2	73.4	63.2	16.8	35.8	93.0	94.4	100	91.8	91.8
Case-Weight perturbation	2.5β	14.4	12.2	15.0	13.0	9.6	11.8	13.0	15.2	12.6	10.8
	5β	34.4	40.6	47.2	48.8	44.4	52.0	50.6	55.6	52.6	50.4
	7.5β	52.8	52.4	55.8	48.6	53.0	78.8	79.2	81.8	79.0	76.4
	10β	78.6	77.4	70.8	26.6	48.0	92.6	94.4	100	92.8	92.6

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.

Lange, K. L., Little, R. & Taylor, J. (1989). Robust statistical modeling using t distribution. *Journal of the American Statistical Association*, **84**, 881–896.

Lee, S. Y. & Xu, L. (2004). R influence analysis of nonlinear mixed-effects models. *Computational Statistics & Data Analysis*, **45**, 321–341.

Lucas, A. (1997). Robustness of the Student t based M-estimator. *Communications in Statistics, Theory and Methods*, **26**, 1165–1182.

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

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.

Matos, L. A., Bandyopadhyay, D., Castro, L. M. & Lachos, V. H. (2015). Influence assessment in censored mixed-effects models using the multivariate Student-t distribution. *Journal of multivariate analysis*, **141**, 104–117.

Mattos, T. B., Garay, A. M. & Lachos, V. (2015). Likelihood based inference for censored linear regression models with scale mixtures of skew-normal distributions. Technical report, Universidad Estadual de Campinas.

Meza, C., Osorio, F. & De la Cruz, R. (2012). Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statistics and Computing*, **22**, 121–139.

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

Santana, L., Vilca, F. & Leiva, V. (2011). Influence analysis in skew-Birnbaum-Saunders regression models and applications. *Journal of Applied Statistics*, **38**, 1633–1649.

Santos, N., López, R. G., Israelian, G., Mayor, M., Rebolo, R., García-Gil, A., de Taoro, M. P. & Randich, S. (2002). Beryllium abundances in stars hosting giant planets. *Astronomy & Astrophysics*, **386**(3), 1028–1038.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

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

Ying-Zi, F., Nian-Sheng, T. & Xing, C. (2009). Local influence analysis of nonlinear structural equation models with nonignorable missing outcomes from reproductive dispersion models. *Computational Statistics & Data Analysis*, **53**, 3671–3684.

Zeller, C. B., Labra, F. V., Lachos, V. H. & Balakrishnan, N. (2010). Influence analyses of skew-normal/independent linear mixed models. *Computational Statistics & Data Analysis*, **54**, 1266–1280.

Zeller, C. B., Lachos, V. H. & Vilca-Labra, F. E. (2011). Local influence analysis for regression models with scale mixtures of skew-normal distributions. *Journal of Applied Statistics*, **38**, 343–368.

Zhu, H. & Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B*, **63**, 111–126.

Zhu, H., Lee, S., Wei, B. & Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika*, **88**, 727–737.

Zhu, H., Ibrahim, J. G. & Shi, X. (2009). Diagnostic measures for generalized linear models with missing covariates. *Scandinavian Journal of Statistics*, **36**(4), 686–712.