Influence Diagnostics for Student-t Censored Linear Regression Models

Monique B. Massuia[a] Celso Rômulo Barbosa Cabral[b] Larissa A. Matos[a]
Víctor H. Lachos[a,*]
[a]*Departamento de Estatística, Universidade Estadual de Campinas, Brazil*
[b]*Departamento de Estatística, Universidade Federal do Amazonas, Brazil*

Abstract

In this paper we extend the censored linear regression model with normal errors to Student-t errors. A simple EM-type algorithm for iteratively computing maximum likelihood estimates of the parameters is presented. In order to examine the performance of the proposed model, case-deletion and local influence techniques are developed in order to show the robust aspect of it against outlying and influential observations. This is made by the analysis of the sensitivity of the EM estimates under some usual perturbation schemes in the model or data and by inspecting some proposed diagnostic graphics. The efficacy of the methodology is verified through the analysis of simulated data sets and modeling a real data set preliminary analyzed under normal errors. The proposed algorithm and methods are implemented in the R package `CensRegMod()`.

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

1 Introduction

The problem of estimation for a regression model where the dependent variable is censored has been studied in different fields, namely, econometric analysis, clinical essays, among many others. For example in econometrics, the study of the labor force participation of married women is usually conducted under the censored Tobit model (Greene, 2012). In this case, the observed response is the wage rate, which is typically considered as censored below zero, i.e., for working women, positive values for the wage rates are registered; whereas for the non-working women, the observed wage rates are zero (Mroz, 1987). In AIDS research, the viral load measures may be subjected 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 Wu, 2010).

In the framework of censored regression (CR) models, the random errors are routinely assumed to have a normal distribution for mathematical convenience (Wei & Tanner, 1990). However, it

*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

is well known that several phenomena are not always in agreement with the assumptions of the normal model, yielding data with a distribution with heavier tails. A good alternative is to consider that the errors have a Student-t distribution, so that the Student-t censored regression (t-CR) model is defined. See Fernández & Steel (1999) for a discussion about inference in this model.

As the classical normal model is very sensible to outlying observations, the assessment of robustness aspects of the parameter estimates is an important concern. The deletion methodology, 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 herein. 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 performing the 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 received considerable attention in the statistical literature in linear regression models. For a review, see Rancel & Sierra (2001). However, for the t-CR model the marginal log-likelihood function is complex for many models, 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 EM algorithm. 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), Lachos *et al.* (2011), Santana *et al.* (2011) and Matos *et al.* (2012), among others. Using this general methodology and also applying the methodology of Lee & Xu (2004), in this paper we develop a local influence approach for the t-CRM and show that it leads to simple influence measures.

The rest of the paper is organized as follows. The t-CR model is defined in Section 2. In Section 3, we develop an EM-type algorithm to proceed maximum likelihood estimation for the parameters of the proposed model. In section 4 we develop influence diagnostic techniques, based on case deletion and local influence approaches. Sections 5 and 6 are dedicated to the analysis of real and simulated data sets, respectively.

2 The t-CR model

2.1 The Student-t and truncated Student-t distribution

Before we talk about the censored regression model, for the sake of completeness, we will give a brief introduction on the truncated Student-t distribution. In the following definitions, $\mathrm{N}(\mu,\sigma^2)$ denotes the normal distribution with mean μ and variance σ^2, $\mathrm{Gamma}(c,d)$ denotes the gamma distribution with mean c/d and variance c/d^2 and $Z \perp U$ means that the random variables Z and U are independent. Also, $\overset{d}{=}$ means "has the same distribution as". First, we give the classical definition of the Student-t distribution as a scale mixture of the normal distribution.

Definition 1. *We say that a random variable X has a Student-t distribution with location parameter μ, scale parameter σ^2 and ν degrees of freedom, denoted by $X \sim \mathrm{t}_\nu(\mu,\sigma^2)$, if*

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

where $Z \sim \mathrm{N}(0,\sigma^2)$*,* $U \sim \mathrm{Gamma}(\nu/2,\nu/2)$ *and* $Z \perp U$*.*

Definition 2. *Let* $X \sim \mathrm{t}_\nu(\mu,\sigma^2)$*. A random variable* Y *has a truncated Student-t distribution in the interval* (a,b) *if* $Y \stackrel{d}{=} X|(X \in (a,b))$*. In this case we will write* $Y \sim \mathrm{Tt}_\nu(\mu,\sigma^2;(a,b))$*.*

To provide a full specification of the distribution of Y, we say that it has ν degrees of freedom and that μ and σ^2 are the parameters *before truncation*. Also, it is allowed $a = -\infty$ or $b = \infty$. As an obvious consequence of this definition, we can obtain the density of $Y \sim \mathrm{Tt}_\nu(\mu,\sigma^2;(a,b))$, given in the next result. Let us denote the density of the Student-t distribution of Definition 1 by $\mathrm{t}_\nu(\cdot|\mu,\sigma^2)$ and the distribution function of the standard Student-t distribution with ν degrees of freedom (that is, with $\mu = 0$ and $\sigma^2 = 1$) by $\mathscr{T}_\nu(\cdot)$.

Proposition 1. *Let* $Y \sim \mathrm{Tt}_\nu(\mu,\sigma^2;(a,b))$*. Then the density of* Y *is*

$$\mathrm{Tt}_\nu(y|\mu,\sigma^2;(a,b)) = \mathrm{t}_\nu(y|\mu,\sigma^2)\left[\mathscr{T}_\nu\left(\frac{b-\mu}{\sigma}\right) - \mathscr{T}_\nu\left(\frac{a-\mu}{\sigma}\right)\right]^{-1}, \quad y \in \mathbb{R}.$$

The following result, provided by Kim (2008), presents the first two moments of the truncated Student-t, being very important for our subsequent exposition. $\Gamma(\cdot)$ denotes the gamma function.

Lemma 1. *If* $Y \sim \mathrm{Tt}_\nu(\mu,\sigma^2;(a,b))$*, then*

$$\begin{aligned}
\mathrm{E}[Y] &= \mu + G(\nu)\left\{(\nu+\alpha^2)^{-(\nu-1)/2} - (\nu+\beta^2)^{-(\nu-1)/2}\right\}\sigma,\ \nu > 1,\\
\mathrm{E}[Y^2] &= \mu^2 + \sigma^2\left\{\frac{\nu}{\nu-2} + G(\nu)\left[\alpha(\nu+\alpha^2)^{-(\nu-1)/2} - \beta(\nu+\beta^2)^{-(\nu-1)/2}\right]\right\}\\
&\quad +2\mu\sigma G(\nu)\left\{(\nu+\alpha^2)^{-(\nu-1)/2} - (\nu+\beta^2)^{-(\nu-1)/2}\right\}, \nu > 2,
\end{aligned}$$

where $G(\nu) = \dfrac{\Gamma((\nu-1)/2)\nu^{\nu/2}}{2(\mathscr{T}_\nu(\beta) - \mathscr{T}_\nu(\alpha))\Gamma(\nu/2)\Gamma(1/2)}$, $\alpha = \frac{a-\mu}{\sigma}$ *and* $\beta = \frac{b-\mu}{\sigma}$.

The following result will be useful for the implementation of the EM algorithm, see Section 3.

Lemma 2. *Let* $Y \sim \mathrm{Tt}_\nu(\mu,\sigma^2;(a,b))$*,* $d^2(\mu,\sigma^2,Y) = (Y-\mu)^2/\sigma^2$ *and* $r > 0$*. Then, for* $k = 0,1,2$*,*

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\nu+1}{\nu+d^2(\mu,\sigma^2,Y)}\right)^r Y^k\right] &= c(\nu,r)\mathrm{E}[X^k]\left[\mathscr{T}_{\nu+2r}\left(\frac{b-\mu}{\sigma^*}\right) - \mathscr{T}_{\nu+2r}\left(\frac{a-\mu}{\sigma^*}\right)\right]\\
&\quad \times\left[\mathscr{T}_\nu\left(\frac{b-\mu}{\sigma}\right) - \mathscr{T}_\nu\left(\frac{a-\mu}{\sigma}\right)\right]^{-1},
\end{aligned}$$

where

$$X \sim \mathrm{Tt}_{\nu+2r}(\mu,\sigma^{*2};(a,b)), \quad \textit{with} \quad \sigma^{*2} = \frac{\nu}{(\nu+2r)}\sigma^2,$$

and

$$c(\nu,r) = \left(\frac{\nu+1}{\nu}\right)^r \frac{\Gamma((\nu+1)/2)\Gamma((\nu+2r)/2)}{\Gamma(\nu/2)\Gamma((\nu+2r+1)/2)}.$$

Proof. We have that

$$\left(\frac{\nu+1}{\nu+d^2(\mu,\sigma^2,Y)}\right)^r \mathrm{t}_\nu(x|\mu,\sigma^2) = c(\nu,r)\mathrm{t}_{\nu+2r}(x|\mu,\sigma^{*2}),$$

which implies

$$\mathrm{E}\left[\left(\frac{\nu+1}{\nu+d^2(\mu,\sigma^2,Y)}\right)^r Y^k\right] = \frac{c(\nu,r)P(W \in (a,b))}{P(Z \in (a,b))}\int_{(a,b)} \frac{1}{P(W \in (a,b))} w^k \mathrm{t}_{\nu+2r}(w|\mu,\sigma^{*2})dw,$$

where $Z \sim \mathrm{t}_\nu(\mu,\sigma)$ and $W \sim \mathrm{t}_{\nu+2r}(\mu,\sigma^{*2})$, and the result follows. □

2.2 The Student-t censored linear regression model

Consider first a linear regression model where the responses are observed with errors which are independent and identically distributed according to a Student-t distribution. To be more precise, let us write

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \;\; \varepsilon_i \sim \mathrm{t}_\nu(0, \sigma^2), \;\; i = 1, \ldots, n, \tag{2}$$

where Y_i is the response for subject i, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of regression parameters and $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{ip})$ is a vector of explanatory variable values. By definition 1, we have that $Y_i \sim \mathrm{t}_\nu(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2)$ for $i = 1, \ldots, n$. We will call (2) as the *t-R model*. Estimation and diagnostic analysis for this model have been widely discussed in the literature (see, for instance, Cysneiros & Paula, 2005).

We are interested in the case were right-censored observations can occur. That is, the observations are of the form

$$Y_{\mathrm{obs}_i} = \begin{cases} \kappa_i & \text{if} \quad Y_i \geq \kappa_i; \\ Y_i & \text{if} \quad Y_i < \kappa_i, \end{cases} \tag{3}$$

$i = 1, \ldots, n$, for some threshold point κ_i. This will be called the *t-CR model*. The interpretation is that the observation i is *right censored* when the sampling period finishes before an event of interest is observed. We have chosen to work with the right censored case, which is the most common in applications, but the results are easily extensible to another censoring types. Note that when $\kappa_i = 0$, $i = 1, \ldots, n$, the proposed t-CR model (2)-(3) is reduced to the Tobit model considered by Arellano-Valle *et al.* (2012) in which an interesting EM algorithm is developed to obtain maximum likelihood estimates.

3 Parameter estimation via the EM-algorithm for the t-CR Model

3.1 A note about fixed degrees of freedom

For the t-CR model given in equations (2) and (3), we will assume that the degrees of freedom ν are fixed and, obviously, some theoretical basis is needed to justify this choice. In this direction, the work of Fernández & Steel (1999) is crucial. They discussed potential problems that may arise in the estimation of the degrees of freedom, in particular for the Student-*t* distribution. This is due to the apparent unboundedness of the likelihood function near the boundary of the parameter space, and hence the maximum likelihood scheme as developed in Lange & Sinsheimer (1993) is questionable because it does not provide sufficient information on whether these estimates correspond to local or global maximas. Interestingly, Lucas (1997) notes that only under the assumption of fixed degrees of freedom, the parameter estimates behave robustly against extreme observations. A plausible (and simple) alternative is to assume that the parameter ν associated with the scale variable U is known, which has been adopted in this work. Recent works in the context of elliptical distributions have also considered the scale parameter to be known. See for instance, Vanegas & Cysneiros (2010) and Meza *et al.* (2012).

3.2 An EM-type Algorithm

In what follows, in general, we use the traditional convention denoting a random variable by an upper case letter and its realization by the correspondent lower case. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$ be the vector with all parameters of the t-CR model. Supposing that are (possibly) m censored values of the characteristic of interest, we can partition the observed sample $\mathbf{y}_{\mathrm{obs}}$ in 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 is given by

$$\ell(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) = \sum_{i=1}^{m} \log\left[\mathscr{T}_\nu\left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta} - \kappa_i}{\sigma}\right)\right] + \sum_{i=m+1}^{n} \log \mathrm{t}_\nu(y_i|\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2).$$

To estimate the parameters of the t-CR model an alternative is to maximize this log-likelihood function directly, a procedure that can be quite cumbersome. Alternatively, our choice is to use the EM algorithm, a classical, reliable, widespread and general framework developed by Dempster *et al.* (1977) to obtain maximum likelihood estimates.

To apply the EM method, we need a representation of the model in terms of missing data. First, observe that, by Definition 1, if $Y_i \sim \mathrm{t}_\nu(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2)$ then

$$Y_i|U_i = u_i \sim \mathrm{N}(\mathbf{x}_i^\top \boldsymbol{\beta}, u_i^{-1}\sigma^2), \; U_i \sim \mathrm{Gamma}(\nu/2, \nu/2). \tag{4}$$

This relation is a convenient stochastic representation of the t-R model, and will be useful in the path E of the algorithm.

In the case of censoring, we can consider the unobserved y_i as a realization of the latent unobservable variable $Y_i \sim \mathrm{t}_\nu(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2)$, $i = 1, \ldots, m$. The key to the development of our EM-type algorithm is to consider the augmented data $\{\mathbf{y}_{\text{obs}}, y_1, \ldots, y_m, u_1, \ldots, u_n\}$, that is, we treat the problem as if $\mathbf{y}_L = (y_1, \ldots, y_m)^\top$ were in fact observed. As a consequence, we can use the representation (4) to obtain the complete-data log-likelihood, given as

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}, \mathbf{y}_L, \mathbf{u}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 + \frac{n}{2}\sum_{i=1}^{n}\log u_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 + \sum_{i=1}^{n}\log h(u_i|\nu),$$

where $\mathbf{u} = (u_1, \ldots, u_n)^\top$ and $h(\cdot|\nu)$ is the Gamma density with both parameters equal to $\nu/2$.

In what follows the superscript (k) indicates the estimate of the related parameter at the stage k of the algorithm. In the path E of the algorithm, we must obtain the so-called Q-function

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\ell_c(\boldsymbol{\theta}|\mathbf{Y}_{\text{obs}}, \mathbf{Y}_L, \mathbf{U})|\mathbf{y}_{\text{obs}}],$$

where $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}$ means that the expectation is being effected using $\boldsymbol{\theta}^{(k)}$ for $\boldsymbol{\theta}$. Observe that the expression of the Q-function is completely determined by the knowledge of the expectations

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

since $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log U_i|y_{\text{obs}_i}]$ and $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[\log h(U_i|\nu)|y_{\text{obs}_i}]$ depend only on ν, which is known. Thus, dropping unimportant constants, the Q-function can be written in a synthetic form as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\mathscr{E}_{2i}(\boldsymbol{\theta}^{(k)}) - 2\mathscr{E}_{1i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i^\top \boldsymbol{\beta} + \mathscr{E}_{0i}(\boldsymbol{\theta}^{(k)})(\mathbf{x}_i^\top \boldsymbol{\beta})^2\right]. \tag{5}$$

This is, undoubtedly, a computationally attractive and quite useful expression to implement the M-step, which consists in maximizing it over $\boldsymbol{\theta}$.

For an uncensored observation i, we have that $Y_{\text{obs}_i} = Y_i \sim \mathrm{t}_\nu(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2)$ and, therefore,

$$\mathscr{E}_{si}(\boldsymbol{\theta}^{(k)}) = y_i^s \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i|y_i], \quad \text{with} \quad \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i|y_i] = \frac{\nu+1}{\nu + d^2(\boldsymbol{\theta}^{(k)}, y_i)}. \tag{6}$$

This result can be obtained using (4) (see Appendix).

For a censored observation i we have that $Y_{\text{obs}_i} = \kappa_i$ iff $Y_i \geq \kappa_i$, such that

$$\begin{aligned}
\mathscr{E}_{si}(\boldsymbol{\theta}^{(k)}) &= \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i Y_i^s | Y_i \geq \kappa_i] = \int\int u_i y_i^s \pi(u_i|y_i, Y_i \geq \kappa_i)\pi(y_i|Y_i \geq \kappa_i) du_i dy_i \\
&= \int y_i^s \left[\int u_i \pi(u_i|y_i) du_i\right] \pi(y_i|Y_i \geq \kappa_i) dy_i \qquad (7)\\
&= \int \frac{(\nu+1)y_i^s}{\nu + d^2(\boldsymbol{\theta}^{(k)}, y_i)} \pi(y_i|Y_i \geq \kappa_i) dy_i \\
&= \mathrm{E}\left[\frac{(\nu+1)Y_i^s}{\nu + d^2(\boldsymbol{\theta}^{(k)}, Y_i)} | Y_i \geq \kappa_i\right], \qquad (8)
\end{aligned}$$

where, the somewhat imprecise but convenient notation $\pi(\cdot|y_i)$, etc. denotes the conditional density of a random variable in general. The equality in third line is true because, if y_i were available, then it would be a realization of a $\mathrm{t}_\nu(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2)$ distribution and, therefore, the inner integral in (7) is equal to the expectation $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i|y_i]$ in (6). Finally, the expectation in (8) can be easily obtained using Lemma 2, because the distribution of $Y_i|Y_i \geq \kappa_i$ is $\mathrm{Tt}_\nu(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2; (\kappa_i, \infty))$.

Our EM-type algorithm can be summarized in the following way.

E-step: Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$. For $i = 1, \ldots, n$.

- If the observation i is uncensored then, for $s = 0, 1, 2$, compute $\mathscr{E}_{si}(\boldsymbol{\theta}^{(k)})$ given in (6);
- If the observation i is censored then, for $s = 0, 1, 2$, compute $\mathscr{E}_{si}(\boldsymbol{\theta}^{(k)})$ in (8) using Lemma 2 with $r = 1$.

M-step: Update $\boldsymbol{\theta}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ over $\boldsymbol{\theta}$, which leads to the following expressions

$$\boldsymbol{\beta}^{(k+1)} = \left(\sum_{i=1}^n \mathscr{E}_{0i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1} \sum_{i=1}^n \mathbf{x}_i \mathscr{E}_{1i}(\boldsymbol{\theta}^{(k)}), \qquad (9)$$

$$\sigma^{2(k+1)} = \frac{1}{n}\sum_{i=1}^n \left[\mathscr{E}_{2i}(\boldsymbol{\theta}^{(k)}) - 2\mathscr{E}_{1i}(\boldsymbol{\theta}^{(k)})\mathbf{x}_i^\top \boldsymbol{\beta}^{(k+1)} + \mathscr{E}_{0i}(\boldsymbol{\theta}^{(k)})(\mathbf{x}_i^\top \boldsymbol{\beta}^{(k+1)})^2\right], \qquad (10)$$

This process is iterated until some distance involving two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta})$, like $||\ell(\boldsymbol{\theta}^{(k+1)}) - \ell(\boldsymbol{\theta}^{(k)})||$ or $||\ell(\boldsymbol{\theta}^{(k+1)})/\ell(\boldsymbol{\theta}^{(k)}) - 1||$, is small enough. This algorithm is implemented as part of the R package `CensRegMod` (Massuia *et al.*, 2012), which can be downloaded freely from the repository CRAN.

It is also important to assess the variability of the EM estimates. Our approach to achieve this is to use an information based method. In this context, the seminal paper of Louis (1982) is crucial, providing a general method to obtain the observed information matrix when the EM algorithm is used to find maximum likelihood estimates in a missing data framework. Combining this method with results of the Appendix B in Lange *et al.* (1989), an asymptotic approximation for the variance of the estimator of the regression parameters of the t-CR model is given by

$$\mathrm{Var}(\widehat{\beta}_i) = \left[\sum_{i=1}^n \frac{(\nu+1)}{(\nu+3)} \frac{\mathbf{x}_i^\top \mathbf{x}_i}{\sigma^2} - \sum_{i=1}^n \frac{\mathbf{x}_i^\top B_i \mathbf{x}_i}{\sigma^4}\right]^{-1}, \qquad (11)$$

where $B_i = \mathrm{Var}\left[\left(\frac{\nu+1}{\nu + d^2(\boldsymbol{\theta}, Y_i)}\right)(Y_i - \mathbf{x}_i^\top \boldsymbol{\beta})|y_{\text{obs}_i}\right]$, which can be easily obtained by using Lemma 2.

4 Diagnostic analysis

Influence diagnostic techniques consist of evaluate the sensitivity of the parameter estimates of a particular model when perturbation occurs either in the data set or in the underlying assumptions of the model. There are primarily two approaches for detecting 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 a comparison of parameter estimates before and after deletion of it. This is made by analyzing one or more fitted models after the exclusion and then assessing by some metrics such as the likelihood distance or the 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, and not its total deletion. By using the results of Subsection 4.1, we will introduce here the case-deletion measures and the local influence measures to the censored data on the basis of the Q-function previously determined. We first consider the case-deletion measures, then the local influence and finally the perturbation schemes used.

4.1 Case-deletion measures

Case-deletion is a classical approach to study the effects of dropping the ith case from the data set. In what follows, $\mathbf{y}_c = \{\mathbf{y}_{\text{obs}}, y_1, \ldots, y_m, u_1, \ldots, u_n\}$ denotes the augmented data set and a quantity with a subscript "$[i]$" denotes the original one with the ith case deleted. Then, $\mathbf{y}_{c[1]} = \{\mathbf{y}_{\text{obs}_{[1]}}, y_2, \ldots, y_m, u_2, \ldots, u_m\}$, with $\mathbf{y}_{\text{obs}_{[1]}} = \{y_{m+2}, \ldots, y_n\}$, and the complete-data log-likelihood function based on the data with the ith case deleted will be denoted by $\ell_c(\boldsymbol{\theta}|\mathbf{y}_{c[i]})$, for instance. Let $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}^{\top}_{[i]}, \widehat{\sigma^2}_{[i]})^{\top}$ be the maximizer of the function $Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}_{\widehat{\boldsymbol{\theta}}}\left[\ell_c(\boldsymbol{\theta}|\mathbf{Y}_{c[i]})|\mathbf{y}_{\text{obs}}\right]$, where $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^{\top}, \widehat{\sigma^2})^{\top}$ is the EM estimate of $\boldsymbol{\theta}$. To assess the influence of the ith case on $\widehat{\boldsymbol{\theta}}$, we compare the difference between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$. If the deletion of a case seriously influences the estimates, more attention need to be paid to that case. Hence, if $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the ith case is regarded as influential. As $\widehat{\boldsymbol{\theta}}_{[i]}$ is needed for every case, the required computational effort can be quite heavy, especially when the sample size is large. Hence, the following one-step pseudo approximation $\widehat{\boldsymbol{\theta}}^{1}_{[i]}$ is used to reduce the burden (see Cook & Weisberg, 1982):

$$\widehat{\boldsymbol{\theta}}^{1}_{[i]} = \widehat{\boldsymbol{\theta}} + \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \tag{12}$$

where

$$\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \text{ and } \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \tag{13}$$

are the Hessian matrix and the gradient vector evaluated at $\widehat{\boldsymbol{\theta}}$, respectively. In particular, the Hessian matrix is an essential element in the methodology developed by Zhu & Lee (2001) in order to obtain the measures for case-deletion diagnostic and for local influence of a specified perturbation scheme. These formulae can be obtained quite easily from relation (5). The latter has its coordinates as follows

$$\begin{aligned}
\dot{Q}_{[i]\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \frac{1}{\widehat{\sigma^2}}E_{1[i]}, \\
\dot{Q}_{[i]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \frac{\partial Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\frac{1}{2\widehat{\sigma^2}}E_{2[i]},
\end{aligned}$$

where

$$E_{1[i]} = \sum_{j\neq i}\left[\mathbf{x}_j\mathscr{E}_{1j}(\widehat{\boldsymbol{\theta}}) - \mathscr{E}_{0j}(\widehat{\boldsymbol{\theta}})\mathbf{x}_j\mathbf{x}_j^\top\widehat{\boldsymbol{\beta}}\right] \quad \text{and} \tag{14}$$

$$E_{2[i]} = \sum_{j\neq i}\left[1-\frac{1}{\widehat{\sigma^2}}\left(\mathscr{E}_{2j}(\widehat{\boldsymbol{\theta}}) - 2\mathscr{E}_{1j}(\boldsymbol{\theta})\mathbf{x}_j^\top\widehat{\boldsymbol{\beta}} + \mathscr{E}_{0j}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_j^\top\widehat{\boldsymbol{\beta}})^2\right)\right]. \tag{15}$$

The second order partial derivatives of $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ evaluated at $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\sigma^2})^\top$ are

$$\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\frac{1}{\widehat{\sigma^2}}\sum_{i=1}^n \mathscr{E}_{0i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i\mathbf{x}_i^\top, \tag{16}$$

$$\begin{aligned}\ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\} &= \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\sigma^2}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\ &= \frac{1}{2\widehat{\sigma^4}}\sum_{i=1}^n\left[1-\frac{2}{\widehat{\sigma^2}}\left(\mathscr{E}_{2i}(\widehat{\boldsymbol{\theta}}) - 2\mathscr{E}_{1i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}} + \mathscr{E}_{0i}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}})^2\right)\right],\end{aligned} \tag{17}$$

$$\ddot{Q}_{\beta\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\} = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma^2}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\frac{1}{\widehat{\sigma^4}}\sum_{i=1}^n\left[\mathbf{x}_i\mathscr{E}_{1i}(\widehat{\boldsymbol{\theta}}) - \mathscr{E}_{0i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}\right]$$

Using expressions (9) and (10), replacing $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\beta}^{(k)^\top}, \sigma^{2(k+1)})^\top$ with $\widehat{\boldsymbol{\theta}}$, we can easily show that $\ddot{Q}_{\beta\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}$ is a null $p-$dimensional vector. This means that the Hessian matrix is block-diagonal of the form

$$\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \text{block diag}\{\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\},$$

where $\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ and $\ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ are given in (16) and (17), respectively.

Applying (12), we can obtain nice formulae for the the one-step approximation of $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\sigma^2}_{[i]})^\top$, $i = 1,\ldots,n$,, viz., the relationships between the parameter estimates for full data and the data with the ith case deleted. They are given in the next Theorem.

Theorem 1. *For the t-CR model, the relationships between the parameter estimates for full data and the data with the ith case deleted are as follows:*

$$\widehat{\boldsymbol{\beta}}_{[i]}^1 = \widehat{\boldsymbol{\beta}} + \left(\sum_{i=1}^n \mathscr{E}_{0i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1} E_{1[i]},$$

$$\widehat{\sigma^2}_{[i]}^1 = \widehat{\sigma^2} + \left[1-\frac{2}{\widehat{\sigma^2}}\left(\mathscr{E}_{2i}(\widehat{\boldsymbol{\theta}}) - 2\mathscr{E}_{1i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}} + \mathscr{E}_{0i}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}})^2\right)\right]^{-1} E_{2[i]},$$

where $E_{1[i]}$ and $E_{2[i]}$ are as in (14) and (15), respectively.

From Theorem 1, case-deletion measures can be developed for assessing influential observations, such as the generalized Cook distance and the likelihood distance (Zhu & Lee, 2001). To assess the influence of the ith case on the EM estimate $\widehat{\boldsymbol{\theta}}$, we need to compare $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$, and if $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the ith 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 & Lee (2001), we consider here the following *generalized Cook distance*:

$$GD_i = (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}(\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}}), \quad i = 1, \ldots, n. \tag{18}$$

Upon substituting (12) into (18), we obtain the approximation

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

Observe that, since $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is a diagonal matrix, GD_i^1 can be decomposed into the sum

$$GD_i^1 = GD_i^1(\boldsymbol{\beta}) + GD_i^1(\sigma^2),$$

where

$$\begin{aligned}
GD_i^1(\boldsymbol{\beta}) &= \dot{Q}_{[i]\beta}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]\beta}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) \\
&= \frac{1}{\widehat{\sigma^2}} E_{1[i]}^\top \left[\sum_{i=1}^n \mathscr{E}_{0i}(\widehat{\boldsymbol{\theta}}) \mathbf{x}_i \mathbf{x}_i^\top \right]^{-1} E_{1[i]} \text{ and} \\
GD_i^1(\sigma^2) &= \dot{Q}_{[i]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) \\
&= \frac{1}{2\widehat{\sigma^2}} \sum_{i=1}^n \left[1 - \frac{2}{\widehat{\sigma^2}} \left(\mathscr{E}_{2i}(\widehat{\boldsymbol{\theta}}) - 2\mathscr{E}_{1i}(\widehat{\boldsymbol{\theta}}) \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}} + \mathscr{E}_{0i}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}})^2\right)\right]^{-1}.
\end{aligned}$$

are measures of the influence of ith case on the estimates of the parameters $\boldsymbol{\beta}$ and σ^2, respectively, being versions of the generalized Cook distance for each situation.

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

$$QD_i = 2\{Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}_{[i]}|\widehat{\boldsymbol{\theta}})\}. \tag{19}$$

We can calculate an approximation of the likelihood displacement QD_i by substituting (12) into (19), resulting in the following approximation QD_i^1 of QD_i:

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

4.2 Local Influence

In this section, we derive the normal curvature of the local influence (Cook, 1986) for some common perturbation schemes either in the model or in the data. For this purpose, we will consider the case-weight and the scale perturbation schemes.

Consider a perturbation vector $\boldsymbol{\omega} = (\omega_1, ..., \omega_g)^\top$ varying in an open region $\boldsymbol{\Omega} \subset \mathbb{R}^g$. Let $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)$ be the complete-data log-likelihood of the perturbed model. We assume that there is a $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ such that $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ for all $\boldsymbol{\theta}$. Let us define

$$\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) &= \mathrm{E}_{\widehat{\boldsymbol{\theta}}}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)|\mathbf{y}_{\text{obs}}] \quad \text{and} \\
\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}) &= \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = (\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})^\top, \widehat{\sigma^2}(\boldsymbol{\omega}))^\top.
\end{aligned}$$

The influence graph is then 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 follows:

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

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

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

Then, it can be shown that

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

where $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is defined in (13).

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 $\boldsymbol{\varepsilon}_k, k = 1,\ldots,g$. Zhu & Lee (2001) proposed to inspect all eigenvectors corresponding to nonzero eigenvalues for capturing more information, according to the following method: let

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

Let $M(0)_l = \sum_{k=1}^{r} \tilde{\zeta}_k \varepsilon_{kl}^2$ be the lth component of $M(0)$. The assessment of influential cases is based on the visual inspection of $M(0)_l, l = 1,\ldots,g$ plotted against the index l. The lth 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 to use the following conformal normal curvature

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

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 lth 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 is the influence of a given case. 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 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, we may use different functions of $M(0)$. For instance, Zhu & Lee (2001) proposed to use $\overline{M(0)}+2SM(0)$ 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, these authors proposed to use $\overline{M(0)}+c^*SM(0)$, where c^* is a selected constant, and depending on the application, c^* may be taken to be any value. In this paper we will use $c^*=3.5$.

4.3 Perturbation schemes

In this section, we will evaluate the matrix $\mathbf{\Delta}_{\boldsymbol{\omega}_0}$ under the following perturbation schemes for the t-CR model: (i) *Case-weight perturbation*, which is appropriate for detecting observations with outstanding contribution on the log-likelihood function and that may exercise high influence on the maximum likelihood estimates; (ii) *Scale perturbation* made on σ^2, which may reveal individuals that are most influential, in the sense of the likelihood displacement on the scale structure.

For each perturbation scheme, we have the partitioned form

$$\mathbf{\Delta}_{\boldsymbol{\omega}_0}=(\mathbf{\Delta}_{\beta}^{\top},\mathbf{\Delta}_{\sigma^2}^{\top})^{\top},$$

where

$$\mathbf{\Delta}_{\beta}=\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^{\top}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)}\in\mathbb{R}^{p\times g}\ \text{ and }\ \mathbf{\Delta}_{\sigma^2}=\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\boldsymbol{\omega}^{\top}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)}\in\mathbb{R}^{1\times g}.$$

Case weight perturbation

Let us consider the the so called *perturbed Q–function*, which consists of an arbitrary attribution of weights to the expected value of the complete-data log-likelihood function. It can be useful in capturing departures in general directions, and is defined as

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})=\mathrm{E}_{\widehat{\boldsymbol{\theta}}}[\ell_c(\boldsymbol{\theta},\boldsymbol{\omega}|\mathbf{Y}_c)|\mathbf{y}_{\text{obs}}]=\sum_{i=1}^{n}\omega_i\mathrm{E}_{\widehat{\boldsymbol{\theta}}}[\ell_i(\boldsymbol{\theta}|\mathbf{Y}_c)|\mathbf{y}_{\text{obs}}]=\sum_{i=1}^{n}\omega_i Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}).$$

Here, $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_n)^{\top}$ and $\ell_i(\boldsymbol{\theta}|\mathbf{Y}_c)$ is the log-likelihood associated to the pair (u_i,y_i). Observe that $\boldsymbol{\omega}_0=(1,\ldots,1)^{\top}\equiv\mathbf{1}_n^{\top}$ corresponds to the non-perturbed model. In addition, it is possible to show that the local influence for this perturbation scheme is equivalent to the deletion method discussed in preceding section. Therefore, for this perturbation scheme, we find the following coordinates of $\mathbf{\Delta}_{\boldsymbol{\omega}_0}$

$$\begin{aligned}\mathbf{\Delta}_{\beta}&=\frac{1}{\widehat{\sigma^2}}\left[\mathbf{X}^{\top}\mathrm{diag}\{\mathscr{E}_1(\widehat{\boldsymbol{\theta}})\}-\mathbf{A}\right];\\ \mathbf{\Delta}_{\sigma^2}&=-\frac{1}{2\widehat{\sigma^2}}\left\{\mathbf{1}_n^{\top}-\frac{1}{\widehat{\sigma^2}}\mathbf{B}^{\top}\right\},\end{aligned}$$

where $\mathbf{A}$ is a matrix with n columns equal to $\mathbf{X}^{\top}\mathrm{diag}\{\mathscr{E}_0(\widehat{\boldsymbol{\theta}})\}\mathbf{X}^{\top}\widehat{\boldsymbol{\beta}}$, $\mathscr{E}_i(\widehat{\boldsymbol{\theta}})=(\mathscr{E}_{i1}(\widehat{\boldsymbol{\theta}}),\ldots,\mathscr{E}_{in}(\widehat{\boldsymbol{\theta}}))^{\top}$, $i=1,2$, $\mathbf{X}$ is a matrix with rows $\mathbf{x}_i^{\top}$ (that is, the design matrix) and $\mathbf{B}$ is a $n-$dimensional vector with coordinates $B_i=\mathscr{E}_{2i}(\widehat{\boldsymbol{\theta}})-2\mathscr{E}_{1i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^{\top}\widehat{\boldsymbol{\beta}}+\mathscr{E}_{0i}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_i^{\top}\widehat{\boldsymbol{\beta}})^2$, $i=1,\ldots,n$.

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 $i = 1,\dots,n$. Under this perturbation scheme, the non-perturbed model is obtained when $\boldsymbol{\omega}_0 = \mathbf{1}_n^\top$. Moreover, the perturbed Q-function is as in (5), with $\sigma^2(\omega_i)$ and $\widehat{\boldsymbol{\theta}}$ replacing σ^2 and $\boldsymbol{\theta}^{(k)}$, respectively. The matrix $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ has its elements as follows:

$$
\begin{aligned}
\boldsymbol{\Delta}_\beta &= \frac{1}{\widehat{\sigma^2}}\left[\mathbf{X}^\top \mathrm{diag}\{\mathscr{E}_1(\widehat{\boldsymbol{\theta}})\} - \mathbf{A}\right];\\
\boldsymbol{\Delta}_{\sigma^2} &= \frac{1}{2\widehat{\sigma^4}}\mathbf{B}^\top.
\end{aligned}
$$

5 Application

Insulation life data with censoring times

In this section we use the data described at Tan *et al.* (2010, Table 2.3). The data set consists of accelerated life tests on electrical insulation in 40 motorettes. Ten motorettes were tested at each of the four temperatures: 150ºC, 170ºC, 190ºC and 220ºC. Testing was terminated at different times at each temperature level.

We applied the EM algorithm for censored data explained in Section 3, considering both cases when the error term follows a normal and a Student-t distribution, respectively N-CR and t-CR models. The reponse, given as $Y = \log_{10}(lifetime)$ of each motorette, is right-censored. Each one of the 40 vectors of explanatory variable values is given by $\mathbf{x}_i^\top = (1, T_i)$, where $T_i = 1000/(temperature + 273.2)$, $i = 1,\dots,40$. The results are shown in Table 1. For the t-CR model, the degrees of freedom were fixed at the value $\nu = 2$. As we can see by the inspection of the

Table 1: Insulation life data. Results of the parameter estimation via EM algorithm

	N-CR Model		t-CR Model	
Parameter	Estimative	SE	Estimative	SE
β_0	-6.2718	1.4044	-5.7817	0.9129
β_1	4.4832	0.6443	4.1964	0.4229
σ^2	0.1466	-	0.0302	-

standard deviations of the parameter estimates, the t-CR model produces more accurate estimates than the N-CR model.

We proceeded a residual analysis in order to identify atypical observations and/or model misspecification, since residuals are measures of agreement between the data and the fitted model. Consider the traditional standardized ordinary (Pearson) residual:

$$r_i = \frac{Y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}, \quad i = 1,\dots,n.$$

Since their introduction by Atkinson (1981), it is a common practice to generate envelopes based on these Pearson residuals in order to assess the quality of the model fit. In our case, the results are shown in Figure 1, and we can see clearly that the t-CR model fits better the data.

The robustness of the t-CR model can be assessed by considering the influence of a single outlying observation on the EM estimate of $\boldsymbol{\theta}$. In particular, we can assess how much the EM

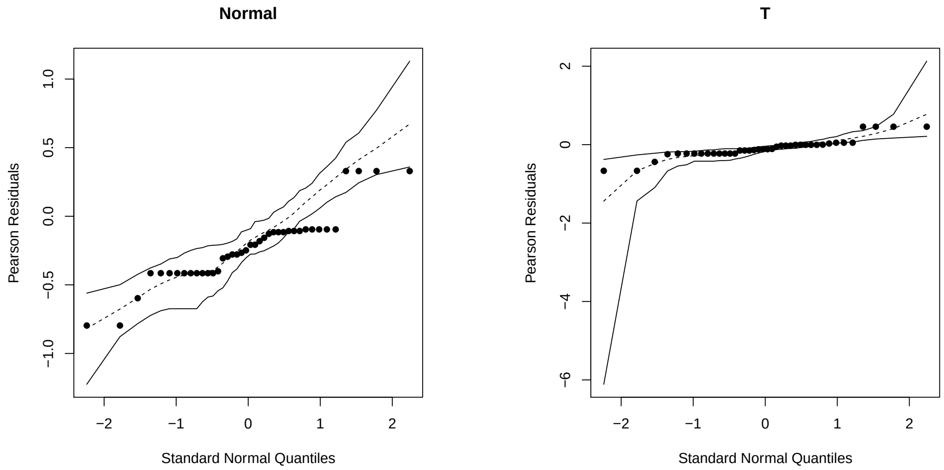

Figure 1: Insulation life data. Simulated envelopes for the Pearson residuals.

estimate of $\boldsymbol{\theta}$ is influenced by a change of δ units in a single observation y_i. Replacing y_i by $y_i(\delta) = y_i + \delta$, let $\widehat{\sigma^2}(\delta)$ and $\widehat{\beta}_j(\delta)$ be the EM estimates of σ^2 and β_j after contamination, $j = 1,2$. We are particularly interested in the relative changes $|\widehat{\sigma^2}(\delta) - \widehat{\sigma^2}|/\widehat{\sigma^2}$ and $|(\widehat{\beta}_j(\delta) - \widehat{\beta}_j)/\widehat{\beta}_j|$. Figure 2 displays the results of the relatives changes of the estimates for different values of δ, under both models, contaminating the observation 20 and varying δ between 0 and 5. As expected, the estimates from the t-CR model are less affected by variations of δ.

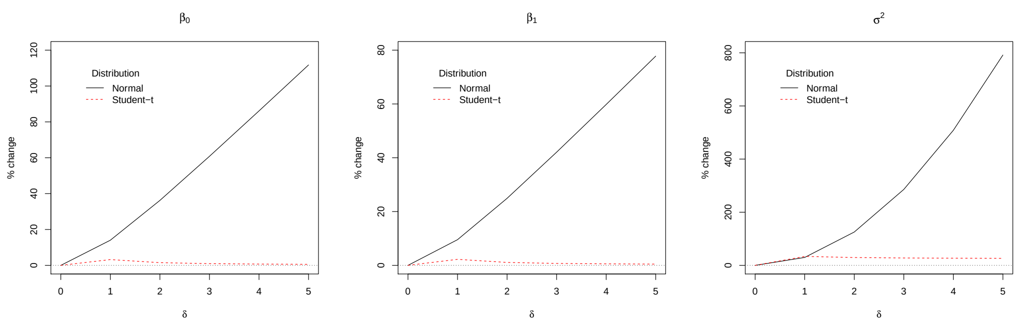

Figure 2: Insulation life data. Relative changes on the maximum likelihood estimation of $\boldsymbol{\theta}$ from the N-CR (solid line) and t-CR (dashed line) models for different contaminations δ.

Diagnostic analysis

In order to identify influential observations we can generate graphs of the generalized Cook distance, as explained in Section 4.2. A high value for GD_i indicates that the ith observation has a high impact on the maximum likelihood estimate of the parameters. Making an adaptation of the suggestion of Barros *et al.* (2010), we can use $2(p+1)/n$ as benchmark for the GD_i, $2p/n$ as benchmark for the $GD_i(\boldsymbol{\beta})$ and $2/n$ as benchmark for the $GD_i(\sigma^2)$, $i = 1,\ldots,n$.

In Figure 3 we note that, under the the normal fit, observations #8 and #9 appear to influence equally the estimation of the parameters (in the first graph, the GD related to observation #1 appears

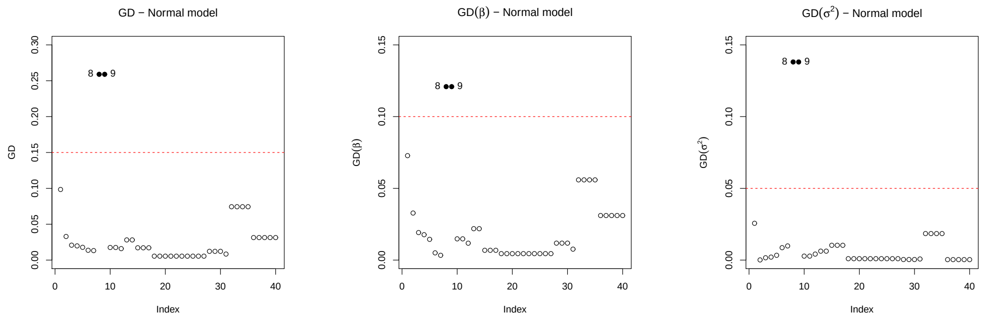

Figure 3: Insulation life data. Considering the N-CR model: (a) approximate generalized Cook's distance GD_i; (b) GD_i for subset $\boldsymbol{\beta}$; (c) GD_i for subset σ^2. The influential observations are numbered.

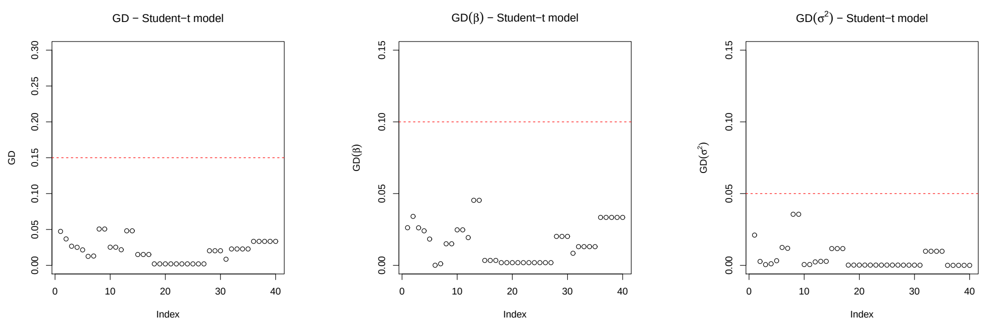

Figure 4: Insulation life data. Considering the t-CR model: (a) approximate generalized Cook's distance GD_i; (b) GD_i for subset $\boldsymbol{\beta}$; (c) GD_i for subset σ^2. If detected, the influential observations were numbered.

to lie above the benchmark, which is 0.15 but, in fact, we have $GD_1 = 0.0984$). In Figure 4, with the t-CR model fitted, the scenario has changed: according to this criterion, there are not observations influencing the maximum likelihood estimation, showing that this model is more robust.

Next, we studied local influence based on $M(0)$ – see Section 4.2. Here we used the criterion $M(0)_i > \overline{M(0)} + 3.5SM(0)$, $i = 1, \ldots, n$ to discriminate whether an observation is influential. Figure 5 displays the results for the N-CR model. Observe that observations #8 and #9 appears as equally influential under the case weight and the scale perturbation; also, the influence of those observations seems not change under both cases. In Figure 6 we see one more evidence that the t-CR model is more robust than the N-CR one: no observation seems to be influential if we fit the first one, no matter what kind of perturbation we apply.

For the purposes of this analysis, we define the relative change (RC) on estimation of some parameter γ as

$$RC_{\hat{\gamma}} = \left| \frac{\hat{\gamma} - \hat{\gamma}_{[i]}}{\hat{\gamma}} \right|,$$

where $\hat{\gamma}_{[i]}$ denotes the maximum likelihood estimate of γ after some set of observations I_i has been removed. Table 2 shows the RC for the estimates of the parameters β_0, β_1 and σ^2. We considered $I_1 = \{\#8\}$, $I_2 = \{\#9\}$ and $I_3 = \{\#8, \#9\}$. Again, the main impression is that observations #8 and #9

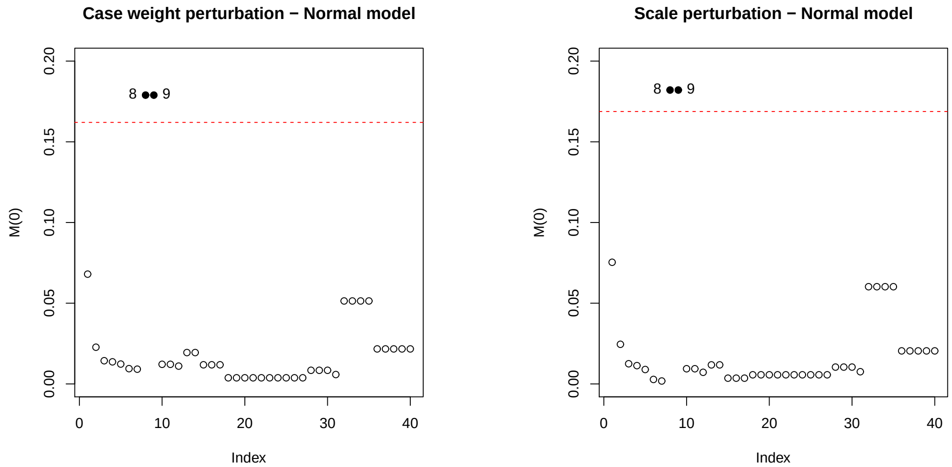

Figure 5: Insulation life data. Considering the N-CR model: index plot of $M(0)$ for assessing local influence on $\boldsymbol{\theta}$ under (a) case weight perturbation; (b) scale perturbation. The influential observations are numbered.

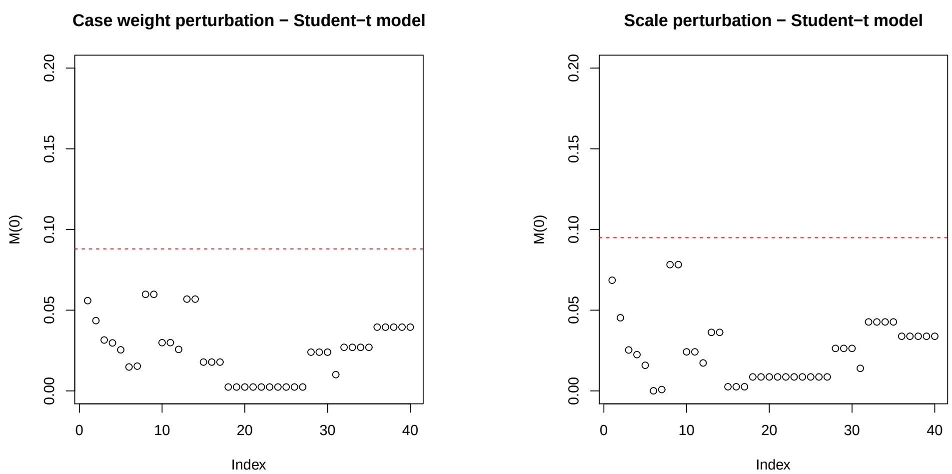

Figure 6: Insulation life data. Considering the t-CR model: index plot of $M(0)$ for assessing local influence on $\boldsymbol{\theta}$ under (a) case weight perturbation; (b) scale perturbation. If detected, the influential observations were numbered.

seem to have high influence on the estimation of β_1 and β_2 under the N-CR model. According to this criterion, it appears that they do not have (or have a little) influence under the t-CR model.

6 Simulation Studies

6.1 Study 1

To assess the performance of our proposed methodology we conducted a simulation study. The goal was to investigate the consequences on parameter inference when the normality assumption

Table 2: RC for the estimates of β_0 and β_1 under the normal and t-CR models (in percentage)

	N-CR Model		t-CR Model	
Dropped	$RC_{\hat{\beta}_0}$	$RC_{\hat{\beta}_1}$	$RC_{\hat{\beta}_0}$	$RC_{\hat{\beta}_1}$
I_1	3.85	2.38	0.93	0.62
I_2	3.85	2.38	0.93	0.62
I_3	8.47	5.24	1.64	1.11

is inappropriate.

First, we considered the t-R model, given in (2). We fixed $n = 100$, $\boldsymbol{\beta}^\top = (\beta_0, \beta_1) = (2, 1)$, $\sigma^2 = 1$ and $\nu = 4$. The design matrix $\mathbf{X}$, that is, the matrix with rows $\mathbf{x}_i^\top$, was constructed in the following way:

$$\mathbf{X} = (\mathbf{1}^\top, \mathbf{t}^\top \otimes \mathbf{1}_{10}), \quad \text{with } \mathbf{t}^\top = (1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8),$$

where $\otimes$ denotes the Kronecker product.

We chose several settings of censoring proportions (5%, 10%, 20% and 50%) to study the effect of the level of censoring on the estimation. In this way, we have 4 different simulation settings with 1000 simulated data sets for each setting. Once the simulated data was generated, we fitted the censored regression model assuming normal and Student-t distributions for the observational errors and, after this, we recorded the parameter estimates values. Table 3 displays these results and some other statistics for the 4 censoring patterns. The average values (MC Mean) and the corresponding standard deviations (MC Sd) of the EM estimates across all samples were computed. For example, MC Mean denotes the averages $\sum_{j=1}^{1000} \widehat{\beta}_{kj}/1000$ and $\sum_{j=1}^{1000} \widehat{\sigma^2}_j/1000$, where $\widehat{\beta}_{kj}$ and $\widehat{\sigma^2}_j$ are the maximum likelihood estimates of β_k and σ^2 obtained at the jth generated sample, $k = 1, 2$. Also were computed the average values of the approximate standard deviations of the EM estimates obtained through the information-based method described in (11) (IM Sd). Finally, MC Coverage is the number of times (divided by 1000) the confidence intervals of the form $parameter\ estimate \pm 1.96Sd(parameter\ estimate)$ cover the true value of the parameter.

From Table 3, we observe that the t-CR model presents a better performance for all levels of censoring – the standard deviations are smaller in all cases. Figures 7–8 show that for the N-CR model there is a strong increase of the bias (the deviations of the parameter estimates from the true value) as well as the mean square error (MSE). Clearly, the t-CR model shows much less bias and smaller MSE values, and consequently more precise estimates. It suggests that a model with heavier tails than the normal one produces more accurate estimates in the context of censored data, since in our analysis all the measures strongly favored the t-CR model.

6.2 Study 2

In this second simulation study we focused our attention on the percentage of change on the estimation of the parameters when the data is perturbed, allowing a comparison of the performances when fitting the N-CR and the t-CR models (when the data is generated from a t-CR model).

With this purpose, considering two settings of censoring proportions (20% and 50%), we generated 100 independent observations from a population following the t-CR model for each setting, and repeated the procedure 1000 times. Once the data was generated, one observation for each data set was contaminated by adding δ units, with $\delta = 0, 1, \ldots, 10$, and then they were modeled

Table 3: Monte Carlo results based on 1000 simulated t-CR samples.

			Simulated t-CR data		
Censoring	Model		β_0	β_1	σ^2
5%	Normal	MC Mean	2.0013	0.9914	1.8031
		MC Sd	(0.4708)	(0.2337)	(0.5265)
		IM Sd	0.4611	0.2327	
		MC Coverage	94.3%	95.1%	
	t-CR	MC Mean	2.0120	0.9923	0.8927
		MC Sd	(0.4184)	(0.2106)	(0.2702)
		IM Sd	0.4328	0.2180	
		MC Coverage	95.1%	95.9%	
10%	Normal	MC Mean	1.9834	0.9980	1.7533
		MC Sd	(0.4756)	(0.2383)	(0.6891)
		IM Sd	0.4560	0.2307	
		MC Coverage	94.7%	95%	
	t-CR	MC Mean	2.0065	0.9946	0.8758
		MC Sd	(0.4285)	(0.2130)	(0.3207)
		IM Sd	0.4231	0.2132	
		MC Coverage	94.2%	95.3%	
20%	Normal	MC Mean	1.9300	1.0262	1.7345
		MC Sd	(0.4852)	(0.2445)	(0.6419)
		IM Sd	0.4598	0.2346	
		MC Coverage	93.8%	94.3%	
	t-CR	MC Mean	1.9751	1.0080	0.8559
		MC Sd	(0.4431)	(0.2216)	(0.3319)
		IM Sd	0.4133	0.2095	
		MC Coverage	93.7%	93.6%	
50%	Normal	MC Mean	1.8641	1.1209	1.9701
		MC Sd	(0.5538)	(0.3139)	(0.9188)
		IM Sd	0.5252	0.2781	
		MC Coverage	94.8%	93.1%	
	t-CR	MC Mean	1.9836	1.0193	1.0813
		MC Sd	(0.4697)	(0.2582)	(0.3861)
		IM Sd	0.5037	0.2780	
		MC Coverage	97.1%	96.6%	

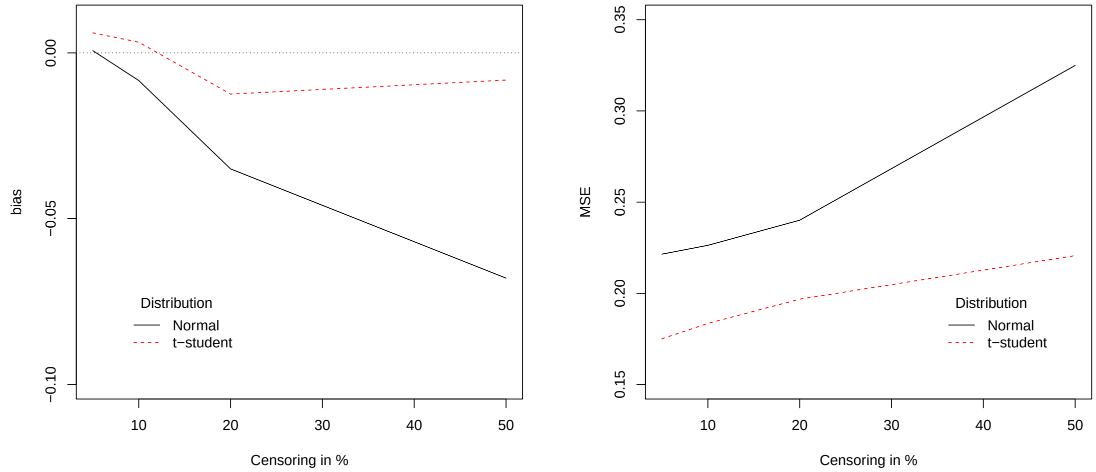

Figure 7: Bias and MSE of $\widehat{\beta}_0$ in comparison with the true value for normal and t-CR models for 4 censoring patterns (5%, 10%, 20%, 50%).

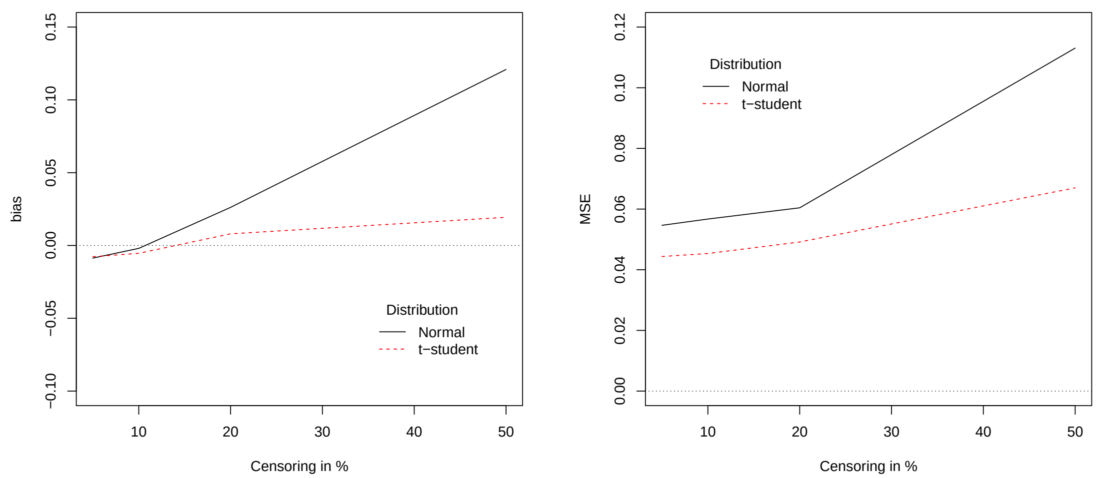

Figure 8: Bias and MSE of $\widehat{\beta}_1$ in comparison with the true value for normal and t-CR models for 4 censoring patterns (5%, 10%, 20%, 50%).

using the N-CR and the t-CR models. The percentage of change on the estimation was taken as the average of the change in each one of the 1000 data sets. The results are shown in figures 9 and 10.

The percentage of change on the estimation of the parameters under the wrong assumption of normality is greater than that one under the right assumption of Student-t errors. With 20% of censoring and the t-CR model fit, the percentage of change is constant for all the contaminations δ while with the N-CR model fit the percentage of chance increases as δ increases (with an exponential pattern for σ^2). With 50% of censoring the percentage of change follows almost the same pattern of increase for both models (with a little bigger increase for σ^2 under normality and $\delta > 7$), but it is always smaller under the t-CR model fit. With this analysis, we see that the wrong assumption of normality prejudices the estimation of the parameters and that the model with Student-t errors deals better than the normal one with contaminated and censored data.

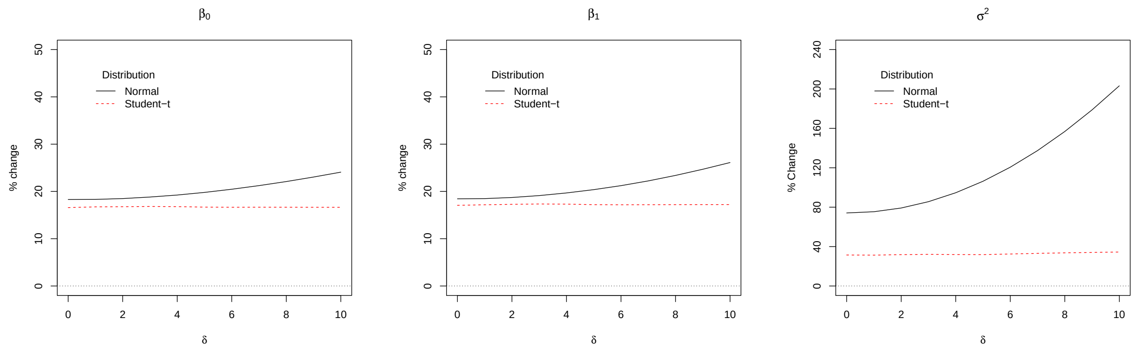

Figure 9: Percentage of change on the estimation of β_0, β_1 and σ^2 in comparison with the true value for the N-CR and t-CR models with 20% of censoring for different contaminations δ.

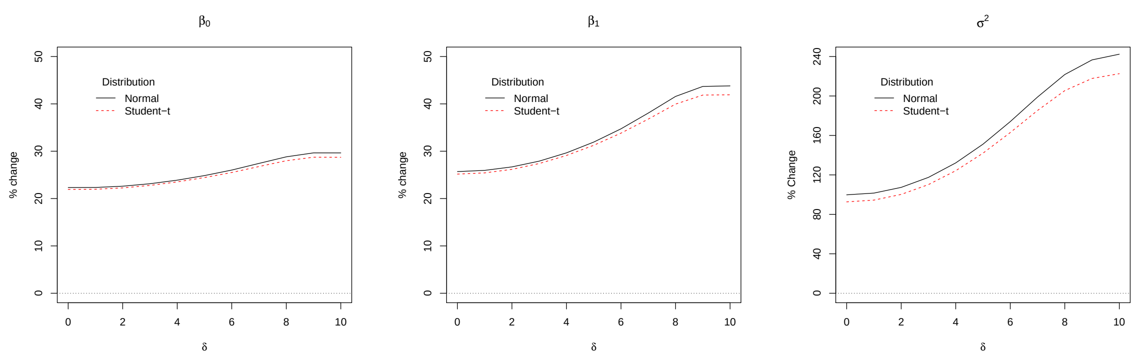

Figure 10: Percentage of change on the estimation of β_0, β_1 and σ^2 in comparison with the true value for the N-CR and t-CR models with 50% of censoring for different contaminations δ.

7 Conclusion

We presented diagnostic analysis in linear regression models with censored responses and observational errors following a Student-t distribution.The approach was based on case-deletion and local influence techniques suggested by Zhu & Lee (2001), that are the counterpart for missing data models of the well-known ones proposed by Cook (1977) and Cook (1986). An EM-type algorithm was obtained, which can be easily implemented using available software, like R and MATLAB. It was implemented as part of the R package `CensRegMod()`. The structure of the complete-data likelihood function, obtained considering as the missing data were in fact observed, is an essential element of the theory. Its simple form allows us to obtain a tractable expression for the Q-function, which is essentially what we need 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 case of the normal curvature expressions. Using the developed methodology, we analyzed a real data set and proceeded two extensive simulation studies. The results showed that the model for censored data with Student-t errors is very robust against outlying observations, outperforming the model with normal errors.

Recently, Genç (2012) has considered the problem of finding the moments of a doubly truncated member of the class of normal/independent (NI) distributions. Therefore, it would be a worthwhile task to investigate the applicability of a likelihood based treatment in the context of NI-CR models.

A Proof of Equation (6)

Supposing that $Y_i \sim \mathrm{t}_\nu(\mathbf{x}_i^\top\boldsymbol{\beta},\sigma^2)$, then we can write $Y_i|U_i=u_i \sim \mathrm{N}(\mathbf{x}_i^\top\boldsymbol{\beta},u_i^{-1}\sigma^2)$, with $U_i \sim$ Gamma$(\nu/2,\nu/2)$. We will prove that

$$\mathrm{E}_{\boldsymbol{\theta}}[U_i|y_i]=\frac{\nu+1}{\nu+d^2(\boldsymbol{\theta},y_i)},$$

where $\boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\sigma^2)^\top$. To facilitate notation, we will write $\mu_i=\mathbf{x}_i^\top\boldsymbol{\beta}$ and $\mathrm{E}[U_i|y_i]$ instead of $\mathrm{E}_{\boldsymbol{\theta}}[U_i|y_i]$. Then,

$$\begin{aligned}
\mathrm{E}[U_i|y_i] &= \frac{1}{\pi(y_i)}\int u_i\pi(y_i|u_i)\pi(u_i)du_i \\
&= \frac{(\nu/2)^{(\nu/2)}}{\sqrt{2\pi\sigma^2}\pi(y_i)\Gamma(\nu/2)}\int u_i^{(\nu+1)/2}\exp\left\{-\frac{u_i}{2}\left[\frac{(y_i-\mu_i)^2}{\sigma^2}+\frac{\nu}{2}\right]\right\}du_i \\
&= \frac{(\nu/2)^{(\nu+1)/2}}{\Gamma\left(\frac{\nu+1}{2}\right)}\left[1+\frac{(y_i-\mu_i)^2}{\nu\sigma^2}\right]^{(\nu+1)/2}\int u_i^{(\nu+1)/2}\exp\left\{-\frac{u_i}{2}\left[\frac{(y_i-\mu_i)^2}{\sigma^2}+\nu\right]\right\}du_i.
\end{aligned}$$

Let us define

$$a-1=\frac{\nu+1}{2} \text{ and } b=\frac{1}{2}\left[\frac{(y_i-\mu_i)^2}{\sigma^2}+\nu\right].$$

Then,

$$\begin{aligned}
\mathrm{E}[U_i|y_i] &= \frac{(\nu/2)^{(\nu+1)/2}}{\Gamma\left(\frac{\nu+1}{2}\right)}\left(\frac{2b}{\nu}\right)^{(\nu+1)/2}\int u_i^{a-1}\exp\{-bu_i\}du_i \\
&= \frac{(\nu/2)^{(\nu+1)/2}2^{(\nu+1)/2}(\nu+1)}{2\nu^{(\nu+1)/2}b}\int\frac{1}{\Gamma(a)}u_i^{a-1}\exp\{-bu_i\}b^a du_i \\
&= \frac{\nu+1}{2b}.
\end{aligned}$$

Acknowledgements

The research of V. H. Lachos was supported by Grant 305054/2011-2 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil). The research of M. B. Massuia was supported by Grant 2011/07978-0 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil). C. R. B. Cabral is deeply indebted to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-Brazil), for financial support via Project PROCAD-2007 and CNPq-Brazil, via Projects Universal and CT-Amazônia.

References

Arellano-Valle, R., Castro, L., González-Farías, G. & Muñoz Gajardo, K. (2012). Student-t censored regression model: properties and inference. *Statistical Methods and Applications*, **DOI: 10.1007/s10260-012-0199-y**.

Atkinson, A. C. (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika*, **68**, 13–20.

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

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

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

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.

Cysneiros, F. & Paula, G. (2005). Restricted methods in symmetrical linear regression models. *Computational Statistics & Data Analysis*, **49**, 689–708.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Fernández, C. & Steel, M. J. F. (1999). Multivariate Student-t regression models: pitfalls and inference. *Biometrika*, **86**, 153.

Genç, A. (2012). Moments of truncated normal/independent distributions. *Statistical Papers*, **DOI 10.1007/s00362-012-0459-9**.

Greene, W. (2012). *Econometric Analysis*. Prentice Hall, NJ.

Kim, H. J. (2008). Moments of truncated Student-t distribution. *Journal of the Korean Statistical Society*, **37**, 81–87.

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. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, **2**, 175–198.

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.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, pages 226–233.

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

Massuia, M. B., Matos, L. A. & Lachos, V. H. (2012). *CensRegMod: Fitting Normal and Student-t censored regression model*. R package version 0.0.

Matos, L. A., Lachos, V. H., Balakrishnan, N. & Labra, F. V. (2012). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistics & Data Analysis*, **doi:10.1016/j.csda.2012.06.021**.

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.

Mroz, T. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, **55**, 765–799.

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.

Rancel, M. M. S. & Sierra, M. A. G. (2001). Regression diagnostic using local influence: a review. *Communications in Statistics - Theory and Methods*, **30**, 799–813.

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

Tan, M. T., Tian, G. & Ng, K. W. (2010). *Bayesian Missing Data Problems: EM, Data Augmentation and Noniterative Computation*. Chapman & Hall/CRC.

Vanegas, L. H. & Cysneiros, F. J. A. (2010). Assessment of diagnostic procedures in symmetrical nonlinear regression models. *Computational Statistics & Data Analysis*, **54**, 1002–1016.

Wei, C. G. & Tanner, M. A. (1990). Posterior computations for censored regression data. *Journal of the American Statistical Association*, **85**, 829–839.

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.

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