Communications for Statistical Applications and Methods
2023, Vol. 30, No. 6, 605–629

https://doi.org/10.29220/CSAM.2023.30.6.605
Print ISSN 2287-7843 / Online ISSN 2383-4757

Influence diagnostics for skew-t censored linear regression models

Marcos S Oliveira[a], Daniela CR Oliveira[a], Victor H Lachos[1,b]

[a]Department of Mathematics and Statistics, Federal University of Sao Joao del-Rei, MG, Brazil;
[b]Department of Statistics, University of Connecticut, Storrs, CT-06269, USA

Abstract

This paper proposes some diagnostics procedures for the skew-t linear regression model with censored response. The skew-t distribution is an attractive family of asymmetrical heavy-tailed densities that includes the normal, skew-normal and student's-t distributions as special cases. Inspired by the power and wide applicability of the EM-type algorithm, local and global influence analysis, based on the conditional expectation of the complete-data log-likelihood function are developed, following Zhu and Lee's approach. For the local influence analysis, four specific perturbation schemes are discussed. Two real data sets, from education and economics, which are right and left censoring, respectively, are analyzed in order to illustrate the usefulness of the proposed methodology.

Keywords: case-deletion model, censored regression, EM-type algorithm, local influence, model perturbation, skewness, skew-t distribution

1. Introduction

The censored regression (CR) model, or the Tobit model, has become quite common in the literature with a wide range of applications. For continuous data, the CR model usually uses the normal distribution (N-CR). However, it is well-known that the normal distribution is sensitive to outliers, thus a large number of parametric models to provide flexibility in modeling data have been investigated in recent years. For instance, Massuia *et al*. (2015) have studied CR models based on the Student's-t distribution (T-CR) and demonstrated the robustness aspects of the T-CR model against outliers through extensive simulations. Note however that the T-CR model is not appropriate when the data, simultaneously, present skewness and heavy tails (large kurtosis).

Recently, Lachos *et al*. (2022) have established a new link between the CR model and asymmetrical heavy tails distributions by using the skew-t (ST) distributions (ST-CR), which allows capturing, simultaneously, skewness and kurtosis and contains, as special cases, the normal (N), Student's-t (T) and skew-normal (SN) distributions. In that paper, an analytically simple EM-type algorithm for computing maximum likelihood (ML) estimates of the ST-CR model is proposed, where they show that the E-step reduces to computing the first two moments of a truncated skew-t distribution with specific parameter. The general formulas for these moments were recently derived explicitly by Lachos *et al*. (2020), thus the proposed EM algorithm is exact and does not require approximations at the E and M steps.

[1]Corresponding author: Department of Statistics, University of Connecticut, 215 Glenbrook Rd. U-4120, Storrs, CT-06269, USA. E-mail: hlachos@uconn.edu

Published 30 November 2023 / journal homepage: http://csam.or.kr

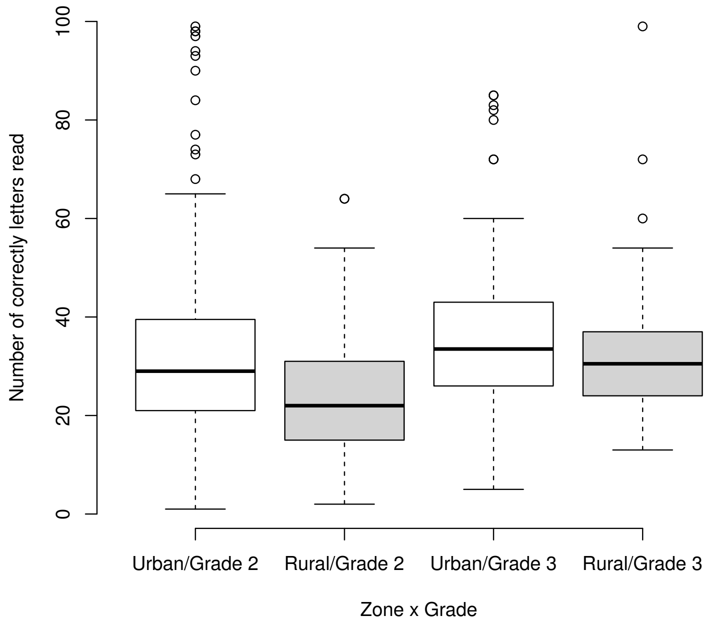

Figure 1: *Boxplot of the response variable in relation to the covariates for LNF data.*

On the other hand, the study of influence analysis is an important and key step in data analysis subsequent to parameter estimation. This can be carried out by conducting an influence analysis for detecting influential observations. There are two primary approaches for detecting influential observations. The first approach is the case-deletion approach and it is an intuitively appealing method (see, Cook and Weisberg, 1982). Deletion diagnostics such as Cook's distance or the likelihood distance have been applied to many statistical models. The second approach, which is a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs, is the local influence approach of Cook (1986). Following the pioneering work of Cook (1986), this method has received considerable attention recently in the statistical literature on CR models; see, for example, Matos *et al.* (2013), Matos *et al.* (2015), Barros *et al.* (2018), Nuñez *et al.* (2021), among many others.

Although several diagnostics studies on CR models have appeared in the literature, to the best of our knowledge, no study seems to have been made on influence diagnostics for ST-CR and certainly not on the local influence analysis. In this paper, for performing diagnostics analysis in the ST-CR model, we use the EM-type algorithm proposed by Lachos *et al.* (2022). Our development is based on Zhu and Lee (2001) approach, which is a method for performing local influence analysis for general statistical models with incomplete data, and it is based on the Q-displacement function that is closely related to the conditional expectation of the complete-data log-likelihood in the E-step of the EM algorithm. Moreover, the case-deletion can be studied by the Q-displacement function following the approach of Zhu *et al.* (2001). A fact to be highlighted is that even robust parameter estimation models (skewed and heavy-tailed) can present unusual observations such as outliers or influential

Table 1: ML estimates, approximate standard errors (SE) and some information criteria for LNF data

Parameter	N-CR		SN-CR		T-CR		ST-CR	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_0	32.460	1.098	12.311	1.152	29.355	0.805	14.958	1.209
β_1	−7.288	1.914	−4.563	1.497	−5.784	1.419	−4.689	1.262
β_2	6.349	1.746	5.934	1.306	6.307	1.445	6.201	1.230
σ^2	283.318	14.499	673.830	45.412	137.457	12.557	358.136	43.547
λ	-	-	4.002	0.632	-	-	2.682	0.506
ν	-	-	-	-	3.5	-	4.8	-
− log-like *	2058.966		2007.595		2023.158		**1995.205**	
AIC	4125.931		4025.191		4056.317		**4002.410**	
BIC	4142.877		4046.373		4077.498		**4027.828**	
CAIC	4146.877		4051.373		4082.498		**4033.828**	
HQIC	4132.575		4033.495		4064.621		**4012.375**	

*log-like: log-likelihood.

observations. Thus, diagnostics methods are still important tools for detecting anomalies in the fitted model. Moreover, we believe that the results of this paper are a necessary supplement of Lachos *et al*. (2022).

The paper is organized as follows. In Section 2, the ST-CR model is defined, and an EM-type algorithm for obtaining the ML estimates are briefly described. In Section 3, we provide a brief sketch of the local influence approach for models with incomplete data, and also develop a methodology pertinent to the ST-CR model. Four different perturbation schemes are considered. Section 4 discusses two applications involving letter-name fluency (LNF) test in Peruvian students and housewife wages. Finally, Section 5 concludes with some discussion and possible directions for future research.

2. The skew-t censored linear regression model

In this section, we consider the skew-t censored linear regression model (ST-CR). Further information on this topic can be found in Lachos *et al*. (2022). To understand this model, it is necessary to first introduce some concepts and terminology. We will begin by defining the skew-normal (SN) distribution. As introduced by Azzalini (1985), a random variable Z has a SN distribution with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma^2 \in (0, \infty)$ and skewness parameter $\lambda \in \mathbb{R}$, denoted by $Z \sim \mathrm{SN}(\mu, \sigma^2, \lambda)$, if its probability density function (pdf) is given by $\phi_{SN}(z \mid \mu, \sigma^2, \lambda) = 2\phi_N(z \mid \mu, \sigma^2)\Phi_N(\lambda(z-\mu)/\sigma) \mid \mu, \sigma^2)$, with $\phi_N(\cdot \mid \mu, \sigma^2)$ and $\Phi_N(\cdot \mid \mu, \sigma^2)$ denoting the pdf and the cumulative distribution function (cdf) of the normal distribution ($\mathrm{N}(\mu, \sigma^2)$), respectively. We denote the cdf of Z by $\Phi_{SN}(\cdot|\mu, \sigma^2, \lambda)$. A useful way to express the stochastic nature of Z can be provided by the following representation

$$Z = \mu + \Delta T + \Gamma^{\frac{1}{2}} T_1, \tag{2.1}$$

where $\Delta = \sigma\delta$, $\Gamma = (1 - \delta^2)\sigma^2$, $\delta = \lambda(1 + \lambda^2)^{-1/2}$, $T = |T_0|$, and T_0 and T_1 are independent standard normal random variables. Here, $|\cdot|$ denotes the absolute value. It should be emphasized that this stochastic representation is valuable not only for generating random samples but also for calculating moments and exploring additional notable characteristics.

Let $Z \sim \mathrm{SN}(0, \sigma^2, \lambda)$ and $U \sim \mathrm{gamma}(\nu/2, \nu/2)$ assuming that Z and U are independent. Here, gamma$(\nu/2, \nu/2)$ denotes the gamma distribution with scale and shape parameters equal to $\nu/2$. We say that the distribution of $Y = \mu + U^{-1/2}Z$ is a skew-t distribution with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma^2 \in (0, \infty)$, skewness parameter $\lambda \in \mathbb{R}$ and degrees of freedom $\nu \in (0, \infty)$. We use

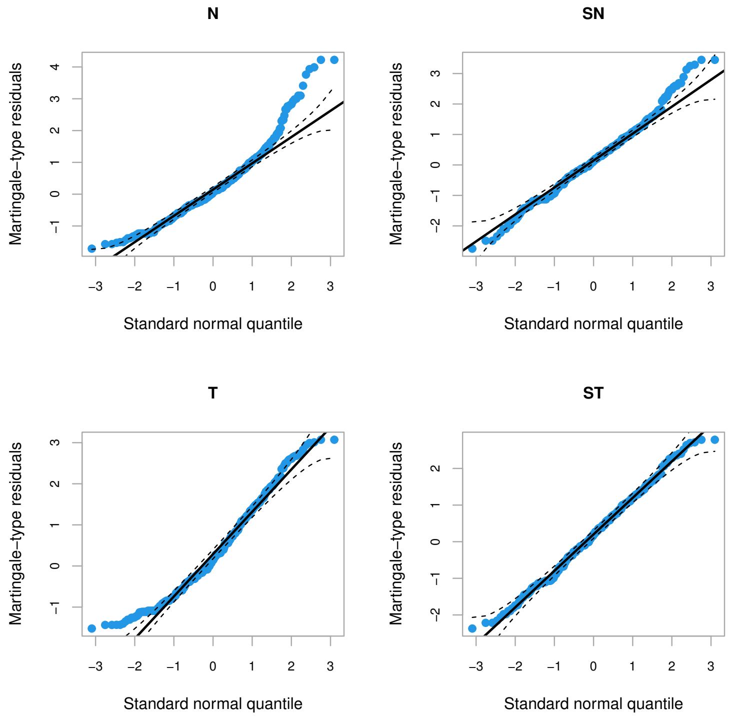

Figure 2: *Martingale-type residuals under the N-CR, SN-CR, T-CR and ST-CR models for LNF data.*

the notation $Y \sim \mathrm{ST}(\mu, \sigma^2, \lambda, \nu)$, with pdf given by

$$\phi_{ST}\left(y \mid \mu, \sigma^2, \lambda, \nu\right) = 2\phi_T\left(y \mid \mu, \sigma^2, \nu\right)\Phi_T\left(\left(\frac{\nu+1}{d+\nu}\right)^{\frac{1}{2}} A \mid \mu, \sigma^2, \nu+1\right),$$

where $A = \lambda(y-\mu)/\sigma$ and $d = (y-\mu)^2/\sigma^2$. In the same way, $\phi_T(\cdot \mid \mu, \sigma^2, \nu)$ and $\Phi_T(\cdot \mid \mu, \sigma^2, \nu)$ denote the pdf and the cdf of the student's-*t* distribution ($\mathrm{T}(\mu, \sigma^2, \nu)$), respectively. Some particular cases of the skew-*t* distribution are the skew-Cauchy distribution ($\nu = 1$) and the student's-*t* distribution ($\lambda = 0$). Also, when $\nu \to \infty$, the skew-normal distribution arises as a limit case. Furthermore, the cdf of a skew-*t* random variable, denoted by $\Phi_{ST}(\cdot|\mu, \sigma^2, \lambda, \nu)$, is given by

$$\Phi_{ST}\left(y \mid \mu, \sigma^2, \lambda, \nu\right) = 2\ \Phi_{T_2}\left(\frac{y-\mu}{\sigma}\mathbf{e}_1 \mid \mathbf{0}, \mathbf{\Sigma}, \nu\right),$$

where $\mathbf{e}_1 = (1, 0)^\top$, $\mathbf{\Sigma} = \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}$, and Φ_{T_2} is the cdf of the bivariate student´s-*t* distribution. Moreover, the conditional distribution of Y given U is

$$Y \mid U = u \sim \mathrm{SN}\left(\mu, u^{-1}\sigma^2, \lambda\right), \quad U \sim \mathrm{gamma}\,(\nu/2, \nu/2)\,. \tag{2.2}$$

For additional properties on the skew-*t* distribution, such as its truncated version, truncated moments, linear transformations, and marginal and conditional distributions, we refer to Lachos *et al.* (2022).

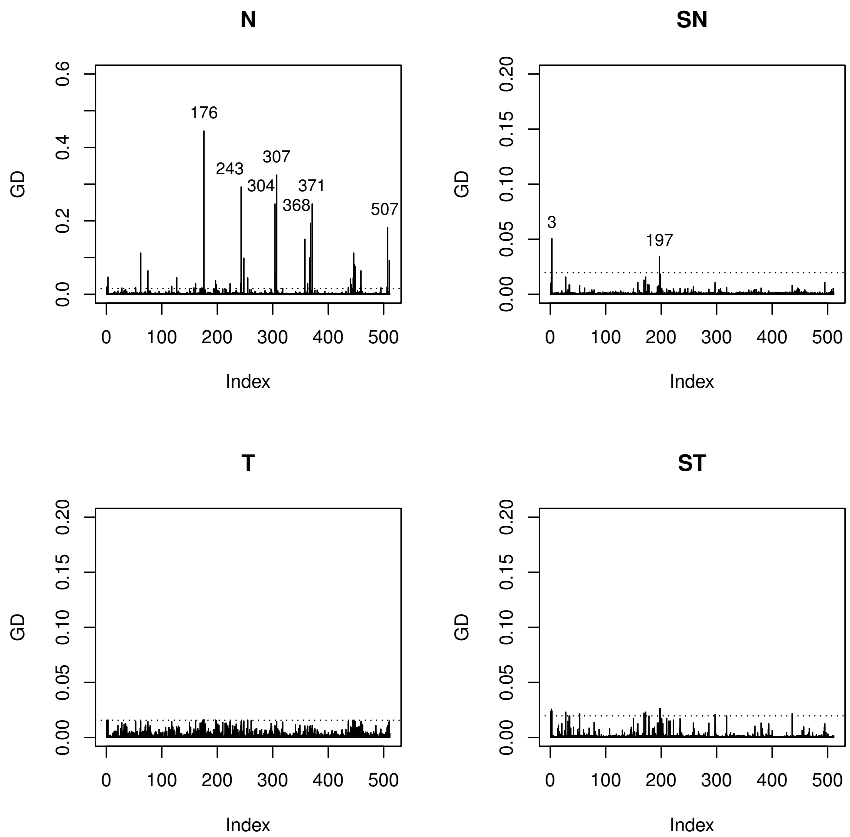

Figure 3: *Approximate generalized Cook's distance (GD_i) under the N-CR, SN-CR, T-CR and ST-CR models for LNF data.*

Lachos *et al*. (2022) considered a linear regression model where the responses are observed with errors which are independent and identically distributed (iid) according to some ST distribution, as follows:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \sigma\varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} \text{ST}(0, 1, \lambda, \nu), \;\; i = 1, \ldots, n, \tag{2.3}$$

where Y_i, $i = 1, \ldots, n$ are the responses, $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$ is unknown regression parameter vector, which contains the intercept β_0, and $\mathbf{x}_i^\top = (1, x_{i1}, \ldots, x_{ip})$ is a vector of known characteristics such that x_{ij} is the value of the j^{th} explanatory variable for subject i. Under this setup, we have that $Y_i \overset{\text{ind}}{\sim} \text{ST}(\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu)$, $i = 1, \ldots, n$. To simplify the mathematical derivations, we will assume that the observations can be left censored, which means that the observations are of the form:

$$Y_{obs_i} = \begin{cases} \kappa_i & \text{if } \; Y_i \le \kappa_i, \\ Y_i & \text{if } \; Y_i > \kappa_i, \end{cases} \tag{2.4}$$

$i = 1, \ldots, n$, for some threshold point κ_i. The model defined in (2.3) and (2.4) is called the skew-t linear censored regression (ST-CR) model. To obtain additional information and a more detailed

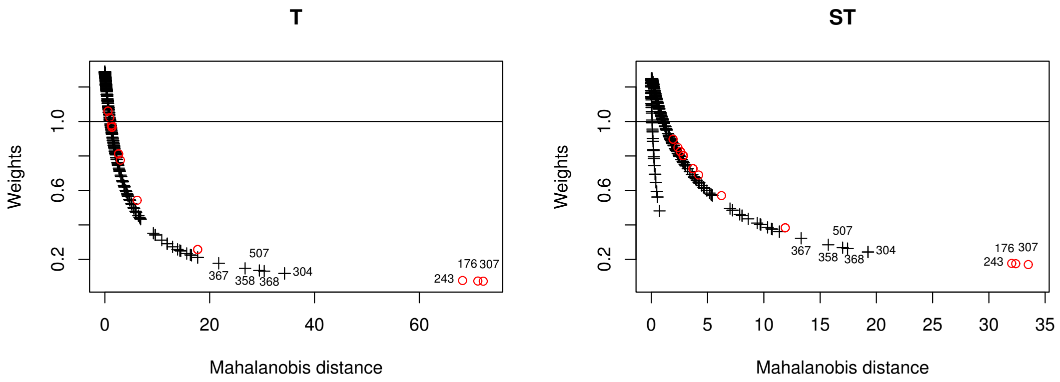

Figure 4: *Estimated u_i under the T-CR and ST-CR models for LNF data. The red ball-shaped points represent the imputed values for the censored values.*

explanation of the topic at hand, see Massuia *et al.* (2017) and Mattos *et al.* (2018). Note that the right censored problem can be represented by a left censored problem by transforming the response Y_{obs_i} to $-Y_{obs_i}$.

Supposing that are m censored values of the characteristic of interest, we can partition the observed sample $\mathbf{y}_{obs}$ in two subsamples of m censored and $n-m$ uncensored values, such that $\mathbf{y}_{obs} = \{\kappa_1, \ldots, \kappa_m, y_{m+1}, \ldots, y_n\}$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda, \nu)^\top$ be the vector with all parameters of the ST-CR model, so the log-likelihood function for $\boldsymbol{\theta}$ is given by

$$\begin{aligned}\ell(\boldsymbol{\theta} \mid \mathbf{y}_{obs}) &= \log\left(\prod_{i=1}^{n}\left[\Phi_{ST}(\kappa_i \mid \mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu)\right]^{\mathbb{I}_i}\left[\phi_{ST}(y_i \mid \mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu)\right]^{1-\mathbb{I}_i}\right)\\ &= \sum_{i=1}^{m}\log\left[\Phi_{ST}(\kappa_i \mid \mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu)\right] + \sum_{i=m+1}^{n}\log\left[\phi_{ST}(y_i \mid \mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu)\right],\end{aligned}$$

where $\mathbb{I}_i = 1$ if $y_i \leq \kappa_i$ and $\mathbb{I}_i = 0$ otherwise. As the observed log-likelihood function involves complex expressions, it is very difficult to work directly with $\ell(\boldsymbol{\theta} \mid \mathbf{y}_{obs})$, either for the ML estimation or for the local influence analysis. For the ST-CR, an EM-type algorithm has been developed by Lachos *et al.* (2022) to perform the ML estimation. In their estimation procedure, the unobserved y_i is considered as a realization of the latent unobservable variable $Y_i \sim \mathrm{ST}(\mathbf{x}_i^\top\boldsymbol{\beta}, \sigma^2, \lambda, \nu)$, $i = 1, \ldots, m$. The key was to consider the augmented data $\{\mathbf{y}_{obs}, y_1, \ldots, y_m, u_1, \ldots, u_n, t_1, \ldots, t_n\}$, that is, they treat the problem as if $\mathbf{y}_L = (y_1, \ldots, y_m)^\top$ were in fact observed. Hence, under the representation (2.2), the EM-type algorithm is applied to the complete-data log-likelihood $\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{y}_L, \mathbf{u}, \mathbf{t})$, given by

$$\begin{aligned}\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_{obs}, \mathbf{y}_L, \mathbf{u}, \mathbf{t}) = C - \frac{n}{2}\log\Gamma + \sum_{i=1}^{n}\log u_i - \frac{1}{2\Gamma}\sum_{i=1}^{n}u_i(y_i - \mathbf{x}_i^\top\boldsymbol{\beta} - \Delta t_i)^2\\ + \sum_{i=1}^{n}\log h(u_i \mid \nu),\end{aligned}$$

where C is a constant that does not depend on the parameter of interest $\boldsymbol{\theta}$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$ is a vector of independent half-normal random variables as defined in (2.1), $\mathbf{u} = (u_1, \ldots, u_n)^\top$ is a vector

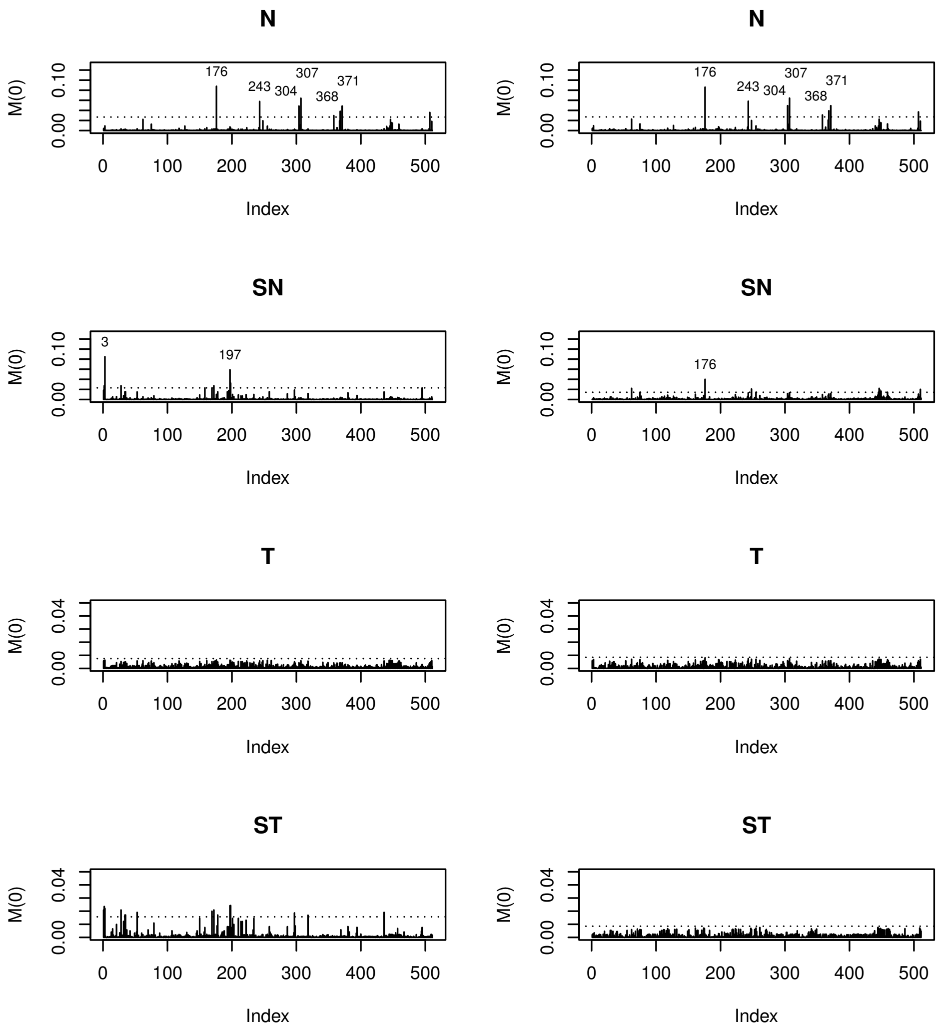

Figure 5: *Index plot of $M(0)$ under case-weights (left) and scale (right) perturbations for the four fitted models. The horizontal lines delimit the Lee and Xu (2004) benchmark for M(0) with $c^* = 3.5$.*

of independent gamma random variables as defined in (2.2), and $h(\cdot|\nu)$ denotes the gamma density with scale and shape parameters equal to $\nu/2$. For the current value $\boldsymbol{\theta}^{(k)}$, where the superscript (k) indicates the estimate of the related parameter at the stage k of the algorithm, the E-step of the EM-type algorithm requires the evaluation of the so-called Q-function:

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

where $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}$ means that the expectation is obtained using $\boldsymbol{\theta}^{(k)}$ instead of $\boldsymbol{\theta}$. Observe that the expression

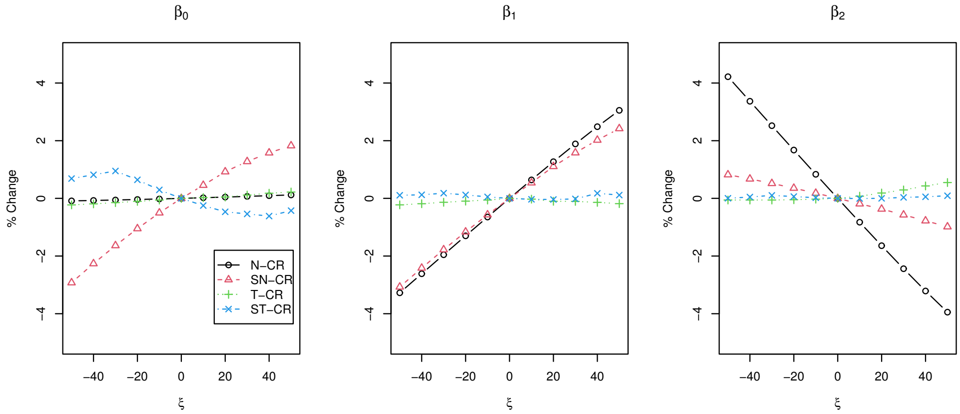

Figure 6: *Relative changes on the ML estimates of β_0, β_1 and β_2 when fitting a N-CR, SN-CR, T-CR, and ST-CR for different contamination of ξ on subject 176. % change = $100 \times ((\hat{\theta}(\xi) - \hat{\theta})/\hat{\theta})$, where $\hat{\theta}$ denotes the original estimate and $\hat{\theta}(\xi)$ the estimate for the contaminated data.*

of the Q-function is completely determined by the knowledge of the expectations

$$\mathcal{E}_{rsi}\left(\boldsymbol{\theta}^{(k)}\right) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[U_i T_i^r Y_i^s \mid y_{obs_i}\right], \quad r, s = 0, 1, 2.$$

Thus, ignoring constants that do not depend on the parameter of interest, the Q-function can be written in a synthetic form as follows:

$$\begin{aligned} Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}\right) = -\frac{n}{2}\log\Gamma - \frac{1}{2\Gamma}\sum_{i=1}^{n}\Big[&\mathcal{E}_{02i}\left(\boldsymbol{\theta}^{(k)}\right) - 2\mathcal{E}_{01i}\left(\boldsymbol{\theta}^{(k)}\right)\mathbf{x}_i^\top\boldsymbol{\beta} + \mathcal{E}_{00i}\left(\boldsymbol{\theta}^{(k)}\right)\left(\mathbf{x}_i^\top\boldsymbol{\beta}\right)^2 \\ &+\Delta^2\mathcal{E}_{20i}\left(\boldsymbol{\theta}^{(k)}\right) - 2\Delta\mathcal{E}_{11i}\left(\boldsymbol{\theta}^{(k)}\right) + 2\Delta\mathcal{E}_{10i}\left(\boldsymbol{\theta}^{(k)}\right)\mathbf{x}_i^\top\boldsymbol{\beta}\Big] \\ &+\sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[\log h(U_i \mid \nu) \mid y_{obs_i}\right], \end{aligned} \tag{2.5}$$

and depending if the observation is censored or not and by using known properties of conditional expectation, the expectations involved in the Q-function (E-step) will take specific, analytic, and closed forms Lachos *et al*. (2022). For instance, for censoring observations, these expectations take the form

$$\begin{aligned} \mathcal{E}_{rsi}\left(\boldsymbol{\theta}^{(k)}\right) &= \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[U_i T_i^r Y_i^s \mid Y_i \le \kappa_i\right] \\ &= \mathrm{E}_{\boldsymbol{\theta}^{(k)}}\left[Y_i^s \mathrm{E}\left[U_i\left[T_i^r \mid U_i, Y_i\right] \mid Y_i\right] \mid Y_i \le \kappa_i\right], \quad r, s = 0, 1, 2, \end{aligned}$$

which can be easily obtained by using Proposition 2 given in Lachos *et al*. (2022). The M-step requires the maximization of (2.5) concerning $\boldsymbol{\theta}$, which leads to closed-form equations. As our main focus here is not on the ML estimation, we refer the interested readers to see Lachos *et al*. (2022) for a detailed discussion of the EM-type algorithm for the ST-CR model. In this work, we suppose that the parameter ν associated with the mixture variable U is known – see, for instance, Lange *et al.* (1989), Berkane *et al*. (1994), Osorio *et al*. (2007), Lucas (1997) and Massuia *et al*. (2015) for an interesting discussion on the assumption of fixed degrees of freedom for the student's-t distribution. In this case, the Akaike information criterion in a grid of values of the degrees of freedom is recommended for determining the optimum value of ν.

Table 2: Means and standard deviations for study variables

	n	Wage	Age	Education	Youngkids	Experience
Full sample	753	-	42.54 (8.07)	12.29 (2.28)	0.24 (0.52)	10.63 (8.07)
Working women	428	4.18 (3.31)	41.97 (7.72)	12.66 (2.29)	0.14 (0.39)	13.04 (8.06)
Non-working women	325	-	43.28 (8.47)	11.80 (2.18)	0.37 (0.64)	7.46 (6.92)

3. Influence diagnostics

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 detecting influential observations. The first one is the case-deletion technique (Cook and Weisberg, 1982), 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 EM 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, $\mathbf{Y}_{c[i]} = (\mathbf{Y}_{obs[i]}, \mathbf{Y}_{L[i]}, \mathbf{U}_{[i]}, \mathbf{T}_{[i]})$ corresponds to the complete data with the i^{th} observation deleted. Let $\hat{\boldsymbol{\theta}}_{[i]} = (\hat{\boldsymbol{\beta}}_{[i]}^{\top}, \hat{\sigma}^2_{[i]}, \hat{\lambda}_{[i]})^{\top}$ be the maximizer of the function $Q_{[i]}(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}) = \mathrm{E}_{\hat{\boldsymbol{\theta}}}[\ell_c(\boldsymbol{\theta} \mid \mathbf{Y}_{c[i]}) \mid \mathbf{y}_{obs[i]}]$, where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^{\top}, \hat{\sigma}^2, \hat{\lambda})^{\top}$ is the ML estimates of $\boldsymbol{\theta}$. To assess the influence of the i^{th} case on $\hat{\boldsymbol{\theta}}$, we compare the difference between $\hat{\boldsymbol{\theta}}_{[i]}$ and $\hat{\boldsymbol{\theta}}$. If deletion of a case seriously influences the estimates, more attention should be paid to that case. In other words, if $\hat{\boldsymbol{\theta}}_{[i]}$ is fairly far from $\hat{\boldsymbol{\theta}}$ in some sense, then the i^{th} case could be considered influential. Since $\hat{\boldsymbol{\theta}}_{[i]}$ is needed for every case, the total computational burden involved can be quite heavy, so the following one-step approximation $\tilde{\boldsymbol{\theta}}_{[i]}$ is used to reduce the burden Cook and Weisberg, 1982:

$$\tilde{\boldsymbol{\theta}}_{[i]} = \hat{\boldsymbol{\theta}} + \left\{-\ddot{Q}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right)\right\}^{-1} \dot{Q}_{[i]}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right), \quad \text{for } i = 1, \ldots, n, \tag{3.1}$$

where

$$\dot{Q}_{[i]}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) = \left.\frac{\partial Q_{[i]}\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{and} \quad \ddot{Q}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) = \left.\frac{\partial^2 Q\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{3.2}$$

are the gradient vector and the Hessian matrix evaluated at $\hat{\boldsymbol{\theta}}$, respectively. In particular, the Hessian matrix is an essential element in the method developed by Zhu *et al.* (2001) (see also Zhu *et al.*, 2009) in order to obtain the measures for case-deletion diagnosis and for local influence of a specified perturbation scheme. These formulas can be obtained quite easily from equation (2.5). The elements of the Hessian matrix are presented in Appendix. The gradient vector, $\dot{Q}_{[i]}(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}) = (\dot{Q}_{[i]\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}), \dot{Q}_{[i]\sigma^2}(\hat{\boldsymbol{\theta}} \mid$

$\hat{\boldsymbol{\theta}}$), $\dot{Q}_{[i]\lambda}(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}})$), have elements given by:

$$\dot{Q}_{[i]\boldsymbol{\beta}}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) = \left.\frac{\partial Q_{[i]}\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{1+\hat{\lambda}^2}{\hat{\sigma}^2}\hat{E}_{1[i]},$$

$$\dot{Q}_{[i]\sigma^2}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) = \left.\frac{\partial Q_{[i]}\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial \sigma^2}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{2\hat{\sigma}^2}\left[(n-1) - \frac{1+\hat{\lambda}^2}{\hat{\sigma}^2}\hat{E}_{2[i]} + \frac{\hat{\lambda}\sqrt{1+\hat{\lambda}^2}}{\hat{\sigma}}\hat{E}_{3[i]}\right],$$

$$\dot{Q}_{[i]\lambda}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) = \left.\frac{\partial Q_{[i]}\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial \lambda}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{(n-1)\hat{\lambda}}{1+\hat{\lambda}^2} - \frac{\hat{\lambda}}{\hat{\sigma}^2}\hat{E}_{2[i]} + \frac{1+2\hat{\lambda}^2}{\hat{\sigma}\sqrt{1+\hat{\lambda}^2}}\hat{E}_{3[i]} - \hat{\lambda}\sum_{j\neq i}\mathcal{E}_{20j}\left(\hat{\boldsymbol{\theta}}\right),$$

where

$$\hat{E}_{1[i]} = \sum_{j\neq i}\left[\mathbf{x}_j\mathcal{E}_{01j}\left(\hat{\boldsymbol{\theta}}\right) - \mathcal{E}_{00j}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_j\mathbf{x}_j^\top\hat{\boldsymbol{\beta}} - \frac{\hat{\sigma}\hat{\lambda}}{\sqrt{1+\hat{\lambda}^2}}\mathbf{x}_j\mathcal{E}_{10j}\left(\hat{\boldsymbol{\theta}}\right)\right],$$

$$\hat{E}_{2[i]} = \sum_{j\neq i}\left[\mathcal{E}_{02j}\left(\hat{\boldsymbol{\theta}}\right) - 2\mathcal{E}_{01j}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_j^\top\hat{\boldsymbol{\beta}} + \mathcal{E}_{00j}\left(\hat{\boldsymbol{\theta}}\right)\left(\mathbf{x}_j^\top\hat{\boldsymbol{\beta}}\right)^2\right] \quad \text{and} \quad \hat{E}_{3[i]} = \sum_{j\neq i}\left[\mathcal{E}_{11j}\left(\hat{\boldsymbol{\theta}}\right) - \mathcal{E}_{10j}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_j^\top\hat{\boldsymbol{\beta}}\right].$$

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 $\hat{\boldsymbol{\theta}}$, we need to compare $\hat{\boldsymbol{\theta}}_{[i]}$ and $\hat{\boldsymbol{\theta}}$. If $\hat{\boldsymbol{\theta}}_{[i]}$ is far from $\hat{\boldsymbol{\theta}}$, in some sense, then the i^{th} case is regarded as influential. Based on the metric for measuring the distance between $\hat{\boldsymbol{\theta}}_{[i]}$ and $\hat{\boldsymbol{\theta}}$ proposed by Zhu *et al.* (2001), we consider here the following generalized Cook's distance:

$$\text{GD}_i = \left(\hat{\boldsymbol{\theta}}_{[i]} - \hat{\boldsymbol{\theta}}\right)^\top\left\{-\ddot{Q}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right)\right\}\left(\hat{\boldsymbol{\theta}}_{[i]} - \hat{\boldsymbol{\theta}}\right), \quad i = 1, \ldots, n. \tag{3.3}$$

Upon substituting (3.1) into (3.3), we obtain the following approximation of the generalized Cook's distance:

$$\text{GD}_i^1 = \dot{Q}_{[i]}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right)^\top\left\{-\ddot{Q}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right)\right\}^{-1}\dot{Q}_{[i]}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right).$$

Another measure of the influence of the i^{th} case is the following Q-distance function, similar to the likelihood distance LD_i (Cook and Weisberg, 1982), defined as:

$$\text{QD}_i = 2\left\{Q\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) - Q\left(\hat{\boldsymbol{\theta}}_{[i]} \mid \hat{\boldsymbol{\theta}}\right)\right\}. \tag{3.4}$$

We can compute an approximation of the likelihood displacement QD_i by substituting (3.1) into (3.4), resulting in the following approximation QD_i^1 of QD_i:

$$\text{QD}_i^1 = 2\left\{Q\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) - Q\left(\tilde{\boldsymbol{\theta}}_{[i]} \mid \hat{\boldsymbol{\theta}}\right)\right\}.$$

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

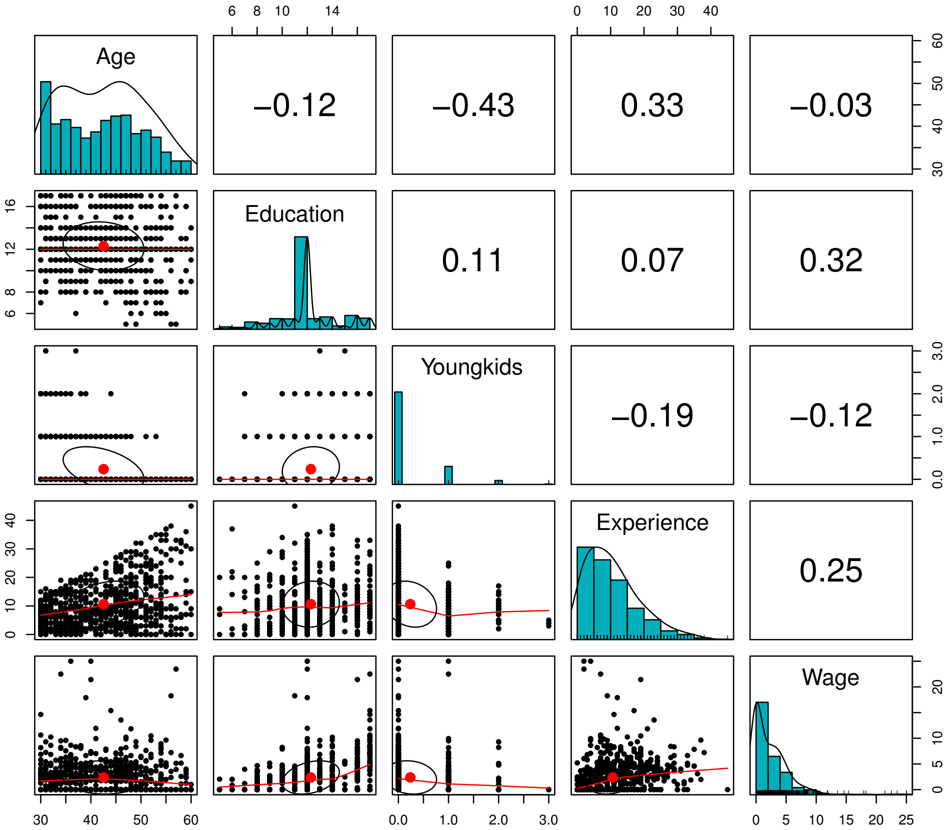

Figure 7: *Scatter-plots, histograms, and correlations for the indicated variables using Mroz data.*

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 $\boldsymbol{\Omega} \subset \mathbb{R}^g$. In general, we have $g = n$. Let $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \mathbf{Y}_c)$ be the complete-data log-likelihood of the perturbed model. We assume there is a $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ such that $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0 \mid \mathbf{Y}_c) = \ell_c(\boldsymbol{\theta} \mid \mathbf{Y}_c)$ for all $\boldsymbol{\theta}$. Let us define

$$Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \hat{\boldsymbol{\theta}}\right) = \mathrm{E}_{\hat{\boldsymbol{\theta}}}\left[\ell_c\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \mathbf{Y}_c\right) \mid \mathbf{y}_{obs}\right] \quad \text{and}$$

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}) = \arg\max_{\boldsymbol{\theta}} \left\{Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \hat{\boldsymbol{\theta}}\right)\right\} = \left(\hat{\boldsymbol{\beta}}(\boldsymbol{\omega})^\top, \hat{\sigma}^2(\boldsymbol{\omega}), \hat{\lambda}(\boldsymbol{\omega})\right)^\top.$$

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:

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

Following the approach of Cook (1986) and Zhu and 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

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

Then, it can be shown that

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

where $\ddot{Q}(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}})$ is as defined in (3.2).

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 $\zeta_1 \geq \cdots \geq \zeta_r > \zeta_{r+1} = \cdots = 0$ and orthonormal eigenvectors $\boldsymbol{\varepsilon}_k$, for $k = 1, \ldots, g$. Zhu and Lee (2001) proposed to inspect all eigenvectors corresponding to nonzero eigenvalues to capture more information, according to the following method:

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

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 and Poon (1999), Zhu and Lee (2001) considered using the following conformal normal curvature:

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

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 and 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 and 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 and 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 and 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. In this paper we use $c^* = 3.5$.

Table 3: ML estimates, SE and some information criteria for Mroz data

Parameter	N-CR		SN-CR		T-CR		ST-CR	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_0	−2.148	1.625	−4.282	1.408	−0.826	1.115	0.382	1.081
β_1	−0.169	0.027	−0.187	0.026	−0.164	0.021	−0.136	0.019
β_2	0.678	0.078	0.610	0.072	0.580	0.056	0.594	0.052
β_3	−2.705	0.408	−2.942	0.400	−2.612	0.323	−2.266	0.300
β_4	0.218	0.027	0.226	0.024	0.214	0.019	0.193	0.019
σ^2	18.996	0.697	31.812	1.513	7.643	0.733	11.507	1.997
λ	-	-	2.363	0.552	-	-	−1.178	0.243
ν	-	-	-	-	3.4	-	2.1	-
− log-like *	1440.062		1416.845		1373.032		**1367.074**	
AIC	2892.125		2847.690		2760.064		**2750.149**	
BIC	2919.869		2880.059		2792.433		**2787.141**	
CAIC	2925.869		2887.059		2799.433		**2795.141**	
HQIC	2902.813		2860.160		2772.534		**2764.400**	

* log-like: log-likelihood.

3.3. Perturbation schemes

We will evaluate the matrix $\boldsymbol{\nabla}$ under the following perturbation schemes for the ST-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; scale perturbation of σ^2, which can reveal individuals that are most influential, in the sense of the likelihood displacement on the scale structure; response perturbation of the response values, which can indicate observations with large influence on their own predicted values; and finally explanatory variables perturbation. For each perturbation scheme, we have the partitioned form:

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

where

$$\boldsymbol{\nabla}_{\boldsymbol{\beta}} = \left.\frac{\partial^2 Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)}, \quad \boldsymbol{\nabla}_{\sigma^2} = \left.\frac{\partial^2 Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \hat{\boldsymbol{\theta}}\right)}{\partial \sigma^2 \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)} \quad \text{and} \quad \boldsymbol{\nabla}_{\lambda} = \left.\frac{\partial^2 Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \hat{\boldsymbol{\theta}}\right)}{\partial \lambda \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)},$$

with $\boldsymbol{\nabla}_{\boldsymbol{\beta}} \in \mathbb{R}^{(p+1)\times g}$, $\boldsymbol{\nabla}_{\sigma^2} \in \mathbb{R}^{1\times g}$ and $\boldsymbol{\nabla}_{\lambda} \in \mathbb{R}^{1\times g}$.

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:

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

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:

$$\nabla_{\boldsymbol{\beta}} = \frac{1+\hat{\lambda}^2}{\hat{\sigma}^2}\left[\mathbf{X}^\top \mathrm{diag}\left\{\mathcal{E}_{01}\left(\hat{\boldsymbol{\theta}}\right)\right\} - \mathbf{X}^\top \mathrm{diag}\left\{\hat{\mathbf{A}}\right\} - \frac{\hat{\sigma}\hat{\lambda}}{\sqrt{1+\hat{\lambda}^2}}\mathbf{X}^\top \mathrm{diag}\left\{\mathcal{E}_{10}\left(\hat{\boldsymbol{\theta}}\right)\right\}\right],$$

$$\nabla_{\sigma^2} = -\frac{1}{2\hat{\sigma}^2}\left[\mathbf{1}_n^\top - \frac{1+\hat{\lambda}^2}{\hat{\sigma}^2}\hat{\mathbf{B}}^\top + \frac{\hat{\lambda}\sqrt{1+\hat{\lambda}^2}}{\hat{\sigma}}\hat{\mathbf{D}}^\top\right] \quad \text{and}$$

$$\nabla_{\lambda} = \frac{\hat{\lambda}}{1+\hat{\lambda}^2}\mathbf{1}_n^\top - \frac{\hat{\lambda}}{\hat{\sigma}^2}\hat{\mathbf{B}}^\top + \frac{1+2\hat{\lambda}^2}{\hat{\sigma}\sqrt{1+\hat{\lambda}^2}}\hat{\mathbf{D}}^\top - \hat{\lambda}\mathcal{E}_{20}^\top\left(\hat{\boldsymbol{\theta}}\right),$$

where $\mathbf{X}$ is a design matrix with rows $\mathbf{x}_i^\top$, $\mathcal{E}_{rs}(\hat{\boldsymbol{\theta}}) = (\mathcal{E}_{rs1}(\hat{\boldsymbol{\theta}}), \ldots, \mathcal{E}_{rsn}(\hat{\boldsymbol{\theta}}))^\top$, $r, s = 0, 1, 2$ and $\mathbf{1}_n$ is a $n \times 1$ vector of ones. $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$ and $\hat{\mathbf{D}}$ are n−dimensional vectors with coordinates $\hat{A}_i = \mathcal{E}_{00i}(\hat{\boldsymbol{\theta}})\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}$, $\hat{B}_i = \mathcal{E}_{02i}(\hat{\boldsymbol{\theta}}) - 2\mathcal{E}_{01i}(\hat{\boldsymbol{\theta}})\mathbf{x}_i^\top\hat{\boldsymbol{\beta}} + \mathcal{E}_{00i}(\hat{\boldsymbol{\theta}})(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}})^2$ and $\hat{D}_i = \mathcal{E}_{11i}(\hat{\boldsymbol{\theta}}) - \mathcal{E}_{10i}(\hat{\boldsymbol{\theta}})\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}$, 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 $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top \in \mathbb{R}^n$. Moreover, the perturbed Q-function is as in (2.5), switching $\sigma^2(\omega_i)$ and $\hat{\boldsymbol{\theta}}$ with σ^2 and $\boldsymbol{\theta}^{(k)}$, respectively. The matrix $\nabla_{\boldsymbol{\omega}_0}$ has the following elements:

$$\nabla_{\boldsymbol{\beta}} = \frac{1+\hat{\lambda}^2}{\hat{\sigma}^2}\left[\mathbf{X}^\top \mathrm{diag}\left\{\mathcal{E}_{01}\left(\hat{\boldsymbol{\theta}}\right)\right\} - \mathbf{X}^\top \mathrm{diag}\left\{\hat{\mathbf{A}}\right\} - \frac{\hat{\sigma}\hat{\lambda}}{2\sqrt{1+\hat{\lambda}^2}}\mathbf{X}^\top \mathrm{diag}\left\{\mathcal{E}_{10}\left(\hat{\boldsymbol{\theta}}\right)\right\}\right],$$

$$\nabla_{\sigma^2} = \frac{1}{2\hat{\sigma}^2}\left[\frac{1+\hat{\lambda}^2}{\hat{\sigma}^2}\hat{\mathbf{B}}^\top - \frac{\hat{\lambda}\sqrt{1+\hat{\lambda}^2}}{2\hat{\sigma}}\hat{\mathbf{D}}^\top\right] \quad \text{and}$$

$$\nabla_{\lambda} = -\frac{\hat{\lambda}}{\hat{\sigma}^2}\hat{\mathbf{B}}^\top + \frac{1+2\hat{\lambda}^2}{2\hat{\sigma}\sqrt{1+\hat{\lambda}^2}}\hat{\mathbf{D}}^\top.$$

3.3.3. Response perturbation

A perturbation of the response variables Y_i, $i = 1, \ldots, n$, can be introduced by replacing Y_{obs_i} by $Y_{obs_i}(\omega_i) = Y_{obs_i} + \omega_i \mathrm{S}_y$, where S_y is a scale factor that can represent the standard deviation of the censored response. Now substituting $Y_{obs_i}(\omega_i)$ into (2.4), we can write the perturbed model as:

$$Y_{obs_i}(\omega_i) = \begin{cases} \kappa_i(\omega_i) & \text{if } \; Y_i \leq \kappa_i, \\ Y_i(\omega_i) & \text{if } \; Y_i > \kappa_i, \end{cases}$$

where $\kappa_i(\omega_i) = \kappa_i - \omega_i \mathrm{S}_y$ and $Y_i(\omega_i) = Y_i - \omega_i \mathrm{S}_y$. Hence, the perturbed Q-function follows (2.5), with $\mathcal{E}_{rsi}(\boldsymbol{\theta}^{(k)}) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i T_i^r Y_i^s \mid y_{obs_i}]$ replaced by $\mathcal{E}_{rsi}(\boldsymbol{\theta}^{(k)}, \omega_i) = \mathrm{E}_{\boldsymbol{\theta}^{(k)}}[U_i T_i^r Y_i^s(\omega_i) \mid y_{obs_i}(\omega_i)]$. Under this perturbation scheme, the vector $\boldsymbol{\omega}_0$, representing no perturbation, is given by $\boldsymbol{\omega}_0 = \mathbf{0} \in \mathbb{R}^n$ and $\nabla_{\boldsymbol{\omega}_0}$

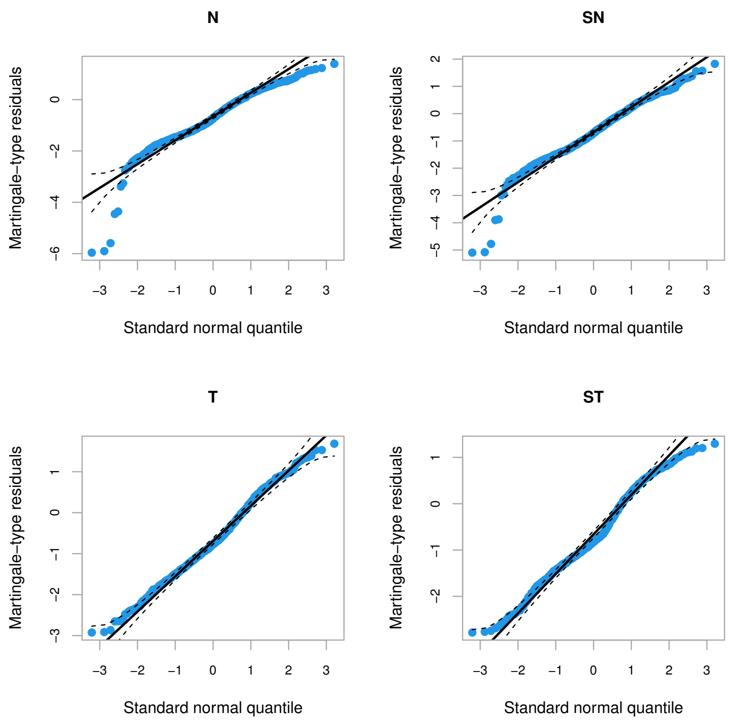

Figure 8: *Martingale-type residuals under the N-CR, SN-CR, T-CR and ST-CR models for Mroz data.*

has the following elements:

$$\nabla_\beta = \frac{1+\hat{\lambda}^2}{\hat{\sigma}^2} \mathrm{S}_y \mathbf{X}^\top \mathrm{diag}\left\{\mathcal{E}_{00}\left(\hat{\boldsymbol{\theta}}\right)\right\},$$

$$\nabla_{\sigma^2} = \frac{\mathrm{S}_y}{\hat{\sigma}^2}\left[\frac{1+\hat{\lambda}^2}{\hat{\sigma}^2}\left(\mathcal{E}_{01}^\top\left(\hat{\boldsymbol{\theta}}\right) - \hat{\mathbf{A}}^\top\right) - \frac{\hat{\lambda}\sqrt{1+\hat{\lambda}^2}}{2\hat{\sigma}}\mathcal{E}_{10}^\top\left(\hat{\boldsymbol{\theta}}\right)\right], \quad \text{and}$$

$$\nabla_\lambda = -\frac{2\hat{\lambda}}{\hat{\sigma}^2}\mathrm{S}_y\left[\mathcal{E}_{01}^\top\left(\hat{\boldsymbol{\theta}}\right) - \hat{\mathbf{A}}^\top\right] + \frac{1+2\hat{\lambda}^2}{\hat{\sigma}\sqrt{1+\hat{\lambda}^2}}\mathrm{S}_y\mathcal{E}_{10}^\top\left(\hat{\boldsymbol{\theta}}\right).$$

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 $\mathbf{x}_{i\omega}^\top = \mathbf{x}_i^\top + \omega_i \mathrm{S}_t \mathbf{1}_t^\top$, S_t is a scale factor that can represent the standard deviation of the t^{th} explanatory variable and $\mathbf{1}_t^\top = (0, \ldots, 1, \ldots, 0)$ is a $1 \times p$ vector with 1 in the t^{th} column, $t = 1, \ldots, p$. Hence, this case covers situations where x is measured with error. The perturbed Q-function is as in (2.5), switching $\mathbf{x}_{i\omega}^\top$ with $\mathbf{x}_i^\top$ and the no perturbation case is obtained by

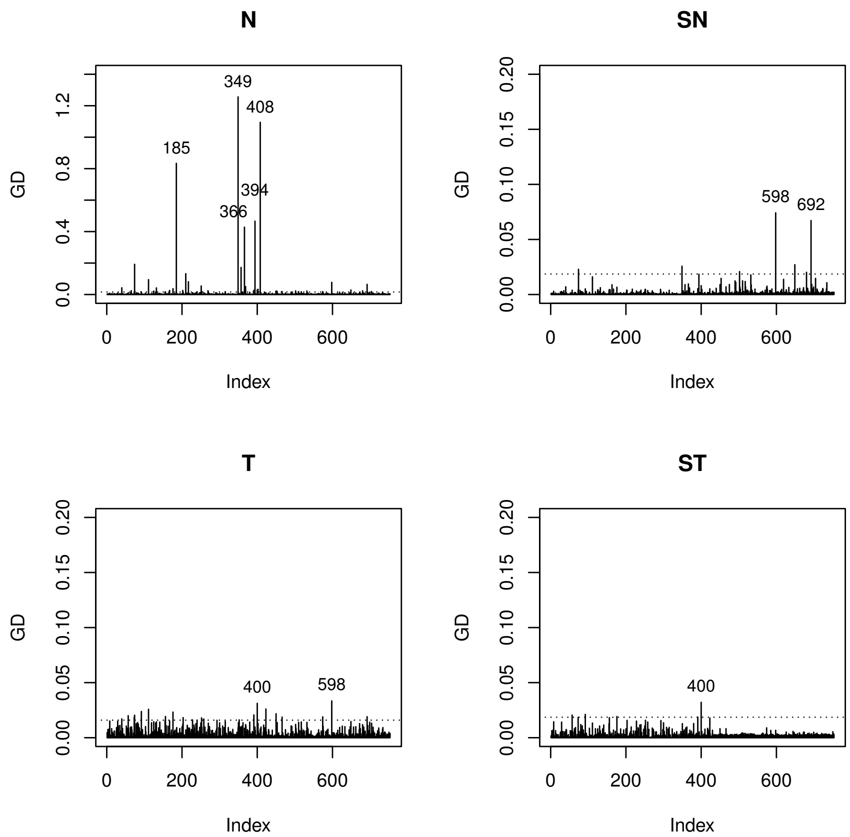

Figure 9: *GD_i under the N-CR, SN-CR, T-CR and ST-CR models for Mroz data.*

taking $\boldsymbol{\omega}_0 = \mathbf{0}$. Under this perturbation scheme, $\boldsymbol{\nabla}_{\boldsymbol{\omega}_0}$ has the following elements:

$$\boldsymbol{\nabla}_{\beta} = \frac{1+\hat{\lambda}^2}{\hat{\sigma}^2} \mathrm{S}_t \mathbf{1}_t \left[\mathcal{E}_{01}^{\top}\left(\hat{\boldsymbol{\theta}}\right) - 2\hat{\mathbf{A}}^{\top} - \frac{\hat{\sigma}\hat{\lambda}}{\sqrt{1+\hat{\lambda}^2}} \mathcal{E}_{10}^{\top}\left(\hat{\boldsymbol{\theta}}\right) \right],$$

$$\boldsymbol{\nabla}_{\sigma^2} = -\frac{\mathrm{S}_t}{\hat{\sigma}^2} \mathbf{1}_t^{\top} \hat{\boldsymbol{\beta}} \left[\frac{1+\hat{\lambda}^2}{\hat{\sigma}^2} \left[\mathcal{E}_{01}^{\top}\left(\hat{\boldsymbol{\theta}}\right) - \hat{\mathbf{A}}^{\top} \right] - \frac{\hat{\lambda}\sqrt{1+\hat{\lambda}^2}}{2\hat{\sigma}} \mathcal{E}_{10}^{\top}\left(\hat{\boldsymbol{\theta}}\right) \right], \quad \text{and}$$

$$\boldsymbol{\nabla}_{\lambda} = \frac{\mathrm{S}_t}{\hat{\sigma}} \mathbf{1}_t^{\top} \hat{\boldsymbol{\beta}} \left[\frac{2\hat{\lambda}}{\hat{\sigma}} \left(\mathcal{E}_{01}^{\top}\left(\hat{\boldsymbol{\theta}}\right) - \hat{\mathbf{A}}^{\top} \right) - \frac{1+2\hat{\lambda}^2}{\sqrt{1+\hat{\lambda}^2}} \mathcal{E}_{10}^{\top}\left(\hat{\boldsymbol{\theta}}\right) \right].$$

Note that it is impossible to give details for all the perturbation schemes that are of interest. However, as long as we can find an appropriate $\boldsymbol{\omega}$, and as long as the perturbed complete-data log-likelihood function $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \mathbf{Y}_c)$ is smooth enough, so that the required derivatives in the diagnostic measures are all well defined, we can conduct the local influence analysis without much difficulty. In the next section, two applications to real data are presented in order to illustrate the performance of the developed methodology. R code for analyzing the application may be downloaded from the third

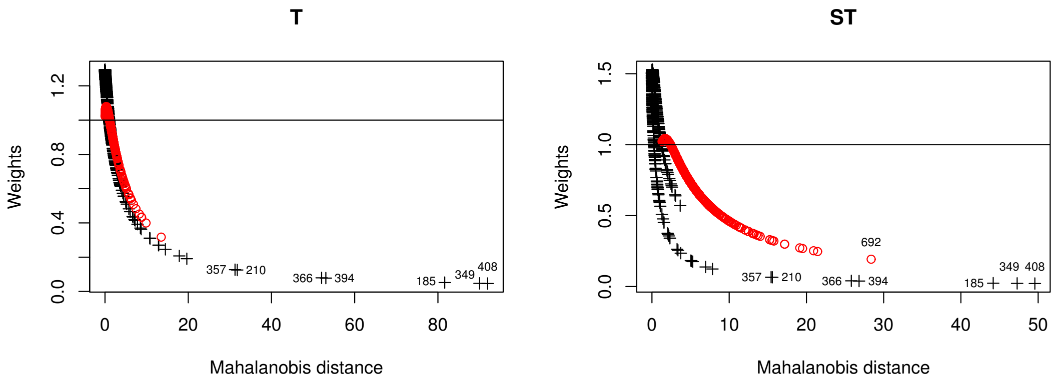

Figure 10: *Estimated u_i under the T-CR and ST-CR models for Mroz data. The red ball-shaped points represent the imputed values for the censored values.*

author's homepage, with website address `https://github.com/hlachos/skewt-censored`.

4. Applications

We illustrate the proposed methods with the analysis of two real data sets previously analyzed using CR models.

4.1. Letter-name fluency data

For illustration purposes, we apply the techniques presented so far to the letter-name fluency (LNF) test in Peruvian students, which is a standardized, individually administered test that provides a measure of letter-name knowledge (LNK) and spelling abilities. In this test, teachers administer timed 1-minute fluency assessments to children, and then compare the results with established norms in order to determine how the students are performing in this task and if they are at risk for future academic problems. LNF is a continuous right censored variable related to the average of letters read correctly in an interval of time and not a discrete variable.

The data were originally reported from the early grade reading assessment (RTI-FDA, 2008). It contains 511 students, with 479 uncensored and 32 right censored observations (6.26%). Consequently, if we are interested in the mean of the LNF response for one group, this quantity could be underestimated due to the presence of censored observations. For that reason, a censored regression model able to take into account observation lying below or above a threshold could be more appropriate for estimating the true mean of the LNF response for different groups of interest. Lachos *et al*. (2022) have analyzed this data set and pointed out the following findings: 1) the mean and the standard deviation of the censored observations are higher in comparison with uncensored (causing the estimated mean for LNF to be underestimated); 2) some degree of right skewness and kurtosis on the response variable (reveling a departure from the normal distribution); and 3) the data are better suited to the ST-CR model with heavy tails than all its competitors (N-CR, SN-CR and T-CR). Here, we revisit this data set with the aim of applying the Zhu and Lee (2001) local influence approach. As in Lachos *et al*. (2022), the proposed censored model is given by:

$$Y_i = \beta_0 + \beta_1 \ \text{Zone}_i + \beta_2 \ \text{Grade}_i + \beta_3 \ \text{Gender}_i + \sigma\epsilon_i, \quad i = 1, \ldots, 511,$$

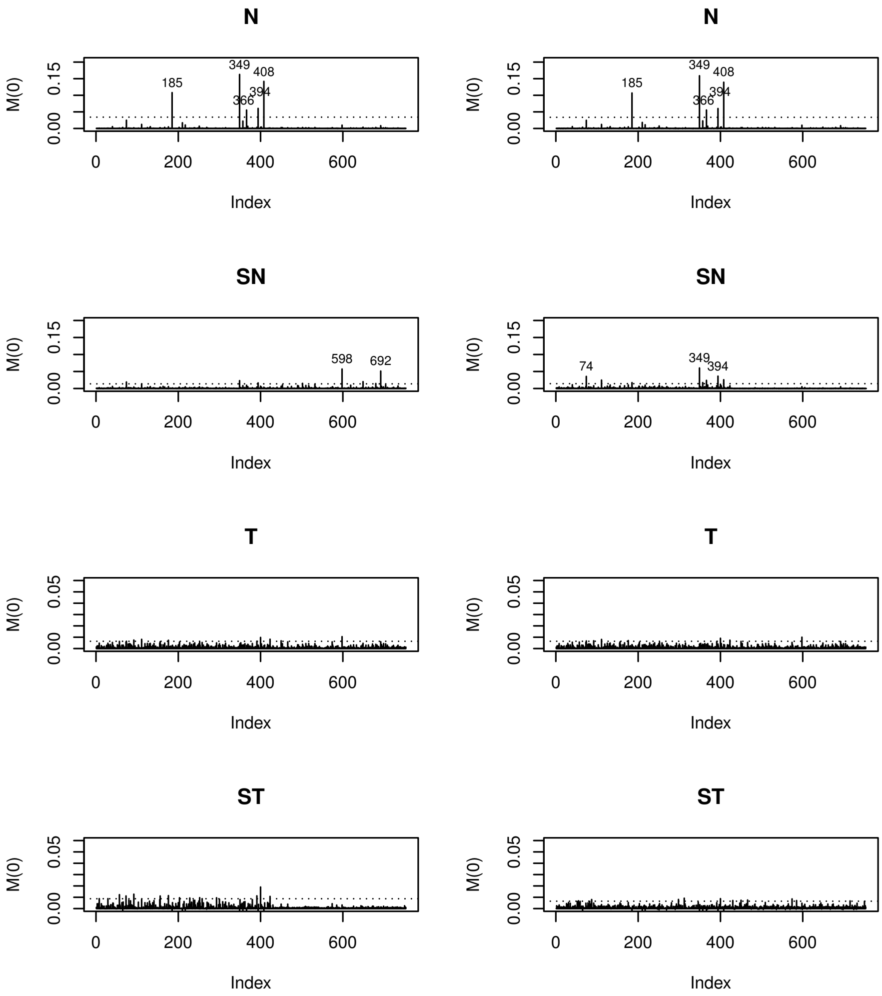

Figure 11: *Index plot of $M(0)$ under case-weights (left) and scale (right) perturbations for the four fitted models for Mroz data. The horizontal lines delimit the Lee and Xu (2004) benchmark for M(0) with $c^* = 3.5$.*

where Y_i is the number of correctly letters read by the i^{th} student in one minute; the zone where the respondent lives (0 = urban, 1 = rural); grade (0 = 2nd grade, 1 = 3rd grade) and gender (0 = male, 1 = female). For the summary of the response variable in the presence (censored) and absence (uncensored) of censoring, we refer the interested reader to see Table 4 in Lachos *et al*. (2022). Based on the results given in Lachos *et al*. (2022), where the covariate gender is non-significant, we will carry out further analyzes, without this covariate.

The application is organized as follows. First, we have fitted the N-CR, SN-CR, T-CR and ST-CR for the LNF data set. Although not being formal tests, we compare the four models by inspecting some selected information criteria and the plots of the Martingale-type residuals. After, in order to identify influential observations, we generate graphs of the generalized Cook's distance (GD_i), as explained in

Section 3.1. Next, we have identified influential observations for the LNF data set using $M(0)$ from the conformal curvature $B_{f_Q,\mathbf{d}_l}$ and the first two perturbation schemes described in Section 3.3. Lastly, we assess the robustness of the ST-CR model by studying the influence of a single outlying observation on the ML estimate of $\boldsymbol{\theta}$.

Figure 1 presents the boxplot of the number of correctly letters read in relation to zone and grade. It appears that the median of the response variable is slightly higher for the urban schools and 3rd grade. Also from this figure, it is possible to identify that the median is higher for the urban, compared to the rural schools, as well as, 3rd grade, compared to 2nd grade.

Table 1 contains the ML estimates for the parameters of the four models, the approximate standard errors (SE) based on the empirical information matrix (Lachos *et al.*, 2022) and some information criteria. The two covariates (zone and grade) are significant in all fitted models. There is agreement in the signs of the $\beta_0 - \beta_2$ between the models, although the estimated coefficients are a little different. Moreover, the SEs for β_1 and β_2 under the ST-CR model are smaller than the other models, indicating that this model produces more accurate ML estimates. According to Lange *et al.* (1989), we choose the value of $\boldsymbol{\nu}$ by maximizing the profile likelihood function as follows: Suppose $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, where $\boldsymbol{\theta}_1$ is of interest and $\boldsymbol{\theta}_2$ is a nuisance parameter, then the profile log-likelihood of $\boldsymbol{\theta}_1$ is $l_p(\boldsymbol{\theta}_1) = \max_{\boldsymbol{\theta}_2} l(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Using this procedure, we found $\nu = 3.5$ for the T-CR model, and we found $\nu = 4.8$ for the ST-CR model. The estimated values found of ν are small, indicating the lack of adequacy of the N-CR model for the LNF data.

We now compare the models by inspecting some information criteria. Four criteria were selected: The Akaike information criteria (AIC, $-2\ell(\hat{\boldsymbol{\theta}}) + 2p$), Bayesian information criterion (BIC, $-2\ell(\hat{\boldsymbol{\theta}}) + \log(n)p$), consistent AIC (CAIC, $-2\ell(\hat{\boldsymbol{\theta}}) + (\log(n)+1)p$) and Hannan-Quinn criterion (HQIC, $-2\ell(\hat{\boldsymbol{\theta}}) + 2\log(\log(n))p$), where p is the number of free parameters in the model. A lower value of these measures indicates that a closer fit of the model to the data. The results are also given in Table 1, where we observe that the ST-CR model outperforms all its competitors (N-CR, SN-CR and T-CR).

In order to study departures from the error assumption as well as the presence of outliers, we analyzed the transformation of the martingale type residual, denoted by r_{MT_i}, proposed by Barros *et al.* (2010) for censored models. These residuals are defined by

$$r_{MT_i} = \operatorname{sign}(r_{M_i}) \sqrt{-2\left[r_{M_i} + \rho_i \log(\rho_i - r_{M_i})\right]},$$

for $i = 1, \ldots, n$, where $r_{M_i} = \rho_i + \log(S(y_i; \hat{\boldsymbol{\theta}}))$ is the martingale residual proposed by Ortega *et al.* (2003) – see more details in Therneau *et al.* (1990), with $\rho_i = 0, 1$ indicating whether the i^{th} observation is censored or not, respectively, $\operatorname{sign}(r_{M_i})$ denoting the sign of r_{M_i} and $S(y_i; \hat{\boldsymbol{\theta}}) = P_{\hat{\boldsymbol{\theta}}}(Y_i > y_i)$ representing the survival function evaluated at y_i and the EM estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. As observed by Ortega *et al.* (2003) these residuals can be used to assess the quality of the model fit. In our case, we generate envelopes based on these residuals for the four models, which are shown in Figure 2. It can be seen that the skew-*t* distribution accommodates the observations in a better way than the other models.

Figure 3 shows the approximate generalized Cook's distance GD_i under the N-CR, SN-CR, T-CR and ST-CR models. A high value for GD_i indicates that the i^{th} observation has a high impact on the maximum-likelihood estimate of the parameters. We can see from these figures that observations 176, 243, 304, 307, 368, 371 and 507 appear to be outliers under the N-CR fit; under the SN-CR fit, we can see observations 3 and 197; on the other hand, no observation is detected under the T-CR and ST-CR fits, showing the robustness of these two heavy-tailed models.

When we use distributions with tails heavier than the normal and skew–normal ones the EM algorithm allows to accommodate discrepant observations attributing to them small weights in the estimation procedure. In Figure 4, we present the Mahalanobis distance, given by $d_i^2 = (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}})/\hat{\sigma}^2$ vs, the estimated weights $u_i = \mathcal{E}_{00i}(\hat{\boldsymbol{\theta}})$, for $i = 1, \ldots, 511$, considering the T-CR and ST-CR models. The weights for the normal and skew–normal distribution are indicated in Figure 4 as a continuous line. Note from these figures that for the T and ST distributions u_i is inversely proportional to the Mahalanobis distance. Therefore, the student-*t* and skew-*t* distributions may naturally attribute different weights to each observation and consequently control the influence of a single observation on the parameter estimates. These results agree with similar considerations, presented in Osorio *et al.* (2007), in a symmetric context.

Next, we conduct a local influence study based on $M(0)$ according to Sections 3.2 and 3.3. Here, we used the criterion $M(0)_l > \overline{M(0)} + 3.5SM(0)$, $l = 1, \ldots, 511$, to discriminate whether an observation is influential. The Figure 5 displays the results for the N-CR, SN-CR, T-CR, and ST-CR models under the case-weight perturbation and scale perturbation. From these figures, it is noted that observations 176, 243, 304, 307, 368, and 371 are identified as influential in both perturbations considered under the N-CR model. Two observations (3 and 197) are identified as influential in the case-weight perturbation and one (176) in the scale perturbation under the SN-CR model. No observation was identified as influential under the T-CR and ST-CR models in both perturbations, assuming $c^* = 3.5$. As expected, the influence of such observations is reduced when we consider distributions with heavier tails than the normal or skew-normal ones. For this data set the student's-*t* and skew-*t* models accommodate slightly better the influential observations.

The robustness of the ST-CR model can be studied through the influence of a single outlying observation on the ML estimate of $\boldsymbol{\theta}$. In particular, we can asses how much the ML estimates of $\boldsymbol{\theta}$ influences by a change of ξ units in a single observation Y_i. We replace a single observation y_i by $y_i(\xi) = y_i + \xi$, and record the relative change in the estimates $((\hat{\theta}(\xi) - \hat{\theta})/\hat{\theta})$, where $\hat{\theta}$ denotes the original estimate and $\hat{\theta}(\xi)$ the estimate for the contaminated data. In this example, we contaminated the observation on subject 176 and varied ξ between -50 and 50 by increments of 10. In Figure 6, we have presented the results of relative changes of the estimate $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$, for different contamination of ξ, under N-CR, SN-CR, T-CR, and ST-CR. As expected, the T-CR and ST-CR models are less adversely affected by variations of ξ than the N-CR and SN-CR model.

4.2. Married women's labour supply

The second application deals with left censored econometric data (Cameron and Trivedi, 2005, Section 16.2, Page 530). We use a real data set previously analyzed by Mroz (1987) consisting of observations on 753 married white women for 21 variables, between 30 and 60 years old in 1975 (interview year: 1976), with 428 (56.84%) of them working at some time during that year and therefore 325 (43.16%) of them have an average hourly wage equal to zero. Thus, we can consider these last ones as left censored with $\kappa_i = 0$, $i = 1, \ldots, n$. We name the data set as Mroz data hereafter. The response variable (Y) is the wife's average hourly wage (in US dollar for the year 1975) and the considered covariates are: Wife's age in years (age), wife's education in years (education), number of children smaller than 6 years old in household (youngkids), number of children between ages 6 and 18 in household (oldkids), and years of wife's previous labor market experience (experience). The sample characteristics are presented in Table 2. Oldkids was hidden, as its mean (1.35) and standard deviation (1.32) are the same for both the full sample and the working woman sample. The data used may be obtained from the `AER` package (see, Kleiber and Zeileis, 2008) with the command `data("PSID1976")` in the `R` software (see, R Core Team, 2022).

Figure 7 shows the scatter-plot matrix, histograms, and correlations for assessing the relationships between variables, simultaneously. Initially, the histogram of wages confirms the asymmetric behavior of this variable. Furthermore, note that the wife's average hourly wage increases as the wife's education in years increases and as her experience in years increases. The largest correlation between the response variable and covariates is 0.32, detected between education and wage, whereas the largest correlation between covariates is −0.43, detected between age and youngkids.

Barros *et al.* (2018) have analyzed this data set with the same covariates and pointed out that the covariate oldkids can be considered non-significant under the Tobit-normal and Tobit-t models. Based on these results, we will carry out the analyzes again, without this covariate, using the left censored N-CR, SN-CR, T-CR and ST-CR models for comparison purposes. As in the first application, our goal is applying the proposed diagnostics techniques, illustrating the robustness of the ST-CR model in relation to its competitors. Therefore, the proposed censored model for the Mroz data is given by

$$Y_i = \beta_0 + \beta_1 \ \text{Age}_i + \beta_2 \ \text{Education}_i + \beta_3 \ \ \text{Youngkids}_i + \beta_4 \ \text{Experience}_i + \sigma\epsilon_i, \quad i = 1, \ldots, 753.$$

The second application is organized as follows. First, we have fitted the N-CR, SN-CR, T-CR and ST-CR for the Mroz data set. Next, we compare the four models by inspecting some selected information criteria. After, in order to identify influential observations, we generate graphs of the generalized Cook's distance (GD_i), as explained in Section 3.1. Lastly, we have identified influential observations for the Mroz data set using $M(0)$ from the conformal curvature $B_{f_Q,\mathbf{d}_l}$ and the case-weight and scale perturbation schemes described in the Section 3.3.

The ML estimates and the corresponding approximate standard errors (SE) based on the empirical information matrix (Lachos *et al.* (2022)) for the coefficients are shown in Table 3. This table also contains some information criteria. The four covariates are significant in all fitted models. The intercept (β_0) is significant only under the SN-CR fit. There is agreement in the signs of the estimated regression coefficients ($\beta_1 - \beta_4$) between the models, and they are relatively closer, compared to the result found in the first application. We noted that the sign of the estimated asymmetry coefficient (λ) has changed in the ST-CR compared to the SN-CR fit, remembering in the ST-CR fit both asymmetry and kurtosis are modeled simultaneously, while in the SN-CR the latter is not taken into account. Again, under the ST-CR model, the SEs for all βs are smaller than the other models. This reinforces the indication that this model produces more accurate maximum likelihood estimates than the N-CR, SN-CR and T-CR models. For this application, we found $\nu = 3.4$ for the T-CR model, and we found $\nu = 2.1$ for the ST-CR model. The estimated values found of ν are again small, indicating the lack of adequacy of the N-CR model for the Mroz data. Finally, we note that the lowest values of the selected information criteria, highlighted in the table, occurred for the ST-CR model. The corresponding plot of the Martingale-type residuals including a simulated envelope are shown in Figure 8.

Figure 9 shows the approximate generalized Cook's distance GD_i under the N-CR, SN-CR, T-CR, and ST-CR models for Mroz data. A diagnostics analysis based on this measure highlights strongly the observations 185, 349, 366, 394, and 408 under the N-CR fit; for the other models, we have fewer highlighted observations: 598 and 692 for SN-CR, 400 and 598 for T-CR, and 400 for ST-CR. For example, the observation 408 corresponds to one married white woman with 36 years, 12 years of education, 1 kid smaller than 6 years old, 4 years of previous labor market experience, and an average hourly wage of 25 dollars; on the other hand, in contrast, the observation 400 corresponds to one married white women with 38 years, 15 years of education, 2 kids smaller than 6 years old, 17 years of previous labor market experience and an average hourly wage of just 5.1 dollars.

The Mahalanobis distance $d_i^2 = (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}})/\hat{\sigma}^2$ vs, the estimated weights $u_i = \mathcal{E}_{00i}(\hat{\boldsymbol{\theta}})$, for $i = 1, \ldots, 753$, considering the T-CR and ST-CR models are shown in Figure 10. The corresponding weights for the normal and skew–normal distribution are indicated as a continuous line in this figure. As in the first application, it is observed that for the T and ST distributions, u_i is inversely proportional to the Mahalanobis distance. This characteristic of these distributions allows controlling the influence of a single observation on the parameter estimates.

Finally, we conduct a local influence study based on $M(0)$ according to the Sections 3.2 and 3.3. Here, we used the criterion $M(0)_l > \overline{M(0)} + 3.5SM(0)$, $l = 1, \ldots, 753$, to discriminate whether an observation is influential. The Figure 11 displays the results for the N-CR, SN-CR, T-CR and ST-CR models under the case-weight perturbation (left side) and scale perturbation (right side). From these figures, it is noted that five observations (185, 349, 366, 394 and 408) are identified as influential under the N-CR model. Note that the observations that were considered influential under the case weight perturbation also were detected under the scale perturbation under N-CR. Besides, the influence of those observations seems not to change under both cases. Others two observations (598 and 692) are identified as influential in the case-weight perturbation and three (74, 349 and 394) in the scale perturbation under the SN-CR model. Again, no observation was identified as influential under the T-CR and ST-CR models in both perturbations, making it possible to conclude that for this data set the student's-*t* and skew-*t* models accommodates slightly better the influential observations.

After all the analyzes performed, the final fitted model selected from our analysis is given by

$$\begin{aligned}\hat{Y}_i \sim \mathrm{ST}\Big(&\hat{\mu} = 0.382 - 0.136\ \mathrm{Age}_i + 0.594\ \ \mathrm{Education}_i - 2.266\ \mathrm{Youngkids}_i + 0.193\ \mathrm{Experience}_i,\\ &\hat{\sigma}^2 = 11.507, \hat{\lambda} = -1.178, \hat{\nu} = 2.1\Big),\end{aligned}$$

$i = 1, \ldots, 753$. The age and youngkids covariates have a negative effect on estimating wife's average hourly wage while the education and experience have a positive effect. The coefficients of the ST-CR model are interpreted in the similar manner to standard regression coefficients. The expected wife's average hourly wage (in 1975 dollars) changes according to the coefficient for each unit increased in the corresponding covariate.

5. Conclusions

Diagnostic analysis is an efficient way to detect influential observations and is an important step in data analysis following parameter estimation. Thus, we believe that this article is a necessary supplement to the work by Lachos *et al.* (2022) by proposing influence diagnostic tools for detecting influential observations in ST-CR models. The diagnostic analysis is based on local influence techniques presented in Zhu and Lee (2001) and Zhu *et al.* (2001), for which explicit expressions are obtained for $\boldsymbol{\nabla}$ under different perturbation schemes, and the Hessian matrix $\ddot{Q}(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}})$. Further, we applied our method to two real data sets (left and right-censored) to illustrate how the procedure developed can be used to evaluate model assumptions, identify outliers and obtain robust parameter estimates. The method proposed in this paper is implemented in `R`, and the code is available for download in the folder Diagnostics from `GitHub` repository (`https://github.com/hlachos/skewt-censored`). A short-term project includes the development of an efficient and reliable `R` package.

Finally, some extensions of the current work include the multivariate ST-CR model (Valeriano *et al.*, 2023) a likelihood-based treatment of ST regressions with informative censoring Lachos *et al.* (2021) or the finite mixtures of ST-CR models (Zeller *et al.*, 2019). An in-depth investigation of such extensions is beyond the scope of the present paper, but certainly an interesting topic for future research.

Acknowledgements

We thank the associate editor and four anonymous referees for their important comments and suggestions which lead to an improvement of this paper. The research of Marcos S. Oliveira was supported by Grant no. 401418/2022-7 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – Brazil. The research of Daniela C. R. Oliveira was supported by Grant no. 401373/2022-3 from CNPq – Brazil. Victor Lachos acknowledges the partial financial support from UConn - CLAS's Summer Research Funding Initiative 2023.

Appendix A: Hessian matrix

The Hessian matrix $\ddot{Q}(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}})$ is obtained from the second-order partial derivatives of $Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}})$, given in (2.5), evaluated at $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\sigma}^2, \hat{\lambda})^\top$, which has elements given by

$$
\begin{aligned}
\ddot{Q}_{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) &= \left.\frac{\partial^2 Q\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1+\hat{\lambda}^2}{\hat{\sigma}^2}\sum_{i=1}^n \mathcal{E}_{00i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i\mathbf{x}_i^\top, \\
\ddot{Q}_{\boldsymbol{\beta}\sigma^2}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) &= \left.\frac{\partial^2 Q\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial\boldsymbol{\beta}\partial\sigma^2}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\
&= -\frac{1+\hat{\lambda}^2}{\hat{\sigma}^2}\left[\frac{1}{\hat{\sigma}^2}\sum_{i=1}^n\left(\mathbf{x}_i\mathcal{E}_{01i}\left(\hat{\boldsymbol{\theta}}\right) - \mathcal{E}_{00i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right) - \frac{1}{2\hat{\sigma}}\frac{\hat{\lambda}}{\sqrt{1+\hat{\lambda}^2}}\sum_{i=1}^n \mathbf{x}_i\mathcal{E}_{10i}\left(\hat{\boldsymbol{\theta}}\right)\right], \\
\ddot{Q}_{\boldsymbol{\beta}\lambda}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) &= \left.\frac{\partial^2 Q\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial\boldsymbol{\beta}\partial\lambda}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{2\hat{\lambda}}{\hat{\sigma}^2}\sum_{i=1}^n\left(\mathbf{x}_i\mathcal{E}_{01i}\left(\hat{\boldsymbol{\theta}}\right) - \mathcal{E}_{00i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right) - \frac{1+2\hat{\lambda}^2}{\hat{\sigma}\sqrt{1+\hat{\lambda}^2}}\sum_{i=1}^n \mathbf{x}_i\mathcal{E}_{10i}\left(\hat{\boldsymbol{\theta}}\right), \\
\ddot{Q}_{\sigma^2}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) &= \left.\frac{\partial^2 Q\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial\sigma^2\partial\sigma^2}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{n}{2\hat{\sigma}^4} - \frac{1+\hat{\lambda}^2}{\hat{\sigma}^6}\sum_{i=1}^n\left(\mathcal{E}_{02i}\left(\hat{\boldsymbol{\theta}}\right) - 2\mathcal{E}_{01i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i^\top\hat{\boldsymbol{\beta}} + \mathcal{E}_{00i}\left(\hat{\boldsymbol{\theta}}\right)\left(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right)^2\right) \\
&\quad + \frac{3\hat{\lambda}\sqrt{1+\hat{\lambda}^2}}{4\hat{\sigma}^5}\sum_{i=1}^n\left(\mathcal{E}_{11i}\left(\hat{\boldsymbol{\theta}}\right) - \mathcal{E}_{10i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right), \\
\ddot{Q}_{\sigma^2\lambda}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) &= \left.\frac{\partial^2 Q\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial\sigma^2\partial\lambda}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{\hat{\lambda}^2}{\hat{\sigma}^4}\sum_{i=1}^n\left(\mathcal{E}_{02i}\left(\hat{\boldsymbol{\theta}}\right) - 2\mathcal{E}_{01i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i^\top\hat{\boldsymbol{\beta}} + \mathcal{E}_{00i}\left(\hat{\boldsymbol{\theta}}\right)\left(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right)^2\right) \\
&\quad - \frac{1+2\hat{\lambda}^2}{2\hat{\sigma}^3\sqrt{1+\hat{\lambda}^2}}\sum_{i=1}^n\left(\mathcal{E}_{11i}\left(\hat{\boldsymbol{\theta}}\right) - \mathcal{E}_{10i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right), \\
\ddot{Q}_{\lambda}\left(\hat{\boldsymbol{\theta}} \mid \hat{\boldsymbol{\theta}}\right) &= \left.\frac{\partial^2 Q\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}\right)}{\partial\lambda\partial\lambda}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{n\left(\hat{\lambda}^2-1\right)}{\left(1+\hat{\lambda}^2\right)^2} - \frac{1}{\hat{\sigma}^2}\sum_{i=1}^n\left(\mathcal{E}_{02i}\left(\hat{\boldsymbol{\theta}}\right) - 2\mathcal{E}_{01i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i^\top\hat{\boldsymbol{\beta}} + \mathcal{E}_{00i}\left(\hat{\boldsymbol{\theta}}\right)\left(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right)^2\right) \\
&\quad + \frac{\hat{\lambda}\left(2\hat{\lambda}^2-3\right)}{\hat{\sigma}\left(1+\hat{\lambda}^2\right)^{\frac{3}{2}}}\sum_{i=1}^n\left(\mathcal{E}_{11i}\left(\hat{\boldsymbol{\theta}}\right) - \mathcal{E}_{10i}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right) - \sum_{i=1}^n \mathcal{E}_{20i}\left(\hat{\boldsymbol{\theta}}\right).
\end{aligned}
$$

References

Azzalini A (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

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

Barros M, Galea M, Leiva V, and Santos-Neto M (2018). Generalized Tobit models: Diagnostics and application in econometrics, *Journal of Applied Statistics*, **45**, 145–167.

Berkane M, Kano Y, and Bentler PM (1994). Pseudo maximum likelihood estimation in elliptical theory: Effects of misspecification, *Computational Statistics and Data Analysis*, **18**, 255–267.

Cameron AC and Trivedi PK (2005). *Microeconometrics: Methods and Applications*, Cambridge University Press, Cambridge.

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

Cook RD and Weisberg S (1982). *Residuals and Influence in Regression*, Chapman & Hall/CRC, Boca Raton, FL.

Kleiber C and Zeileis A (2008). *Applied Econometrics with R*, Springer-Verlag, New York, Available from: ISBN 978-0-387-77316-2

Lachos VH, Garay A, and Cabral CR (2020). Moments of truncated skew-normal/independent distributions, *Brazilian Journal of Probability and Statistics*, **34**, 478–494.

Lachos VH, Prates MO, and Dey DK (2021). Heckman selection-t model: Parameter estimation via the EM-algorithm, *Journal of Multivariate Analysis*, **184**, 104737.

Lachos VH, Bazan JL, Castro LM, and Park J (2022). The skew-t censored regression model: Parameter estimation via an EM-type algorithm, *Commnications for Statistical Applications and Methods*, **29**, 333–351.

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

Lee SY and Xu L (2004). Rinfluence analysis of nonlinear mixed-effects models, *Computational Statistics and 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 MB, Cabral CRB, Matos LA, and Lachos VH (2015). Influence diagnostics for student-t censored linear regression models, *Statistics*, **49**, 1074–1094.

Massuia MB, Garay AM, Lachos VH, and Cabral CRB (2017). Bayesian analysis of cen- sored linear regression models with scale mixtures of skew-normal distributions, *Statistics and Its Interface*, **10**, 425–439.

Matos LA, Prates MO, Chen MH, and Lachos VH (2013). Likelihood based inference for linear and nonlinear mixed-effects models with censored response using the multivariate-t distribution, *Statistica Sinica*, **23**, 1323–1345.

Matos LA, Bandyopadhyay D, Castro LM, and Lachos VH (2015). Influence assessment in censored mixed-effects models using the multivariate student-t distribution, *Journal of Multivariate Analysis*, **141**, 104–117.

Mattos TdB, Garay AM, and Lachos VH (2018). Likelihood-based inference for censored linear regression models with scale mixtures of skew-normal distributions, *Journal of Applied Statistics*, **45**, 2039–2066.

Mroz TA (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions, *Econometrica*, **55**, 765–799.

Nunez LM, Lachos VH, Galarza CE, and Matos LA (2021). Estimation and diagnostics for partially linear censored regression models based on heavy-tailed distributions, *Statistics and Its Interface*, **14**, 165–182.

Ortega EMM, Bolfarine H, and Paula GA (2003). Influence diagnostics in generalized log-gamma regression models, *Computational Statistics and Data Analysis*, **42**, 165–186.

Osorio F, Paula GA, and Galea M (2007). Assessment of local influence in elliptical linear models with longitudinal structure, *Computational Statistics and Data Analysis*, **51**, 4354–4368.

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

R Core Team (2022). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna.

RTI-FDA (2008). Snapshot of School Management Effectiveness: Peru Pilot Study (Technical report), USAID.

Therneau TM, Grambsch PM, and Fleming TR (1990). Martingale-based residuals for survival models, *Biometrika*, **77**, 147–160.

Valeriano KA, Galarza CE, Matos LA, and Lachos VH (2023). Likelihood-based inference for the multivariate skew-t regression with censored or missing responses, *Journal of Multivariate Analysis*, **196**, 105174.

Zeller CB, Cabral CRB, Lachos VH, and Benites L (2019). Finite mixture of regression models for censored data based on scale mixtures of normal distributions, *Advances in Data Analysis and Classification*, **13**, 89–116.

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

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

Received December 26, 2022; Revised September 30, 2023; Accepted October 3, 2023