UNICAMP

Universidade Estadual de Campinas

Instituto de Matemática, Estatística
e Computação Científica

Thalita do Bem Mattos

Robust estimation in regression models for censored data

Estimação robusta em modelos de regressão para dados censurados

CAMPINAS

2016

Thalita do Bem Mattos

Robust estimation in regression models for censored data

Estimação robusta em modelos de regressão para dados censurados

Dissertation presented to the Institute of Mathematics, Statistics and Scientific Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Statistics.

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestra em Estatística.

Orientador: Prof. Dr. Víctor Hugo Lachos Dávila

Coorientador: Prof. Dr. Aldo William Medina Garay

Este exemplar corresponde à versão final da dissertação defendida pela aluna Thalita do Bem Mattos, e orientada pelo Prof. Dr. Víctor Hugo Lachos Dávila.

Campinas

2016

Agência(s) de fomento e nº(s) de processo(s): CAPES

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Maria Fabiana Bezerra Muller - CRB 8/6162

M436r Mattos, Thalita do Bem, 1991-
Robust estimation in regression models for censored data / Thalita do Bem Mattos. – Campinas, SP : [s.n.], 2016.

Orientador: Víctor Hugo Lachos Dávila.
Coorientador: Aldo William Medina Garay.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Modelos lineares (Estatística). 2. Misturas de escalas. 3. Algoritmos de esperança-maximização. 4. Influência local. 5. Modelo de eliminação de casos. 6. Aproximação estocástica. I. Lachos Dávila, Víctor Hugo,1973-. II. Medina Garay, Aldo William. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Estimação robusta em modelos de regressão para dados censurados
Palavras-chave em inglês:
Linear models (Statistics)
Scale mixtures
Expectation-maximization algorithms
Local influence
Case-deletion model
Stochastic approximation
Área de concentração: Estatística
Titulação: Mestre em Estatística
Banca examinadora:
Víctor Hugo Lachos Dávila [Orientador]
Márcia D'Elia Branco
Filidor Edilfonso Vilca Labra
Data de defesa: 17-02-2016
Programa de Pós-Graduação: Estatística

Dissertação de Mestrado defendida em 17 de fevereiro de 2016 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). VÍCTOR HUGO LACHOS DÁVILA

Prof(a). Dr(a). MÁRCIA D'ELIA BRANCO

Prof(a). Dr(a). FILIDOR EDILFONSO VILCA LABRA

A Ata da defesa com as respectivas assinaturas dos membros
encontra-se no processo de vida acadêmica do aluno.

Aos meus pais Marilene e Milton,

à minha irmã, Areta,

e ao meu sobrinho, Luccas.

Dedico com muito amor e carinho.

Acknowledgement

Firstly I thank God for never abandon me, giving me courage and strength throughout this journey.

I want to express my sincere gratitude to my parents Marilene and Milton, for the love, dedication and for believing in me. For abdicate their own dreams in my favor. I am grateful to my sister Areta and my nephew Luccas for all the support and caring.

Also would like to express my deeply gratitude to my advisor Victor Hugo Lachos Dávila for his guidance, patience, support and trust.

My appreciation to my co-advisor Aldo Medina Garay for giving me important suggestions and make important corrections.

I thank all my friends and everyone else who helped contribute with this work.

Lastly, I am really grateful to CAPES for economic support during this master.

Resumo

Neste trabalho estudamos alguns aspectos de estimação e diagnóstico de influência global e local em modelos de regressão robustos com respostas censuradas sob a classe de distribuições de misturas de escala skew-normal (SMSN) (Lachos et al., 2010). A SMSN é uma classe atraente de distribuições assimétricas com caudas pesadas que inclui a skew-normal, skew-t, skew-slash, skew-normal contaminada e toda a família de distribuições de misturas de escala normal (SMN) como casos especiais. As estimativas de máxima verossimilhança (ML) dos parâmetros são obtidas utilizando uma aproximação estocástica do algoritmo EM (SAEM). Esta abordagem nos permite estimar os parâmetros de interesse de forma eficiente, assim como, obter os erros padrão, predições de valores não observados (censuras) e a função de log-verossimilhança. Para analisar o desempenho do modelo proposto, técnicas de deleção de casos e influência local são desenvolvidas para mostrar a robustez contra outlier e observações influentes. Os métodos propostos são verificados através da análise de vários estudos de simulação e aplicando em dois conjuntos de dados reais.

Palavras-chave: modelos de regressão censurados; caudas pesadas; algoritmo SAEM; distribuições de misturas de escala skew-normal; modelo de deleção de casos, influência local.

Abstract

In this work, we study studied some aspects of estimation and diagnostics of global and local influence in robust regression models with censored responses under the class of scale mixtures of the skew-normal (SMSN) distribution (Lachos et al., 2010). The SMSN is an attractive class of asymmetrical heavy-tailed densities that includes the skew-normal, skew-t, skew-slash, skew-contaminated normal and the entire family of scale mixtures of normal (SMN) distributions as special cases. The estimates of maximum likelihood (ML)of the parameters are obtained using a stochastic approximation of the EM (SAEM) algorithm. This approach allows us to estimate the parameters of interest efficiently, as well as obtain standard errors, predictions of unobservable values (censoring) and the log-likelihood function. To examine the performance of the proposed model, case-deletion and local influence techniques are developed to show its robust aspect against outlying and influential observations. The proposed methods are verified through the analysis of several simulation studies and applying in two real datasets.

Keywords: censored regression model; heavy tails; SAEM algorithm; scale mixtures of skew-normal distributions; case-deletion model; local influence.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Regression models with censored dependent variable (CR models) are applied in different fields, such as econometric analysis, astrophysics, clinical testing, among many others. In econometrics, for example, the study of the labor force participation of married women is usually conducted under the ordinary Tobit model (Greene, 2012). In AIDS research, the viral load measures can be subject to some upper and lower detection limits, below or above which they are not quantifiable. As a result, the viral load responses are either left or right censored depending on the diagnostic assays used (see, for instance, Wu, 2010).

The cases of censored responses with normal observational errors, denoted by N-CR, has been studied extensively in the literature, see for example, Nelson (1977), Thompson and Nelson (2003), Park et al. (2007) and Vaida and Liu (2009), to mention a few. However, several phenomena are not always in agreement with the assumptions of the normal model. To deal with this problem, some proposals have been made in the literature to replace the normality assumption with more flexible classes of distributions. For instance, Fernández and Steel (1999) discuss some inferential procedures in regression models with Student-t distribution for the errors. Ibacache-Pulgar and Paula (2011), propose local influence measures in the Student-t partially linear regression model. Recently, Arellano-Valle et al. (2012) and Massuia et al. (2014) proposed an extension of the CR model with normal errors (N-CR) to Student-t (T-CR) errors.

Other existing methods for robust estimation are based on the class of scale mixtures of normal (SMN) distributions introduced by Andrews and Mallows (1974). For example, Garay et al. (2015) proposed a robust CR model where the observational errors follow a SMN distribution (SMN-CR model). These distributions have heavier tails than the normal one, so they seem to be a reasonable choice for robust inference. They include as special cases many symmetric distributions, such as the normal (N), Pearson type VII (P-VII), Student-t (T), slash (SL) and contaminated normal (CN). Although these models are attractive, there is a need to check the distributional assumptions of the model errors because these can present skewness and heavy tail behavior, simultaneously. To overcome the problem of atypical data in an asymmetrical context, Branco and Dey (2001) proposed the class of scale mixtures of skew-normal (SMSN) distributions. This class of distributions contains the entire family of SMN distributions, and skewed versions of classic asymmetric distributions such as the skew-normal (SN), skew-t (ST), skew-slash (SSL) and skew contaminated normal (SCN) distributions. Recently, Massuia et al. (2015) developed a Bayesian framework for CR models by assuming that the random errors follow a SMSN distribution.

Motivated by this, in this work, we propose a robust parametric approach of the censored linear regression models based on the SMSN distributions, denoted by SMSN-CR, from a likelihood-based perspective, extending the previous cited works of Arellano-Valle et al. (2012), Massuia et al. (2014), Garay et al. (2015) and supplementing the work by Massuia et al. (2015). In addition, we suggest a fast estimation procedure to obtain the maximum likelihood (ML) estimates of the parameters, using a stochastic approximation of the EM (SAEM) algorithm, proposed by Delyon et al. (1999). The SAEM algorithm has been proved to be more computationally efficient than the classical Monte Carlo EM (MCEM) algorithm due to reuse of simulations from one iteration to the next in the smoothing phase of the algorithm (Jank, 2006). In this work we also develop influence diagnostic techniques, based on case deletion and local influence approaches, proposed by Cook and Weisberg (1982) and Cook (1986).

Based on what was discussed up to here, we will present a brief description of the family of SMSN distributions. Next, we present an outline of the EM and SAEM algorithms, a summary of analysis of influence and, finally, a organization of dissertation.

1.1 Scale mixtures of skew-normal (SMSN) distributions

In order to define the SMSN-CR model, we first make some remarks related to the SMSN class of distributions. This class of distributions was proposed by Branco and Dey (2001) and is a group of skew-thick-tailed distributions that are useful for robust inference and that contain as proper elements the SN, ST, SSL, SCN distributions and the entire family of SMN distributions proposed by Andrews and Mallows (1974) (see also, Lange and Sinsheimer, 1993). Thus, in the following we present some definitions where we explain first the fundamental concept of the SN distribution proposed by Azzalini (1985), and its relation with the SMSN class of distributions.

Definition 1. *A random variable Z has a skew-normal distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ, denoted by $Z \sim SN(\mu, \sigma^2, \lambda)$, if its probability density function (pdf) is given by*

$$f_{SN}(z|\mu, \sigma^2, \lambda) = 2\phi(z|\mu, \sigma^2)\Phi\left(\frac{\lambda(z-\mu)}{\sigma}\right), \qquad z \in \mathbb{R}, \tag{1.1.1}$$

where $\phi(\,\cdot\,|\mu, \sigma^2)$ denotes the density of the univariate normal distribution with mean μ and variance $\sigma^2 > 0$ and $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard univariate normal distribution.

Definition 2. *A random variable Y has a SMSN distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ, denoted by $SMSN(\mu, \sigma^2, \lambda; H)$, if it has the following stochastic representation:*

$$Y = \mu + \kappa(U)^{1/2}Z, \quad U \perp Z, \tag{1.1.2}$$

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

The random variable U is known as *the scale factor* and its cdf $H(\,\cdot\,|\boldsymbol{\nu})$ is called the *mixing distribution function.* Note that when $\lambda = 0$, the SMSN family reduces to the symmetric class of SMN distributions.

Using the representation given in Equation (1.1.2), we observe that

$$Y|U = u \sim SN(\mu, \kappa(u)\sigma^2, \lambda)$$

and integrating out U from the joint density of Y and U leads to the following marginal density of Y:

$$f_{SMSN}(y|\mu, \sigma^2, \lambda; H) = 2\int_0^\infty \phi(y|\mu, \kappa(u)\sigma^2)\Phi\left(\frac{\lambda(y-\mu)}{\sigma\kappa(u)^{1/2}}\right) dH(u). \tag{1.1.3}$$

Another important class of distribution, which will be useful for implementing the SAEM algorithm, is the truncated SMSN distributions, given by the following definition:

Definition 3. *Let $W \sim$ SMSN$(\mu, \sigma^2, \lambda; H)$ and $\mathbb{P}(a < W < b) > 0$, with $a < b$. A random variable Y has a truncated SMSN distribution in the interval $\lfloor a, b \rfloor$, denoted by $Y \sim$ TSMSN$(\mu, \sigma^2, \lambda; H, \lfloor a, b \rfloor)$, if it has the same distribution as $W|W \in \lfloor a, b \rfloor$. Here $\lfloor a, b \rfloor$ means that each extreme of the interval can be either open or closed.*

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

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

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

The following lemmas show a convenient stochastic representation of a SMSN random variable as well as its cdf. These lemmas will be useful to implement the proposed SAEM algorithm.

Lemma 1. *The random variable $Y \sim$ SMSN$(\mu, \sigma^2, \lambda; H)$, has a stochastic representation given by*

$$Y = \mu + \Delta T + \kappa(U)^{1/2}\tau^{1/2}T_1, \tag{1.1.4}$$

where $\Delta = \sigma\delta$, $\tau = (1-\delta^2)\sigma^2$, $\delta = \dfrac{\lambda}{\sqrt{1+\lambda^2}}$, $T = \kappa(U)^{1/2}|T_0|$, T_0 and T_1 are independent standard normal random variables and $|\cdot|$ denotes absolute value.

Proof. See Basso et al. (2010). □

The representation given in Lemma 1 is very appropriate to derive some mathematical properties and can be used to simulate pseudo-realizations of Y. It is important to stress that this representation was used by Basso et al. (2010) in the context of finite mixtures of SMSN distributions and by Cancho et al. (2011), Garay et al. (2011) and Labra et al. (2012) in the context of non-linear regression models for complete data. For instance, from Equation (1.1.2), we have the following hierarchical representation:

$$\begin{aligned} Y|T=t, U=u &\sim N(\mu+\Delta t\ ,\ \kappa(u)\tau), \\ T|U=u &\sim TN(0, \kappa(u)\ ;\ \lfloor 0, \infty \rfloor), \\ U &\sim H(\ \cdot\ ; \boldsymbol{\nu}), \end{aligned} \tag{1.1.5}$$

where $\mathrm{TN}(\mu, \sigma^2\ ;\ \lfloor a, b \rfloor)$ denotes a normal distribution with mean μ and variance σ^2 truncated in the interval $\lfloor a, b \rfloor$.

Lemma 2. *Let $Y \sim SMSN(\mu, \sigma^2, \lambda; H)$. Then, the cdf of Y can be written in the following way:*

$$F_{SMSN}(y \mid \mu, \sigma^2, \lambda\ ;\ H) = \int_0^{\infty} 2\Phi_2\Big(\boldsymbol{y}(u)^* | \boldsymbol{\mu}^*, \boldsymbol{\Sigma}\Big) dH(u), \tag{1.1.6}$$

where

$$\boldsymbol{y}(u)^* = (\kappa(u)^{-1/2} y, 0)^\top, \quad \boldsymbol{\mu}^* = (\mu, 0)^\top, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & -\delta\sigma \\ -\delta\sigma & 1 \end{pmatrix} \tag{1.1.7}$$

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

Proof. See Appendix 1 in Massuia et al. (2015). □

1.1.1 Particular cases of SMSN distributions.

Although we can deal with any $\kappa(\cdot)$ function, we restrict our attention to the case where $\kappa(u) = 1/u$, since it leads to good mathematical properties. Moreover, the scale factor U can be discrete or continuous and the form of the SMSN distribution is determined by its distribution. We take into account four members of SMSN class: skew–normal, skew–t, skew–slash and skew contaminated normal distributions. For each specific SMSN distribution described below, we compute its cdf, which is useful to evaluate the likelihood function related to CR models.

- *The skew–t distribution.* Denoted by $Y \sim ST(\mu, \sigma^2, \lambda; \nu)$, this case arises when we consider $U \sim Gamma(\nu/2, \nu/2)$ in Definition 2. Thus, the density of Y takes the form

$$f_{ST}(y|\mu, \sigma^2, \lambda; \nu) = \frac{2\,\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}\left(1+\frac{d(y)^2}{\nu}\right)^{-\frac{\nu+1}{2}} T_1\left(\lambda\, d(y)\sqrt{\frac{\nu+1}{\nu+d(y)^2}} \mid 0, 1, \nu+1\right), \quad y \in \mathbb{R}, \tag{1.1.8}$$

  where $d(y) = (y-\mu)/\sigma$. A particular case of the skew-t distribution is the skew–Cauchy distribution, when $\nu = 1$. Also, when $\nu \to \infty$, we get the skew-normal distribution as the limiting case. Using Lemma 2, we obtain the following expression for the cdf of Y:

$$F_{ST}(y \mid \mu, \sigma^2, \lambda\,; \nu) = 2\,T_2\left(\boldsymbol{y(u)}^* \mid \boldsymbol{\mu}^*, \boldsymbol{\Sigma}, \nu\right), \tag{1.1.9}$$

  where $\boldsymbol{y(u)}^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ are as in (1.1.7) and $T_m(\cdot \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \nu)$ represents the cdf of the m-variate Student-t distribution with location vector $\boldsymbol{\mu}_0$, scale matrix $\boldsymbol{\Sigma}_0$ and ν degrees of freedom. The proof of these results are given in Massuia et al. (2015).

- *The skew–slash distribution.* Denoted by $Y \sim SSL(\mu, \sigma^2, \lambda; \nu)$, in this case we consider $U \sim Beta(\nu, 1)$ with $\nu > 0$ in Definition 2. The density of Y is given by

$$f_{SSL}(y|\mu, \sigma^2, \lambda; \nu) = 2\nu \int_0^1 u^{\nu-1}\phi(y|\mu, u^{-1}\sigma^2)\Phi(u^{1/2}A(y))du, \quad y \in \mathbb{R}, \tag{1.1.10}$$

  where $A(y) = \lambda(y-\mu)/\sigma$. The cdf of the skew-slash distribution does not have a closed form expression. However, using Lemma 2, we can write it in terms of the following integral, which can be approximated by numerical methods:

$$F_{SSL}(y|\mu, \sigma^2, \lambda; \nu) = \int_0^\infty 2\nu\Phi_2\Big(\boldsymbol{y}(u)^*|\boldsymbol{\mu}^*, \boldsymbol{\Sigma}\Big)u^{\nu-1}\,du, \tag{1.1.11}$$

  where $\boldsymbol{y}(u)^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ are as in (1.1.7).

- *The skew-contaminated normal distribution.* Denoted by $Y \sim SCN(\mu, \sigma^2, \lambda; (\nu, \gamma))$, here U is a discrete random variable taking one of two states of $\boldsymbol{\nu} = (\nu, \gamma)^\top$. In this case the pdf of U is given by

$$U = \begin{cases} \gamma & \text{with probability } \nu; \\ 1 & \text{with probability } 1-\nu. \end{cases}$$

It follows immediately that

$$f_{SCN}(y|\mu,\sigma^2,\lambda;\boldsymbol{\nu}) = 2\{\nu\phi(y|\mu,\gamma^{-1}\sigma^2)\Phi(\gamma^{1/2}A(y)) + (1-\nu)\phi(y|\mu,\sigma^2)\Phi(A(y))\} \tag{1.1.12}$$

and

$$F_{SCN}(y|\mu,\sigma^2,\lambda;\boldsymbol{\nu}) = 2\{\nu\Phi_2\Big(\gamma^{1/2}\boldsymbol{y}^*|\boldsymbol{\mu}^*,\boldsymbol{\Sigma}\Big) + (1-\nu)\Phi_2\Big(\boldsymbol{y}^*|\boldsymbol{\mu}^*,\boldsymbol{\Sigma}\Big)\}, \tag{1.1.13}$$

where $A(y) = \lambda(y-\mu)/\sigma$.

- *The skew-normal distribution.* This distribution is obtained when $U = 1$ (a degenerated random variable) in Definition 2. The density of Y was defined in (1.1.1) and clearly, from Lemma 2, the cdf of Y is given by

$$F(y) = 2\Phi_2\Big(\boldsymbol{y}^*|\boldsymbol{\mu}^*,\boldsymbol{\Sigma}\Big), \tag{1.1.14}$$

where $\boldsymbol{y}^* = (y,0)^\top$ and $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}$ are as in (1.1.7).

In Table 1.1, we present the expected values $k_m = E[U^{-m/2}]$ for all the SMSN distributions discussed above, which are useful to define the SMSN-CR model.

Table 1.1: $k_m = E[U^{-m/2}]$ for different SMSN models.

Model	k_m
SN	1
ST	$\left(\frac{\nu}{2}\right)^{m/2}\frac{\Gamma\left(\frac{\nu-m}{2}\right)}{\Gamma(\nu/2)}$
SSL	$\frac{2\nu}{\nu-m/2}$
SCN	$\frac{\nu}{\gamma^{m/2}}+1-\nu$

1.2 Algorithms for ML estimation

In models with non-observed or incomplete data, the EM algorithm is a very popular iterative optimization strategy commonly used. This algorithm has many attractive features such as numerical stability and simplicity of implementation, and its memory requirements are quite reasonable (Couvreur, 1996). Letting $\mathbf{y}_{comp} = (\mathbf{y}^m, \mathbf{y}^o)$ the complete data vector, where $\mathbf{y}^m$ represents the missing data and $\mathbf{y}^o$ the observed data respectively and $\ell_{comp}(\boldsymbol{\theta}|\mathbf{y}_{comp})$ the complete data log-likelihood function, then the EM-algorithm proceeds in two steps:

- **E-step:** Let $\widehat{\boldsymbol{\theta}}^{(j)}$ be the current $j-$th step estimate of $\boldsymbol{\theta}$. By using the property of conditional expectation, we can compute the $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)})$ function by

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)}) = E\left[\ell_{comp}(\boldsymbol{\theta}|\mathbf{y}_{comp})|\mathbf{y}^o, \widehat{\boldsymbol{\theta}}^{(j)}\right]. \tag{1.2.1}$$

- **M-step:** Maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)})$ with respect to $\boldsymbol{\theta}$, obtaining $\widehat{\boldsymbol{\theta}}^{(j+1)}$.

As mentioned by Meza et al. (2012), each iteration of the EM algorithm increases the likelihood function $\ell(\boldsymbol{\theta}|\mathbf{y}^o)$ and the EM sequence $\boldsymbol{\theta}^{(j)}$ converges to a stationary point of the observed likelihood under mild regularity conditions (for more details see Wu (1983) and Vaida (2005)).

For cases in which the E-step has no analytic form, Wei and Tanner (1990) proposed the Monte Carlo EM (MCEM) algorithm, in which the E-step is replaced by a Monte Carlo approximation based on a large number of independent simulations of the missing data. In order to reduce the number of required simulations compared to the MCEM algorithm, Delyon et al. (1999) proposed the stochastic approximation version of the EM algorithm, the so-called SAEM algorithm, which consists of replacing the E-step by a stochastic approximation, obtained using simulated data, while the M-step is unchanged. The SAEM algorithm consists, at each iteration, of successively simulating the missing data with the conditional distribution, and updating the unknown parameters of the model. Thus, the $j-$th iteration of SAEM algorithm consists of the following steps:

- **S-step:** Draw the missing data $\mathbf{y}^{m(j)}$ with the conditional distribution $p(\mathbf{y}^m|\mathbf{y}^o, \widehat{\boldsymbol{\theta}}^{(j-1)})$.

- **AE-step:** Update $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)})$ according to

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)}) \approx Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j-1)}) + \gamma_j \left[\frac{1}{m}\sum_{\ell=1}^{m} \ell_{comp}(\boldsymbol{\theta}|\mathbf{y}^o, \mathbf{y}^{m(j)}) - Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j-1)})\right]. \quad (1.2.2)$$

- **M-step:** Maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(j)})$ with respect to $\boldsymbol{\theta}$ obtaining $\widehat{\boldsymbol{\theta}}^{(j+1)}$,

where γ_j is a decreasing sequence of positive numbers such that

$$\sum_{j=1}^{\infty} \gamma_j = \infty \quad \text{and} \quad \sum_{j=1}^{\infty} \gamma_j^2 < \infty, \quad (1.2.3)$$

as presented by Kuhn and Lavielle (2004).

Thus, the SAEM algorithm performs a Monte Carlo E-step, like MCEM, but with a small and fixed Monte Carlo sample sizes ($m \leq 20$). This is possible because unlike the traditional EM algorithm and its variants, the SAEM algorithm uses not only the current simulation of the missing data at the $j-$iteration, denoted by $\mathbf{y}^m$, but also some or all previous simulations, where this *'memory'* property is set by the smoothing parameter γ_j.

Note, in Equation (1.2.2), that sequence γ_j has a strong impact on the speed of convergence of the algorithm. Thus, if the smoothing parameter γ_j is equal to 1 for all j, the SAEM algorithm will have *'no memory'*, and will be equivalent to the MCEM algorithm. The SAEM with no memory will converge quickly (convergence in distribution) to a solution neighborhood, but the algorithm with memory will converge slowly (almost sure convergence) to the ML solution. As suggested by Galarza et al. (2015), we use the following choice of the smoothing parameter:

$$\gamma_j = \begin{cases} 1, & \text{for} \quad 1 \leq j \leq cS, \\ \frac{1}{j-cS}, & \text{for} \quad cS+1 \leq j \leq S, \end{cases}$$

where S is the maximum number of iterations, and c a cutoff point ($0 \leq c \leq 1$) that determines the percentage of initial iterations with no memory. For example, if $c = 0$, the algorithm will have memory for all iterations, and hence will converge slowly to the ML estimates. If $c = 1$, the algorithm will have no memory, and so will converge quickly to a solution neighborhood. For the first case, S would need to be large in order to achieve the ML estimates. For the second, the algorithm will yield a Markov Chain where, after applying a *burn-in* and *thinning*, the mean of the chain observations can be a reasonable estimate. A number c between 0 and 1 ($0 < c < 1$) will assure an initial convergence in distribution to a solution neighborhood for the first cS iterations and an almost sure convergence for the rest of the iterations. Hence, this combination will lead to a fast algorithm with good estimates. To implement SAEM, the user must fix several constants matching the number of total iterations S and the cutoff point c that defines the start of the smoothing step of the SAEM algorithm. However, those parameters will vary depending of the model and the data. To determinate those constants, a graphical approach is recommended to monitor the convergence of the estimates for all the parameters, and if possible, to monitor the difference (relative difference) between two successive evaluations of the log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{y}_o)$, given by:

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

respectively.

1.3 Diagnostic Analysis

The statistical models are important tools to extract and understand essential features of a dataset. An important stage in the analysis is the verification of possible deviations from the assumptions made in the model, such as the existence of extreme observations with some interference in the results of the fit. The elements of the dataset that effectively control aspects of the analysis, are said influential if they produce changes in the analysis result when deleted or subjected to some type of disturbance.

In the statistical literature have developed two main approaches for the detection of influential observations. The first approach is the case-deletion model in which the impact of dropping an individual observation in the prediction is directly assessed by measures such as the generalized Cook's distance and likelihood distance (see, Cook (1977) and Cook and Weisberg (1982)). The second approach is local influence technique Cook (1986), used to assess the stability of the estimation outputs with respect to the model inputs.

Following the pioneering work of Cook (1986), this area of research has received considerable attention in the recent statistical literature; see, for example Lesaffre and Verbeke (1998), Zhu and Lee (2001), Lee and Xu (2004) and Osorio et al. (2007), amongst others. A study of influence diagnostic in linear regression models under SMSN distributions has been presented by Zeller et al. (2011). However, to the best of our knowledge, there are no previous studies of diagnostic analysis for censored linear regression models based in the SMSN distributions.

1.4 Organization of the Dissertation

The results contained in this dissertation are organized into four chapters. In Chapter 2, we describe the SMSN-CR model and the ML estimation procedure based in the SAEM algorithm and we discuss how to obtain the approximated standard errors. To conclude this chapter we examine the performance of the proposed methods through various simulation studies as well as the analysis of a real dataset.

In Chapter 3, we develop influence diagnostic techniques, based on case-deletion and local influence approaches. The methodology is illustrated using a motivating astrophysical dataset and we present a simulation study evaluating the efficiency of our method in detecting outliers under various degrees of data perturbation and censoring.

Finally, the Chapter 4 presents some concluding remarks, with some possible directions for future research.

Chapter 2

The SMSN censored linear regression model

2.1 Introduction

Censored regression models, in general, are based on the development of the so called Tobit model, which is constructed in terms of the normal assumption (Tobin, 1958). However, many models do not fit the assumption of normality. Thus, in recent years several authors have studied CR models for statistical modeling of censored datasets involving observed variables with heavier tails than the normal distribution. For instance, Arellano-Valle et al. (2012) and Massuia et al. (2014) proposed an extension of the CR model with normal errors (N-CR) to Student-t (T-CR) errors. Garay et al. (2015) proposed a robust CR model where the observational errors follow a SMN distribution (SMN-CR model). More recently, Massuia et al. (2015) developed a Bayesian framework for CR models by assuming that the random errors follow a SMSN distribution. In this chapter, we suggest an attractive ML estimation procedure for CR models considering the SMSN class of distributions, extending the works by Arellano-Valle et al. (2012), Massuia et al. (2014), Garay et al. (2015) and supplementing the work by Massuia et al. (2015) from a likelihood-based perspective.

A typical algorithm for ML estimation in models involving the class of SMSN distributions is the EM algorithm and its variants. See, for instance, Basso et al. (2010), Lachos et al. (2010) and Garay et al. (2011). However, in some cases EM-type algorithms are not appropriate due to the computational difficulty in the E-step, which involves the computation of expected quantities that cannot be obtained analytically and must be calculated using stochastic simulation. To deal with this problem, Delyon et al. (1999) proposed a stochastic approximation version of the EM algorithm, the so-called SAEM algorithm. This algorithm consists of replacing the E-step by a stochastic approximation obtained using simulated data, while the M-step remains unchanged. Jank (2006) showed that the computational effort of SAEM is much smaller and reaches convergence in just a fraction of the simulation size when compared to Monte Carlo EM (MCEM). This is due the memory effect contained in the SAEM method, in which the previous simulations are considered in the computation of the posterior ones. In this chapter, we develop a full likelihood approach for SMSN-CR models, including the implementation of the SAEM algorithm for ML estimation with the likelihood function, predictions of unobservable values of the response and the asymptotic standard errors as a byproduct.

2.2 Model specification

The SMSN-CR model that we are going to discuss is defined by:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{2.2.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of regression parameters, Y_i is a response variable and $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is a vector of explanatory variables for subject i.

In this work, we assume that

$$\varepsilon_i \sim SMSN\left(-\sqrt{\frac{2}{\pi}}k_1\Delta, \sigma^2, \lambda; H\right), \quad i = 1, \ldots, n, \tag{2.2.2}$$

are independent random variables. The value of the location parameter, $-\sqrt{\frac{2}{\pi}}k_1\Delta$, of ε_i is chosen in order to obtain $E[\varepsilon_i] = 0$, as in the normal model. For more details, see Lemma 1 in Basso et al. (2010). Thus, when the moments exist, we have

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

where $E[Y_i] = \mathbf{x}_i^\top \boldsymbol{\beta}$ and $Var[Y_i] = k_2\sigma^2 - \dfrac{2k_1^2\Delta^2}{\pi}$, for $i = 1, \ldots, n$. The values of k_1 and

k_2 are given in Table 1.1, for particular cases of SMSN distributions.

In this work we are interested in the situation in which the response variable is not fully observed for all subjects. Thus, for the $i-$th subject and assuming left-censoring, Y_i is a latent variable and the observed data (V_i, ρ_i) is of the form

$$V_i = \begin{cases} c_i & \text{if} \quad \rho_i = 1 \quad (\text{i.e. } Y_i \le c_i); \\ Y_i & \text{if} \quad \rho_i = 0 \quad (\text{i.e. } Y_i > c_i), \end{cases} \tag{2.2.3}$$

for some known threshold point c_i, $i = 1, 2, \ldots, n$.

The SMSN-CR model is defined by combining (2.2.1)–(2.2.3). The log-likelihood function of $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma^2, \lambda, \nu\right)^\top$ given the observed data $(\boldsymbol{v}, \boldsymbol{\rho})$, is

$$\ell(\boldsymbol{\theta}|\boldsymbol{v}, \boldsymbol{\rho}) = \log\left\{\prod_{i=1}^{n}\left[F_{SMSN}\left(\frac{v_i - \mathbf{x}_i^\top\boldsymbol{\beta}}{\sigma}|\boldsymbol{\theta}; H\right)\right]^{\rho_i}\left[f_{SMSN}(v_i|\boldsymbol{\theta}; H)\right]^{1-\rho_i}\right\}, \tag{2.2.4}$$

where $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_n)$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ is the observed sample of $\mathbf{V} = (V_1, V_2, \ldots, V_n)$. Thus, $\rho_i = 1,\ 0$ indicating whether the $i-$th observation is censored, i.e. $Y_i \le c_i$, or not respectively. $f_{SMSN}(\cdot|\boldsymbol{\theta}; H)$ and $F_{SMSN}(\cdot|\boldsymbol{\theta}; H)$ represent the pdf and cdf of the SMSN class, respectively.

For simplicity, we will assume the data are left censored, and develop the SAEM algorithm for ML estimation. Extensions, to right censored data are immediate.

2.2.1 ML estimation via the SAEM algorithm

In this subsection we consider the ML estimation of the parameters in the SMSN-CR models introduced in Section 1.2. In particular, we show how to implement the SAEM algorithm for the particular cases of the SMSN class, that is, the SN, ST, SSL and SCN distributions.

Let $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^\top, \Delta, \tau, \nu)^\top$ be the vector of parameters in focus, which has a one-to-one correspondence with the original vector of parameters $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma^2, \lambda, \nu\right)^\top$, since

$$\Delta = \sigma\frac{\lambda}{\sqrt{\lambda^2 + 1}} = \sigma\delta \quad \text{and} \quad \tau = \left(1 - \delta^2\right)\sigma^2 = \frac{\sigma^2}{\lambda^2 + 1},$$

we can obtain σ^2 and λ from Δ and τ considering

$$\sigma^2 = \tau + \Delta^2 \quad \text{and} \quad \lambda = \Delta/\sqrt{\tau}. \tag{2.2.5}$$

We observe that a useful straightforward result, used by Basso et al. (2010) and Massuia et al. (2015), is that the conditional distribution of T_i given y_i and u_i is $TN(\mu_{Ti} - \sqrt{\frac{2}{\pi}}k_1, u_i^{-1}M_T^2; \lfloor -\sqrt{\frac{2}{\pi}}k_1, \infty \rfloor)$, where

$$\mu_{T_i} = \frac{\Delta}{\Delta^2 + \tau}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) \quad \text{and} \quad M_T^2 = \frac{\tau}{\Delta^2 + \tau}. \tag{2.2.6}$$

In order to implement the SAEM algorithm, we consider a data augmentation scheme that consists of assuming that the latent variables (*missing data*) in the model, given by the vector of censored responses $\mathbf{Y} = (y_1, y_2, \ldots, y_n)^\top$, the vector $\mathbf{t} = (t_1, t_2, \ldots, t_n)^\top$ and $\mathbf{u} = (u_1, u_2, \ldots, u_n)^\top$ - see representation (1.1.4) - can be observed. Thus, considering the observed data $(\mathbf{V}, \boldsymbol{\rho})$ and the latent variables $(\mathbf{Y}, \mathbf{t}, \mathbf{u})$, we define the complete data by $\mathbf{Y}_{comp} = (\mathbf{V}^\top, \boldsymbol{\rho}^\top, \mathbf{Y}^\top, \mathbf{t}^\top, \mathbf{u}^\top)^\top$. Then, it is easy to derive the complete data log-likelihood, defined by $\ell_{comp}(\boldsymbol{\theta}^*|\mathbf{Y}_{comp})$, using the representation (1.1.5) as:

$$\ell_{comp}(\boldsymbol{\theta}^*|\mathbf{Y}_{comp}) \propto cte - \frac{n}{2}\log\tau - \frac{1}{2\tau}\sum_{i=1}^{n} u_i(y_i - \mathbf{x}_i^\top\boldsymbol{\beta} - \Delta t_i)^2 + \sum_{i=1}^{n}\log h(u_i|\boldsymbol{\nu}), \tag{2.2.7}$$

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

- **E-step:** Given the current estimate $\boldsymbol{\theta}^{*(j)} = (\beta^{(j)\top}, \Delta^{(j)}, \tau^{(j)}, \nu^{(j)})^\top$ at the $j-$th iteration, we obtain the conditional expectation of the complete data log-likelihood function (Q-function), which is given by

$$\begin{aligned} Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{*(j)}) &= E\left[\ell_{comp}(\boldsymbol{\theta}^*|\mathbf{Y}_{comp})|\mathbf{V}, \boldsymbol{\rho}, \boldsymbol{\theta}^{*(j)}\right] \\ &= cte - \frac{n}{2}\log(\tau) - \frac{1}{2\tau}\sum_{i=1}^{n}\left[\mathcal{E}_{02i}(\boldsymbol{\theta}^{*(j)}) - 2\mathcal{E}_{01i}(\boldsymbol{\theta}^{*(j)})\mathbf{x}_i^\top\boldsymbol{\beta}\right. \\ &+ \mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)})(\mathbf{x}_i^\top\boldsymbol{\beta})^2 - 2\Delta\mathcal{E}_{11i}(\boldsymbol{\theta}^{*(j)}) + 2\Delta\mathcal{E}_{10i}(\boldsymbol{\theta}^{*(j)})\mathbf{x}_i^\top\boldsymbol{\beta} \\ &+ \left.\Delta^2\mathcal{E}_{20i}(\boldsymbol{\theta}^{*(j)})\right] + \mathrm{E}\left[\log\{h(U_i|\boldsymbol{\nu})\}|V_i, \rho_i, \boldsymbol{\theta}^{*(j)}\right]. \end{aligned} \tag{2.2.8}$$

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

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

as well as

$$\mathrm{E}[\log\{h(U_i|\boldsymbol{\nu})\}|V_i, \rho_i].$$

As presented by Basso et al. (2010), considering known properties of conditional expectation and Equation (2.2.6), we obtain

$$\mathcal{E}_{10i}(\boldsymbol{\theta}^{*(j)}) = \mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)})\mu_{T_i}^{(j)} + M_T^{(j)}\psi_i^{(j)}, \tag{2.2.9}$$

$$\mathcal{E}_{20i}(\boldsymbol{\theta}^{*(j)}) = \mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)})\mu_{T_i}^{2(j)} + M_T^{2(j)} + M_T^{(j)}\mu_{T_i}^{(j)}\psi_i^{(j)}, \tag{2.2.10}$$

where

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

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

a) *For an uncensored observation i:*

In this case we have that $\rho_i = 0$, thus

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

and, therefore,

$$\mathcal{E}_{rsi}(\boldsymbol{\theta}^{*(j)}) = y_i^s\mathcal{E}_{r0i}(\boldsymbol{\theta}^{*(j)}), \tag{2.2.11}$$

where $\mathcal{E}_{r0i}(\boldsymbol{\theta}^{*(j)})$ can be obtained using equations (2.2.9)-(2.2.10) and the results given by Basso et al. (2010). Thus, for example,

∗ **For the skew-t case**

$$\mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)}) = \frac{2^2\nu^{\nu^{(j)}/2}\Gamma\left(\frac{\nu^{(j)}+3}{2}\right)(\nu^{(j)}+\mathrm{d}^{(j)}(y_i))^{-\frac{\nu^{(j)}+3}{2}}}{f_{ST}(y_i)\Gamma\left(\frac{\nu^{(j)}}{2}\right)\sqrt{\pi}(\tau^{(j)}+\Delta^{2(j)})^{1/2}} \times T\left(\sqrt{\frac{\nu^{(j)}+3}{\nu^{(j)}+\mathrm{d}^{(j)}(y_i)}}A_i^{*(j)};\nu^{(j)}+3\right),$$

$$\psi_i^{(j)} = \frac{2\Gamma(\frac{\nu^{(j)}+2}{2})\nu^{\nu^{(j)}/2}(\nu^{(j)}+\mathrm{d}^{(j)}(y_i)+A_i^{2*(j)})^{-\frac{\nu^{(j)}+2}{2}}}{f_{ST}(y_i)\Gamma\left(\frac{\nu^{(j)}}{2}\right)\pi(\tau^{(j)}+\Delta^{2(j)})^{1/2}},$$

as defined in (1.1.8), $f_{ST}(\cdot)$ represents the pdf of skew-t distribution and $T(\cdot;\nu)$ is the cdf of the standard Student-t distribution.

∗ **For the skew-slash case**

$$\mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)}) = \frac{\nu^{(j)}2^{\nu^{(j)}+2}\Gamma\left(\frac{2\nu^{(j)}+3}{2}\right)P_1\left(\frac{2\nu^{(j)}+3}{2},\frac{\mathrm{d}^{(j)}(y_i)}{2}\right)\mathrm{d}^{(j)}(y_i)^{-\frac{2\nu^{(j)}+3}{2}}}{f_{SSL}(y_i)\sqrt{\pi}(\tau^{(j)}+\Delta^{2(j)})^{1/2}} \times E\left[\Phi(S_i^{(j)1/2}A_i^{*(j)})\right],$$

$$\psi_i^{(j)} = \frac{\nu^{(j)}2^{\nu^{(j)}+1}\Gamma\left(\frac{2\nu+2}{2}\right)}{f_{SSL}(y_i)\pi(\tau^{(j)}+\Delta^{2(j)})^{1/2}}\left(\mathrm{d}^{(j)}(y_i)+A_i^{2*(j)}\right)^{-\frac{2\nu^{(j)}+2}{2}} \times P_1\left(\frac{2\nu^{(j)}+2}{2},\frac{\mathrm{d}^{(j)}(y_i)+A_i^{2*(j)}}{2}\right),$$

where $S_i^{(j)} \sim Gamma\left(\frac{2\nu^{(j)}+3}{2},\frac{\mathrm{d}^{(j)}(y_i)}{2}\right)\mathbb{I}_{(0,1)}$ is a truncated gamma distribution on $(0,1)$, with the parameter values in parentheses before truncation and $P_x(a,b)$ denotes the cdf of the $Gamma(a,b)$ evaluated at x . As defined in (1.1.10), $f_{SSL}(\cdot)$ represents a density of skew-slash distribution.

∗ **For the skew contaminated normal case**

$$\mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)}) = \frac{2}{f_{SCN}(y_i)}\left\{\nu^{(j)}\gamma^{(j)}\phi\left(y_i;\mu^{*(j)},\gamma^{-1(j)}(\tau^{(j)}+\Delta^{2(j)})\right)\Phi\left(\gamma^{1/2}A_i^{*(j)}\right) +(1-\nu^{(j)})\phi\left(y_i;\mu^{*(j)},\tau^{(j)}+\Delta^{2(j)}\right)\Phi\left(A_i^{*(j)}\right)\right\}$$

$$\psi_i^{(j)} = \frac{2}{f_{SCN}(y_i)}\left\{\nu^{(j)}\gamma^{(j)}\phi\left(y_i;\mu^{*(j)},\gamma^{-1(j)}(\tau^{(j)}+\Delta^{2(j)})\right)\Phi\left(\gamma^{1/2}A_i^{*(j)}\right) +(1-\nu^{(j)})\phi\left(y_i;\mu^{*(j)},\tau^{(j)}+\Delta^{2(j)}\right)\phi\left(A_i^{*(j)}\right)\right\},$$

where $f_{SCN}(\cdot)$ represents the pdf of the skew contaminated normal distribution, as defined in (1.1.12).

In all cases described before, $\mu^{*(j)} = \mathbf{x}_i^\top \boldsymbol{\beta}^{(j)} - \sqrt{\frac{2}{\pi}} k_1 \Delta^{(j)}, \quad A^{*(j)} = \frac{\mu_{T_i}^{(j)}}{M_T^{(j)}}$ and $\mathrm{d}^{(j)}(y_i) = \frac{(y_i - \mu^{*(j)})}{\sqrt{\tau^{(j)} + \Delta^{2(j)}}}$ represents the Mahalanobis distance. Thus, in each step, the conditional expectations $\mathcal{E}_{00i}(\boldsymbol{\theta}^{(j)})$ and $\psi_i^{(j)}$ can be easily obtained.

For the skew–t and skew contaminated normal distributions we have computationally attractive expressions that can be easily implemented. However, this is not the case for the skew–slash one, where Monte Carlo integration can be employed, as suggested by Basso et al. (2010) and Lachos et al. (2010).

b) *For a censored observation* i:

In this case, we have that $\rho_i = 1$, i.e. $Y_i \leq c_i$, therefore

$$\mathcal{E}_{rsi}\left(\boldsymbol{\theta}^{*(j)}\right) = \mathrm{E}[U_i T_i^r Y_i^s | V_i, Y_i \leq c_i, \boldsymbol{\theta}^{*(j)}] \text{ with } r, s = 0, 1, 2. \tag{2.2.12}$$

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

* **S-step (Sampling)**
  Let $\mathbf{Y}^{(c)} = \left(Y_1^{(c)}, Y_2^{(c)}, \ldots, Y_{n^c}^{(c)}\right)$ the vector of n^c censored cases, where $Y_i^{(c)}$ is generated from TSMSN $\left(\mathbf{x}_i^\top \boldsymbol{\beta} - \sqrt{\frac{2}{\pi}} k_1 \Delta, \tau + \Delta^2, \lambda; H, \lfloor -\infty, c_i \rfloor\right)$ for $i = 1, \ldots, n^c$. Thus, the new vector of observations
  $\mathbf{Y}^{(l,j)} = (Y_{i1}^{(l,j)}, \ldots, Y_{in^c}^{(l,j)}, Y_{n_i^c+1}, \ldots, Y_n)$ is a sample generated for the n^c censored cases and the observed values (uncensored cases), for $l = 1, \ldots, m$. Subsection 2.2.2 describes the details of the methods used to generate from the random variable $\mathbf{Y}^{(c)}$.

* **AE-step (Approximation Expectation)**
  Since we have the sequence $\mathbf{Y}^{(l,j)}$, at the j-th iteration, considering equations (2.2.9)-(2.2.10) and the results given in Basso et al. (2010), we replace the conditional expectations in (2.2.11) by the following stochastic approximations:

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

for $r, s = 0, 1, 2$.

An advantage of the SAEM algorithm is that even though it performs a Monte Carlo E-step, it requires a small and fixed Monte Carlo sample size, making it much faster than MCEM. Some authors claim that $m = 10$ is large enough, but to be more conservative, we chose $m = 20$.

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

$$\boldsymbol{\beta}^{(j+1)} = \left(\sum_{i=1}^{n} \mathcal{E}_{00i}(\boldsymbol{\theta}^{*(j)})(\mathbf{x}_i\mathbf{x}_i^\top)\right)^{-1} \left[\sum_{i=1}^{n} \mathbf{x}_i \mathcal{E}_{01i}(\boldsymbol{\theta}^{*(j)}) - \Delta \sum_{i=1}^{n} \mathbf{x}_i \mathcal{E}_{10i}(\boldsymbol{\theta}^{*(j)})\right];$$

$$\Delta^{(j+1)} = \frac{\sum_{i=1}^{n} \mathcal{E}_{11i}(\boldsymbol{\theta}^{*(j)}) - \sum_{i=1}^{n} \mathcal{E}_{10i}(\boldsymbol{\theta}^{*(j)})(\mathbf{x}_i^\top \boldsymbol{\beta}^{(j+1)})}{\sum_{i=1}^{n} \mathcal{E}_{20i}(\boldsymbol{\theta}^{*(j)})};$$

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

- **CML-step:** We estimates $\boldsymbol{\nu}$ by maximizing the actual marginal log-likelihood function, obtaining

$$\boldsymbol{\nu}^{(j+1)} = \text{argmax}_{\boldsymbol{\nu}} \left\{\sum_{i=1}^{n} \log\left[F_{SMSN}\left(v_i|\boldsymbol{\theta}; H\right)\right]^{\rho_i} + \sum_{i=1}^{n} \log\left[f_{SMSN}(v_i|\boldsymbol{\theta}; H)\right]^{1-\rho_i}\right\}.$$

Note that $\sigma^{2(j+1)}$ and $\lambda^{(j+1)}$ can be recovered using (2.2.5). The more efficient CML-step can be easily accomplished by using, for instance, the *optim* routine in the *R* software (R Development Core Team, 2015).

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

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

is small enough. We have adopted this strategy to update the estimate of $\boldsymbol{\nu}$, by direct maximization of the marginal log-likelihood, circumventing the computation of

$\mathrm{E}_{\boldsymbol{\theta}^{(j)}}[\log\{h(U_i|\boldsymbol{\nu})\}|\mathbf{V},\boldsymbol{\rho}]$.

In order to make our proposed algorithm more informative for the reader, in Figure 2.1 we present a flow diagram, which reports all the steps needed to implement the SAEM algorithm.

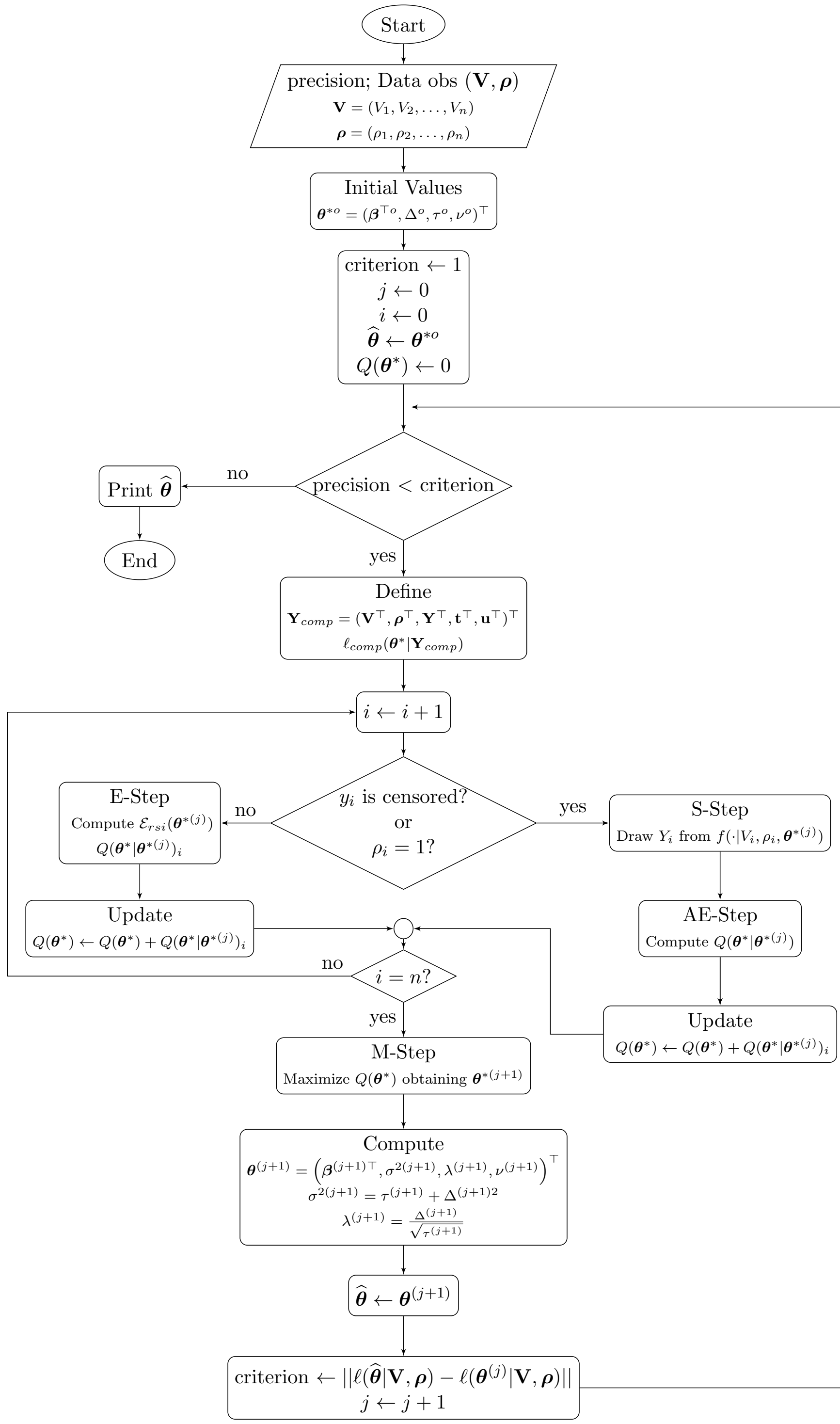

Figure 2.1: Flow diagram of the SAEM algorithm.

2.2.2 Computational aspects

The convergence of the SAEM algorithm is ensured by a careful choice of the simulation data method. Thus, in this subsection, we describe two simulation methods to generate random samples from the random variable $Y \sim \text{TSMSN}(\mu, \sigma^2, \lambda; H, \lfloor a, b \rfloor)$. We concentrate on the truncated skew normal (TSN), truncated skew-t (TST), truncated skew slash (TSSL) and truncated skew contaminated normal (TSCN) distributions.

We use the sampling/importance resampling method (Method 1), proposed by Rubin (1987) and Rubin et al. (1988), to generate samples from the TSN and TST models. For the TSSL and TSCN models we use the stochastic representation of a SMSN random variable, given in Lemma 1 (Method 2). In the following, we present a brief description of those two methods:

- **Method 1**

  The sampling/importance resampling (SIR) method is useful to generate an approximate independent and identically distributed (*iid*) sample of size m, from the target density $f(y)$, where $y \in \mathcal{S}_Y \subseteq \mathbb{R}$. Thus, let $g(y)$ a proposal density with the same support $\mathcal{S}_Y$. The method consists of two steps:

  – **Step 1.**(*Sampling*) Generate a random sample $Y_1,\ Y_2, \cdots, Y_J$ from $g(y)$ and construct weights
  $$W(Y_j) = \frac{f(Y_j)}{g(Y_j)}, \ \ j = 1, \ldots, J$$
  and probabilities
  $$\pi_j = \frac{W(Y_j)}{\sum_{j=1}^{J} W(Y_j)}, \ \ j = 1, \ldots, J.$$

  – **Step 2.**(*Importance resampling*) Draw m values ($m << J$) $Y_1^*, ..., Y_m^*$ from the J values $Y_1, Y_2, \ldots, Y_J$ with respective probabilities $\pi_1,\ \pi_2, \ldots \pi_J$. In practice, Rubin (1987) suggested $J/m = 20$.

  For the TSN model, the target density $f(\cdot)$ is a truncated skew normal distribution $\text{TSN}\left(\mu, \sigma^2, \lambda; \lfloor a, b \rfloor\right)$ and as proposal density $g(\cdot)$, we utilize truncated normal distribution, $\text{TN}\left(\mu, \sigma^2; \lfloor a, b \rfloor\right)$. For TST model, the target density $f(\cdot)$ is a truncated skew-t distribution $\text{TST}\left(\mu, \sigma^2, \lambda, \nu; \lfloor a, b \rfloor\right)$ and as the proposal density $g(\cdot)$, we utilize the truncated t distribution, $\text{Tt}\left(\mu, \sigma^2, \nu; \lfloor a, b \rfloor\right)$.

- **Method 2**

  In this case, we need to generate samples from the random variable

$$Y \sim \text{TSMSN}(\mu, \sigma^2, \lambda; H, \lfloor a, b \rfloor).$$

  Then, since U is a positive random variable, we have that

$$a \; < \; Y \; < \; b,$$

  which implies

$$(a-\mu)U^{1/2} \; < \; (Y-\mu)U^{1/2} \; < \; (b-\mu)U^{1/2}.$$

  Considering the stochastic representation given in (1.1.2), we have that $Z = (Y - \mu)U^{1/2}$, where $Z \sim \text{SN}(0, \sigma^2, \lambda)$. Thus,

$$(a-\mu)U^{1/2} \; < \; Z \; < \; (b-\mu)U^{1/2}.$$

  Therefore, the algorithm to generate random samples of TSSL and TSCN models is as follows:

  - **Step 1.** Generate a random sample $U_1, U_2, \ldots, U_m$ from $H(\,\cdot\,|\boldsymbol{\nu})$.
  - **Step 2.** Generate a random sample $Z_1, Z_2, \ldots, Z_m$ from $\text{TSN}(0, \sigma^2, \lambda; [\gamma_1, \gamma_2])$, where $\gamma_1 = (a-\mu)U^{1/2}$ and $\gamma_2 = (b-\mu)U^{1/2}$, using Method 1.
  - **Step 3.** Using the stochastic representation given in (1.1.2), set $Y = \mu + U^{1/2}Z$.

  Consequently, we draw $y_i^{(j)}$ from $f(y_i|\boldsymbol{\theta}^{*(j)}, V_i, \rho_i)$ in the S-step.

2.2.3 Model selection

Because there is no universal criterion for mixture model selection, we chose three criteria to compare the models considered in this work. These are the Akaike information criterion (AIC) (Akaike, 1974), the Bayesian information criterion (BIC) (Schwarz, 1978) and the efficient determination criterion (EDC) (Bai et al., 1989). Like the more popular AIC and BIC criteria, EDC has the form

$$-2\ell(\widehat{\boldsymbol{\theta}}) + \rho c_n,$$

where $\ell(\boldsymbol{\theta})$ is the actual log-likelihood, ρ is the number of free parameters that have to be estimated in the model and the penalty term c_n is a convenient sequence of positive numbers. Here, we use $c_n = 0.2\sqrt{n}$, a proposal that was considered in Basso et al. (2010) and Cabral et al. (2012). We have $c_n = 2$ for AIC, $c_n = \log n$ for BIC, where n is the sample size.

2.3 Approximated standard errors

Standard errors of the ML estimates can be approximated by the inverse of the observed information matrix, but there is generally no closed form. Thus, we consider the same strategy used by Meilijson (1989), Lin (2010) and Garay et al. (2015) to get approximate standard errors of the parameter estimates based on the empirical information matrix.

Let $(\mathbf{V}, \boldsymbol{\rho})$ be the vector of observed data. So, considering $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \lambda, \boldsymbol{\nu})$, $\mathbf{Y}_{comp} = (\mathbf{V}^\top, \boldsymbol{\rho}^\top, \mathbf{Y}^\top, \mathbf{t}^\top, \mathbf{u}^\top)^\top$ and relations described in the Equation (2.2.5), the empirical information matrix is defined as

$$\mathbf{I}_e\left(\boldsymbol{\theta}|\mathbf{V}, \boldsymbol{\rho}\right) = \sum_{i=1}^{n} \mathbf{s}\left(V_i, \rho_i|\boldsymbol{\theta}\right) \mathbf{s}^\top\left(V_i, \rho_i|\boldsymbol{\theta}\right) - \frac{1}{n}\mathbf{S}\left(\mathbf{V}, \boldsymbol{\rho}|\boldsymbol{\theta}\right) \mathbf{S}^\top\left(\mathbf{V}, \boldsymbol{\rho}|\boldsymbol{\theta}\right),$$

where $\mathbf{S}^\top\left(\mathbf{V}, \boldsymbol{\rho}|\boldsymbol{\theta}\right) = \sum_{i=1}^{n} \mathbf{s}\left(V_i, \rho_i|\boldsymbol{\theta}\right)$. It is noted from the result of Louis (1982) that, the individual score can be determined as

$$\mathbf{s}\left(V_i, \rho_i|\boldsymbol{\theta}\right) = \frac{\partial \ell(\boldsymbol{\theta}|V_i, \rho_i)}{\partial \boldsymbol{\theta}} = \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})}{\partial \boldsymbol{\theta}}|V_i, \rho_i, \boldsymbol{\theta}\right]. \tag{2.3.1}$$

Thus, substituting the ML estimates of $\boldsymbol{\theta}$ in (2.3.1), the empirical information matrix $\mathbf{I}_e\left(\boldsymbol{\theta}|\mathbf{V}, \boldsymbol{\rho}\right)$ is reduced to

$$\mathbf{I}_e\left(\widehat{\boldsymbol{\theta}}|\mathbf{V}, \boldsymbol{\rho}\right) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^\top, \tag{2.3.2}$$

where $\widehat{\mathbf{s}}_i = \left(\widehat{\mathbf{s}}_{\boldsymbol{\beta}_i}, \widehat{\mathbf{s}}_{\sigma_i^2}, \widehat{\mathbf{s}}_{\boldsymbol{\lambda}_i}, \widehat{\mathbf{s}}_{\boldsymbol{\nu}_i}\right)$ is an individual score vector and

$$\widehat{\mathbf{s}}_{\boldsymbol{\beta}_i} = \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})}{\partial \boldsymbol{\beta}}|V_i, \rho_i, \widehat{\boldsymbol{\theta}}\right] = \frac{1+\widehat{\lambda}^2}{\widehat{\sigma}^2}\left(\mathbf{x}_i\mathcal{E}_{01i}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}} - \widehat{\sigma}\frac{\widehat{\lambda}}{\sqrt{1+\widehat{\lambda}^2}}\mathbf{x}_i\mathcal{E}_{10i}(\widehat{\boldsymbol{\theta}})\right),$$

$$\widehat{\mathbf{s}}_{\sigma_i^2} = \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})}{\partial \sigma^2}|V_i, \rho_i, \widehat{\boldsymbol{\theta}}\right] = -\frac{1}{2\widehat{\sigma}^2} + \frac{1+\widehat{\lambda}^2}{2\widehat{\sigma}^4}\left(\mathcal{E}_{02i}(\widehat{\boldsymbol{\theta}}) - 2\mathcal{E}_{01i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}} + \mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}})^2\right)$$
$$-\frac{\widehat{\lambda}\sqrt{(1+\widehat{\lambda}^2)}}{2\widehat{\sigma}^3}\left(\mathcal{E}_{11i}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{10i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}\right),$$

$$\widehat{\mathbf{s}}_{\lambda_i} = \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})}{\partial \lambda}|V_i, \rho_i, \widehat{\boldsymbol{\theta}}\right] = \frac{\widehat{\lambda}}{1+\widehat{\lambda}^2} + \frac{\widehat{\lambda}}{\widehat{\sigma}^2}\left(\mathcal{E}_{02i}(\widehat{\boldsymbol{\theta}}) - 2\mathcal{E}_{01i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}} + \mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}})^2\right)$$

$$+\frac{1+2\widehat{\lambda}^2}{\widehat{\sigma}\sqrt{1+\widehat{\lambda}^2}}\left(\mathcal{E}_{11i}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{10i}(\widehat{\boldsymbol{\theta}})\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}\right) - \widehat{\lambda}\mathcal{E}_{20i}(\widehat{\boldsymbol{\theta}}),$$

$$\widehat{\mathbf{s}}_{\boldsymbol{\nu}_i} = \mathrm{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})}{\partial \boldsymbol{\nu}}|V_i, \rho_i, \widehat{\boldsymbol{\theta}}\right] = \mathrm{E}\left[\frac{\partial \log\left(f(U_i|\boldsymbol{\nu})\right)}{\partial \boldsymbol{\nu}}|V_i, \rho_i, \widehat{\boldsymbol{\theta}}\right],$$

where $\ell_c(\boldsymbol{\theta}|\mathbf{Y}_{comp_i})$ is the log-likelihood formed from the single complete observation i and $\mathcal{E}_{rsi}(\boldsymbol{\theta}^{*(k)}) = \mathrm{E}[U_iT_i^rY_i^s|V_i, \rho_i, \boldsymbol{\theta}^{*(k)}]$. It is important to stress that the standard error of $\boldsymbol{\nu}$ depends heavily on the calculation of $\mathrm{E}\left[\log(U_i)\,|V_i, \rho_i, \widehat{\boldsymbol{\theta}}\right]$, which relies on computationally intensive Monte Carlo integrations. In our analysis, we focus solely on comparing the standard errors of $\boldsymbol{\beta}^\top$, σ^2 and λ.

2.4 Simulation studies

In order to examine the performance of our proposed models and algorithm, we present three simulation studies. The first compare the performance of the estimates for SMSN-CR models in the presence of outliers on the response variable. The second study shows that our proposed SAEM algorithm estimates do provide good asymptotic properties. In the the third study we show the consistency of the approximate standard errors for the ML estimates of parameters. All computational procedures were implemented using the `R` software (R Development Core Team, 2015). We performed all Monte Carlo simulation studies considering the model SMSN-CR, defined by combining (2.2.1)–(2.2.3) where $\boldsymbol{\beta}^\top = (\beta_1, \beta_2) = (1, 4)$, $\sigma^2 = 2$, $\lambda = 4$ and $\mathbf{x}_i^\top = (1, x_i)$. The values x_i, $i = 1, \ldots, n$, were generated independently from a uniform distribution on the interval (2,20) and those values were kept constant throughout the experiment. For all simulation studies, were considered a random sample with censoring levels $p = 0\%,\ 8\%,\ 20\%$ and 35% (i.e., 0%, 8%, 20% and 35% of the observations in each dataset were censored respectively). In addition, we also choose the parameters $m = 20,\ c = 0.3$ and $S = 400$ for the SAEM implementation.

2.4.1 Robustness of the SAEM estimates (Simulation study 1)

The purpose of this simulation study is to compare the performance of the estimates for some censored regression models in the presence of outliers on the response variable. We consider the different cases of the SMSN-CR models with fixed $\boldsymbol{\nu}$, i.e., SN-CR, ST-CR ($\nu = 3$), SSL-CR ($\nu = 3$) and SCN-CR ($(\nu, \gamma) = (0.1, 0.1)$).

For this case, we generated 200 samples of size $n = 300$ under the SN-CR model with $\varepsilon_i \sim \text{SN}(-\sqrt{\frac{2}{\pi}}k_1\Delta, \sigma^2, \lambda)$ and four levels of percentage to the response variable with left censored values, in each sample. To assess how much the SAEM estimates are influenced by the presence of outliers, we replaced the observation y_{150} by $y_{150}(\vartheta) = y_{150} + \vartheta$, with $\vartheta = 1, 2, \ldots, 10$. For each replication, we obtained the parameter estimates with and without outliers, under the four SMSN-CR models. We are interested in evaluating the relative change in the estimates as a ϑ function. Given $\boldsymbol{\theta} = (\beta_1, \beta_2, \sigma^2, \lambda)$, the relative change is defined by

$$RC\left(\widehat{\theta}_i(\vartheta)\right) = \left|\frac{\left(\widehat{\theta}_i(\vartheta) - \widehat{\theta}_i\right)}{\widehat{\theta}_i}\right|,$$

where $\widehat{\theta}_i(\vartheta)$ and $\widehat{\theta}_i$ denote the SAEM estimates of θ_i with and without perturbation, respectively.

Figure 2.2 show the average values of the relative changes undergone by all the parameters, for the censoring level of 8%. We note that for all parameters, the average relative changes suddenly increase under SN-CR model as the ϑ value grows. In contrast, for the SMSN-CR models with heavy tails, namely the ST-CR, SSL-CR and SCN-CR, the measures vary little, indicating they are more robust than the SN-CR model in the ability to accommodate discrepant observations. We also conducted simulations with three censoring rates ($p = 0\%,\ 20\%$ and 35%), obtaining similar results, as shown in Figures A.1, A.2 and A.3 in Appendix A.

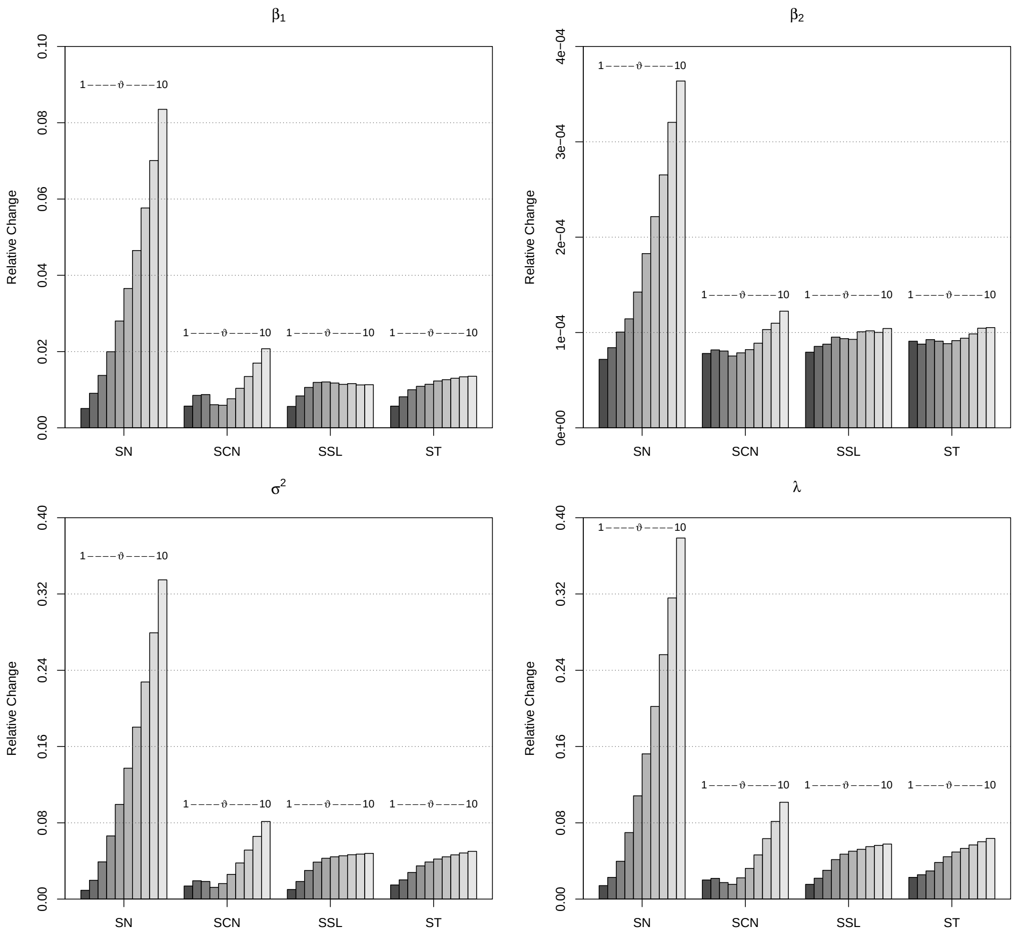

Figure 2.2: Simulation study 1. Average relative changes on estimates for different perturbations ϑ and censoring level $p = 8\%$.

2.4.2 Asymptotic properties (Simulation study 2)

In this simulation study, the main focus is to evaluate the finite-sample performance of the parameter estimates. To do so we generated left-censored samples from the SMSN-CR model with the different censoring levels $p = 8\%,\ 20\%$ and 35% and sample sizes fixed at $n = 50, 150, 300, 450, 600$ and 750. For each combination of sample size and censoring level, we generated 500 samples from the SMSN-CR models, under four different situations: SN-CR, ST-CR ($\nu = 3$), SSL-CR ($\nu = 4$) and SCN-CR ($\boldsymbol{\nu}^\top = (0.1, 0.1)$).

As in Garay et al. (2015), to evaluate the estimates obtained by the proposed SAEM algorithm, we compared the bias (Bias) and the mean square error (MSE) for each parameter over the 500 replicates. They are defined as

$$\text{Bias}(\theta_i) = \frac{1}{500}\sum_{j=1}^{500}(\widehat{\theta}_i^{(j)} - \theta_i) \quad \text{and} \quad \text{MSE}(\theta_i) = \frac{1}{500}\sum_{j=1}^{500}(\widehat{\theta}_i^{(j)} - \theta_i)^2,$$

where $\widehat{\theta}_i^{(j)}$ is the estimate of θ_i from the j-th sample for $j = 1, \ldots, 500$.

Analyzing Figures 2.3 and 2.4, for the censoring level $p = 8\%$, it can be seen that the Bias and MSE tend to zero in all SMSN-CR models when n increases. Thus, as a general rule the results indicate that the ML estimates of the SMSN-CR models do provide good asymptotic properties. We also performed simulations with two higher censoring rates ($p = 20\%$ and 35%) and the patterns of convergence still behaved well (See Figures A.4 - A.7 given in Appendix A for more details).

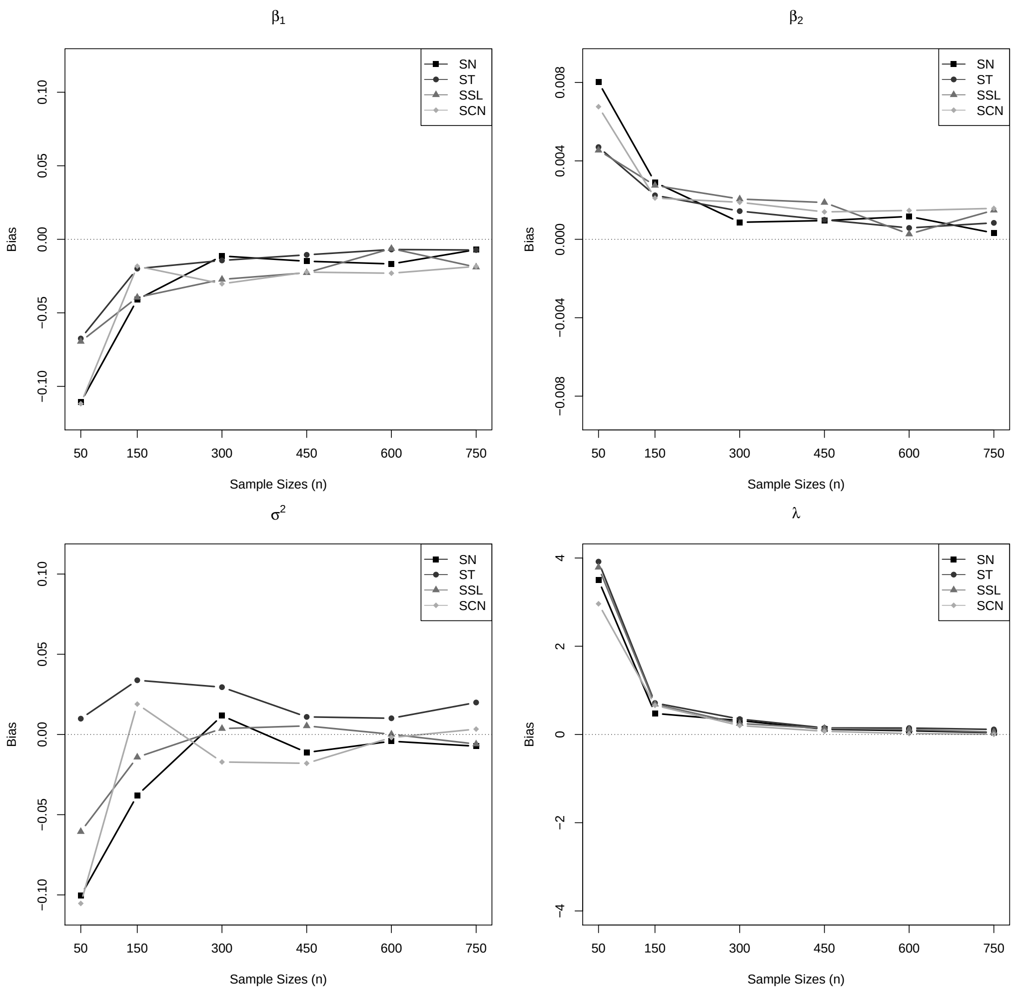

Figure 2.3: Simulation study 2. Bias of parameters β_1, β_2, σ^2 and λ for SMSN models with level of censoring $p = 8\%$.

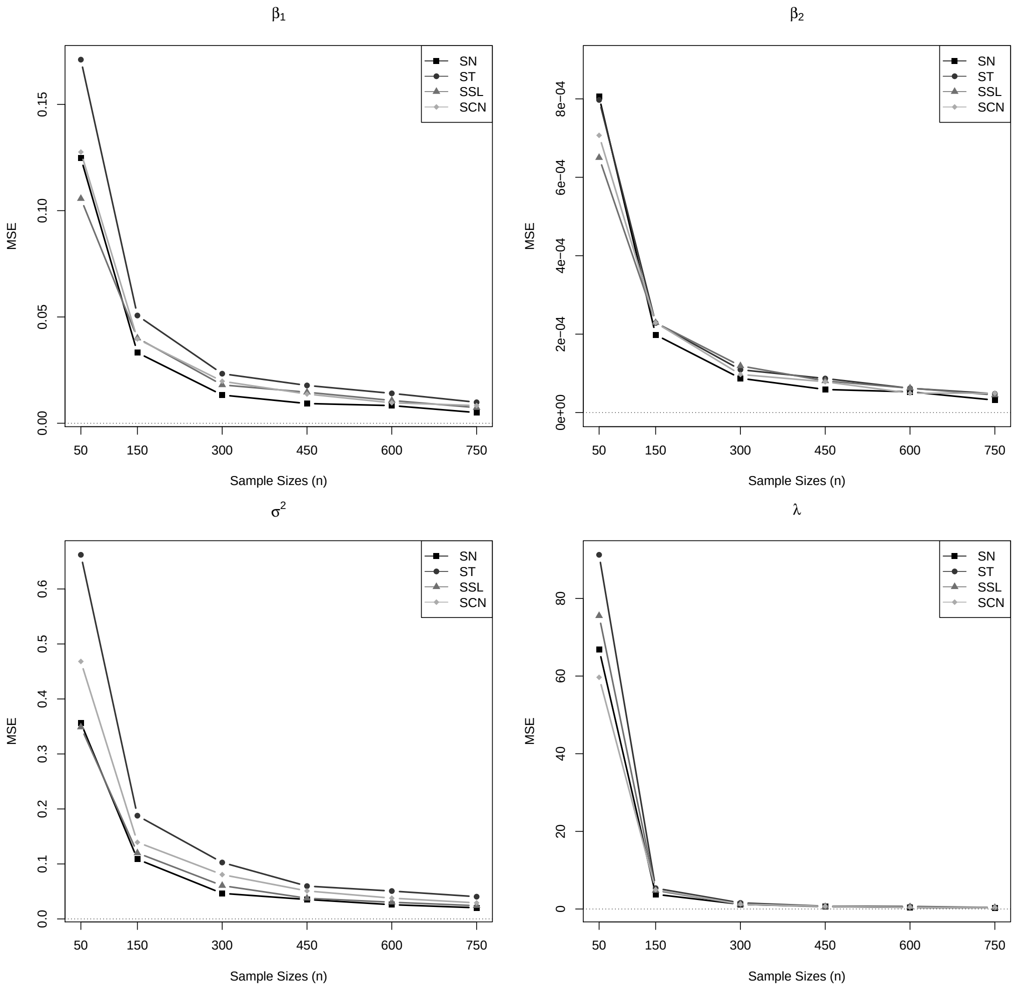

Figure 2.4: Simulation study 2. MSE of parameters β_1, β_2, σ^2 and λ for SMSN models with level of censoring $p = 8\%$.

2.4.3 Consistency of the estimates of the standard errors (Simulation study 3)

The design considered in this simulation study is the same as used in Subsection 2.4.1. Here, we examine the consistency of the approximation method, suggested in Section 2.3, to get the standard errors (SE) of ML estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma^2}, \widehat{\lambda})$ for the SMSN-CR models, considering four censoring levels $p = 0\%,\ 8\%,\ 20\%$ and 35%.

We generated 500 random samples of size $n = 450$ for the different SMSN-CR models: SN-CR, ST-CR ($\nu = 3$) , SSL-CR ($\nu = 4$) and SCN-CR ($\boldsymbol{\nu}^\top = (0.1, 0.1)$). For each sample, we obtained the ML estimates of $\boldsymbol{\theta} = (\beta_1, \beta_2, \sigma^2, \lambda)$, their SE using the technique proposed in Section 2.3 and the 95% normal approximation confidence intervals for each parameter, i.e., $\widehat{\theta} \pm 1.96$SE.

Considering all the ML estimates obtained (across 500 samples), we computed:

- the Monte Carlo standard deviation of $\widehat{\theta}_i$, defined by

$$\text{MC-Sd} = \sqrt{\frac{1}{499}\left[\sum_{j=1}^{500}\left(\widehat{\theta}_i^{(j)}\right)^2 - 500\left(\overline{\widehat{\theta}_i}\right)^2\right]} \quad \text{where} \quad \overline{\widehat{\theta}_i} = \frac{1}{500}\sum_{j=1}^{500}\widehat{\theta}_i^{(j)};$$

- the average values of the approximate standard errors of the SAEM estimates obtained through the method described in Subsection 2.2.1 using the empirical information matrix, denoted by AV-SE, and

- the percentage of times that the confidence intervals cover the true value of the parameter (COV MC).

Table 2.1 shows that in general, the COV MC for the parameters is quite stable for the censoring levels $p = 0\%,\ 8\%,$ and 20%, but it tends to be lower than the nominal level (95%) when considering a high level of censoring, say $p = 35\%$. This table also provides the average values of the approximate standard errors of the EM estimates obtained through the information-based method described in Subsection 2.2.1 (AV SE) and the Monte Carlo standard deviation (MC Sd) for the parameters. Table 2.1 also reveals that the estimation method of the standard errors provides relatively close results for the SMSN models, indicating that the proposed empirical information matrix(Equation 2.3.2) is reliable.

Table 2.1: Simulation study 3. Results based on 500 simulated samples. MC Sd, AVE SE and COV MC are the respective average of the standard deviations, the average of the approximate standard error obtained through the information-based method and the coverage probability from fitting SMSN-CR models under various levels of censoring proportion.

Cens. Level	$\widehat{\theta}_i$	SN-CR			ST-CR		
		MC-Sd	AV-SE	COV MC	MC-Sd	AV-SE	COV MC
0%	$\widehat{\beta}_1$	0.0897	0.0847	92.60%	0.1099	0.1126	95.40%
	$\widehat{\beta}_2$	0.0070	0.0066	92.80%	0.0072	0.0076	97.00%
	$\widehat{\sigma}^2$	0.1803	0.1828	95.40%	0.2422	0.2411	95.40%
	$\widehat{\lambda}$	0.7739	0.7590	95.60%	0.7978	0.7884	95.00%
8%	$\widehat{\beta}_1$	0.0952	0.0981	96.00%	0.1331	0.1290	94.60%
	$\widehat{\beta}_2$	0.0076	0.0077	94.40%	0.0093	0.0089	93.80%
	$\widehat{\sigma}^2$	0.1878	0.1878	92.80%	0.2443	0.2484	95.60%
	$\widehat{\lambda}$	0.8127	0.7856	95.80%	0.8468	0.8460	95.00%
20%	$\widehat{\beta}_1$	0.1319	0.1314	94.20%	0.1612	0.1562	94.00%
	$\widehat{\beta}_2$	0.0095	0.0096	95.00%	0.0106	0.0105	94.20%
	$\widehat{\sigma}^2$	0.1827	0.2035	97.60%	0.2902	0.2653	92.60%
	$\widehat{\lambda}$	0.7963	0.8822	97.60%	0.9836	0.9104	92.80%
35%	$\widehat{\beta}_1$	0.1922	0.1881	94.60%	0.2452	0.2344	90.80%
	$\widehat{\beta}_2$	0.0134	0.0130	94.40%	0.0158	0.0153	93.40%
	$\widehat{\sigma}^2$	0.2185	0.2262	95.20%	0.2866	0.2835	93.40%
	$\widehat{\lambda}$	0.8434	0.9606	94.60%	0.7929	0.9635	92.20%
Cens. Level	$\widehat{\theta}_i$	SCN-CR			SSL-CR		
		MC-Sd	AV-SE	COV MC	MC-Sd	AV-SE	COV MC
0%	$\widehat{\beta}_1$	0.0982	0.0995	95.40%	0.1001	0.0953	92.80%
	$\widehat{\beta}_2$	0.0072	0.0074	96.40%	0.0076	0.0075	94.40%
	$\widehat{\sigma}^2$	0.2260	0.2184	94.40%	0.2030	0.1921	92.80%
	$\widehat{\lambda}$	0.8461	0.8028	95.00%	0.8115	0.7716	95.00%
8%	$\widehat{\beta}_1$	0.1152	0.1151	93.80%	0.1188	0.1176	94.40%
	$\widehat{\beta}_2$	0.0088	0.0086	95.20%	0.0088	0.0088	93.20%
	$\widehat{\sigma}^2$	0.2251	0.2244	93.80%	0.1940	0.1988	95.60%
	$\widehat{\lambda}$	0.8343	0.8303	94.20%	0.7410	0.8023	96.60%
20%	$\widehat{\beta}_1$	0.1372	0.1405	92.60%	0.1484	0.1523	94.80%
	$\widehat{\beta}_2$	0.0103	0.0102	93.20%	0.0105	0.0109	94.60%
	$\widehat{\sigma}^2$	0.2498	0.2431	95.80%	0.2080	0.2129	94.60%
	$\widehat{\lambda}$	0.8607	0.8978	95.20%	0.8307	0.8785	97.00%
35%	$\widehat{\beta}_1$	0.2032	0.2122	88.20%	0.2037	0.2125	93.20%
	$\widehat{\beta}_2$	0.0139	0.0142	90.60%	0.0139	0.0145	93.40%
	$\widehat{\sigma}^2$	0.2657	0.2653	93.80%	0.2304	0.2354	93.20%
	$\widehat{\lambda}$	0.8156	0.9845	93.80%	0.8894	0.9914	96.40%

2.5 Application: Wage rate dataset

In this section we provide an application of the results derived in the previous sections using the data described by Mroz (1987). The dataset consists of 753 married white women with ages between 30 and 60 years old in 1975, with 428 women who worked at some point during that year. The response variable is the wage rate, which represents a measure of the wage of the housewife known as the average hourly earnings. If the wage rates are set equal to zero, these wives did not work in 1975. Therefore, these observations are considered left censored at zero.

The variables involved in the study were:

- y_i: defined as the average hourly earnings (wage rates);
- x_{i1}: wife's age;
- x_{i2}: wife's years of schooling;
- x_{i3}: the number of children younger than six years old in the household;
- x_{i4}: the number of children between the ages of six and nineteen.

These data were analyzed by Arellano-Valle et al. (2012) using the Student-t censored regression model; by Garay et al. (2015) considering SMN-CR models and, more recently by Massuia et al. (2015) to evaluated the performance of the SMSN-CR models from a Bayesian perspective. Here, we revisit this dataset in order to evaluate the performance of the proposed SAEM algorithm to obtain ML estimates.

Table 2.2 contains the ML estimates for the parameters of the four models, i.e., SN-CR, ST-CR, SSL-CR and SCN-CR models, together with their corresponding standard errors calculated via the empirical information matrix. For the ST and SSL models, the estimated value of ν is small, indicating the lack of adequacy of the skew-normal (and normal) assumption for the wage rates dataset. Moreover, the results obtained under SN-CR and ST-CR models are consistent with those presented in Massuia et al. (2015). The SCN-CR and SSL-CR models presented estimates for λ closed to zero, indicating coherence with the results presented in Garay et al. (2015). Table 2.3 compares the fit of the four SMSN models using the model selection criteria discussed in Subsection 2.2.3. Note that the SMSN distributions with heavy tails have better fit than the SN model.

Particularly, the ST distribution fits the data better than the other three distributions. The comparison process is conducted now considering the symmetric SMN distributions (vide Garay et al. (2015)) and we observe that according model selection criteria the ST-CR model still presents a better overall fit than the other four models (see Table A.1 given in Appendix A).

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

$$r_{MT_i} = \text{sign}(r_{M_i})\sqrt{-2[r_{M_i} + \rho_i \log(\rho_i - r_{M_i})]}, \quad i = 1, \ldots, n. \tag{2.5.1}$$

where $r_{M_i} = \rho_i + \log(S(y_i; \widehat{\boldsymbol{\theta}}))$ is the martingale residual, with $\rho_i = 0, 1$ indicating whether the observation is censored or not, respectively. $S(y_i, \widehat{\boldsymbol{\theta}})$ is the SAEM estimate of the survival function of y – see more details in Ortega et al. (2003) and Garay et al. (2015). The normal probability plot of the MT residuals with generated envelopes is presented in Figure 2.5. From this figure, we note that the SMSN-CR models with heavy tails present better fit than the SN-CR model.

Table 2.2: Wage rate dataset. Parameter estimates of the SMSN-CR models and SE for Wage rate dataset.

	SN-CR		**ST-CR**		**SCN-CR**		**SSL-CR**	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_1	-1.3355	1.7627	-4.1685	1.4392	-1.3291	1.5026	-1.3489	1.4221
β_2	-0.1185	0.0272	-0.0722	0.0223	-0.1061	0.0229	-0.1053	0.0226
β_3	0.6917	0.0809	0.6541	0.0576	0.6490	0.0611	0.6434	0.0620
β_4	-3.2502	0.4345	-2.5956	0.3291	-3.0685	0.3642	-3.0480	0.3628
β_5	-0.2602	0.1433	-0.2676	0.1136	-0.3016	0.1199	-0.2901	0.1176
σ^2	32.8512	2.0202	19.4969	3.1730	11.8519	3.7339	6.7930	1.1830
λ	1.5454	0.4412	-1.6976	0.2942	0.1273	1.4542	-0.2144	0.3698
ν	-	-	2.5000	-	0.0537	-	1.45	-
γ	-	-	-	-	0.0645	-	-	-

Table 2.3: Wage rate dataset. Model selection criteria (values in bold correspond to the best model).

Criteria	SN-CR	ST-CR	SCN-CR	SSL-CR
log-likelihood	-1470.617	**-1410.583**	-1430.992	-1435.426
AIC	2955.234	**2837.166**	2879.984	2884.852
BIC	2987.602	**2874.159**	2921.601	2917.220
EDC	2979.651	**2865.071**	2911.378	2909.269

2.6 Conclusions

In this chapter, we have proposed a linear regression models with censored responses based on scale mixtures of skew-normal distributions, denoted by SMSN-CR, as a replacement for the conventional choice of normal (or symmetric) distribution for censored linear models. Our results generalize the works of Barros et al. (2010), Arellano-Valle et al. (2012), Massuia et al. (2014) and Garay et al. (2015) from a frequentist point of view. In the context of SMSN-CR models, a Bayesian analysis was developed recently by Massuia et al. (2015). However, to the best of our knowledge, there are no previous studies of a likelihood based perspective related to this topic.

In order to explore the performance of our proposed models and SAEM algorithm, we developed three simulation studies. The study compared the performance of the estimates for SMSN-CR models in the presence of outliers on the response variable. The second study showed that our proposed SAEM algorithm estimates do provide good asymptotic properties. The third study showed the consistency of the approximate of standard errors for the ML estimates of parameters. We also applied our method to the wage rate dataset of Mroz (1987), in order to illustrate how the procedure developed can be used to evaluate model assumptions and obtain robust parameter estimates. As expected, our proposed SMSN-CR with heavy tails models, as ST-CR, SSL-CR and SCN-CR models, present better results than the SN-CR model. It is interesting to note that the ST-CR model still presents a better overall fit than the symmetrical SMN-CR models.

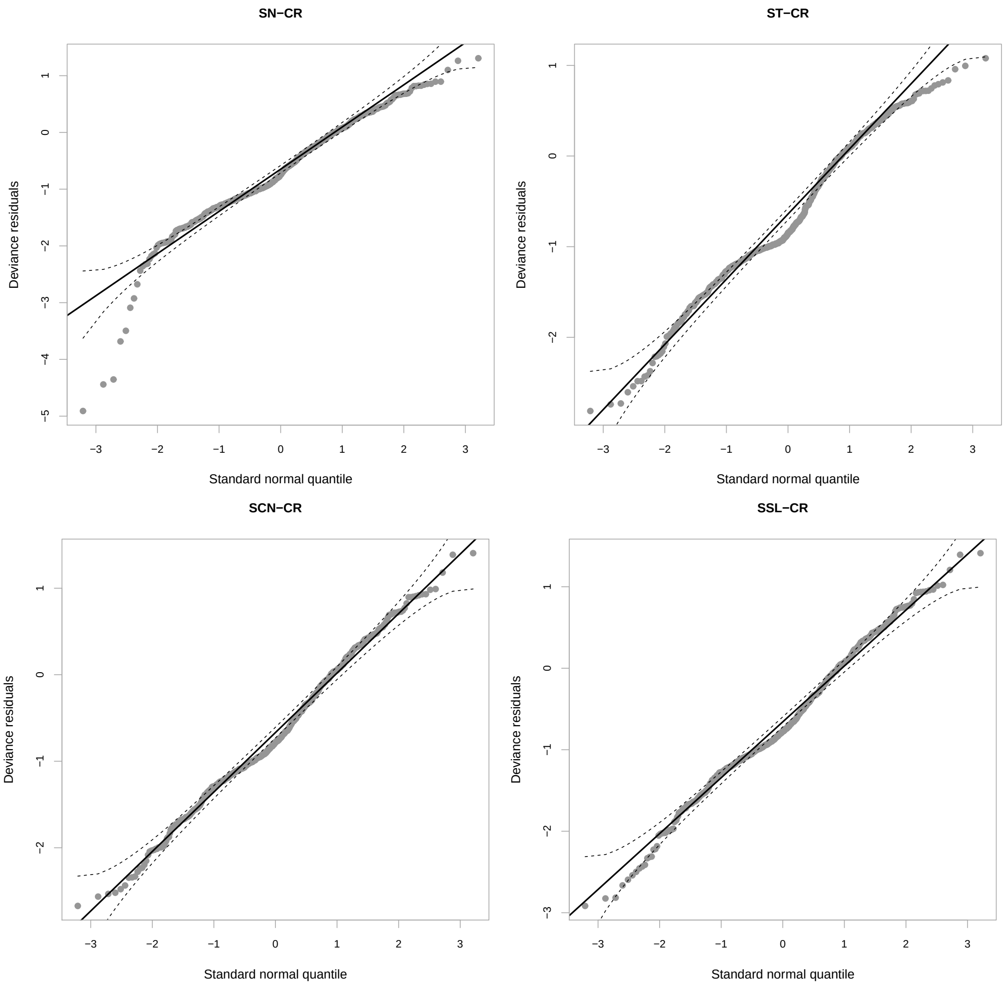

Figure 2.5: Wage rate dataset. Envelopes of the MT residuals for the SMSN-CR models

Chapter 3

Diagnostic Analysis

3.1 Introduction

In the framework of censored regression models, the random errors are routinely assumed to have a normal distribution for mathematical convenience. However, it is well known that several phenomena do not always fit under the assumptions of the normal model, yielding data with a distribution having simultaneously heavier tails and skewness. A good alternative is to consider observational errors with scale mixtures of skew normal (SMSN) distributions, so that the SMSN censored regression (SMSN-CR) model is defined. In Chapter 2 we developed a full likelihood approach for SMSN-CR models, including the implementation of the SAEM algorithm for maximum likelihood (ML) estimation with the likelihood function, predictions of unobservable values of the response and asymptotic standard errors as byproducts.

Since the classic normal model is very sensitive to outlying observations, the assessment of robustness aspects of the parameter estimates is an important concern. The deletion method, which consists of studying the impact on the parameter estimates after dropping individual observations, is probably the most employed technique to detect influential observations – see Cook and Weisberg (1982) and the references therein. Nevertheless, research on the influence of small perturbations in the model/data on the parameter estimates has received increasing attention in recent years. This can be achieved by performing local influence analysis, a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs. Following the pioneering work of Cook (1986), this area of research has received considerable attention in the statistical literature in linear regression models. However, for the SMSN-CR model the marginal log-likelihood function is complex for many applications, and a direct application of Cook's approach may be very difficult, because these measures involve the first and second partial derivatives of this function. The work of Zhu and Lee (2001) presents an approach to perform local influence analysis for general statistical models with missing data by working with a Q-displacement function, closely related to the conditional expectation of the complete-data log-likelihood at the E-step of the SAEM algorithm. This approach produces results very similar to those obtained from Cook's method. Moreover, the case-deletion can be studied by the Q-displacement function following the approach of Zhu et al. (2001) and Zhu et al. (2009). So, we develop in this chapter methods to obtain case-deletion measures and local influence measures by using the method of Zhu et al. (2001) (see also Zhu and Lee, 2001; Lee and Xu, 2004) in the context of regression models with censored data. This method or modifications of it have been applied successfully to perform influence analysis in several regression models, see for example Bolfarine et al. (2007), Ying-Zi et al. (2009), Zeller et al. (2010), Zeller et al. (2011), Lachos et al. (2011), Santana et al. (2011), Matos et al. (2013), among others. Using this general method and also applying the method of Lee and Xu (2004), in this chapter we develop a local influence approach for SMSN-CR models and show that it leads to simple influence measures.

3.2 Case-deletion Measures

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

$$\widetilde{\boldsymbol{\theta}}_{[i]} = \widehat{\boldsymbol{\theta}} + \left\{-\ddot{Q}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)\right\}^{-1} \dot{Q}_{[i]}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right), \quad \text{for} \quad i = 1, 2, \ldots, n, \tag{3.2.1}$$

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

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

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

where

$$\widehat{E}_{1[i]} = \sum_{j\neq i}\left[\mathbf{x}_j\mathcal{E}_{01j}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{00j}(\widehat{\boldsymbol{\theta}})\mathbf{x}_j\mathbf{x}_j^\top\widehat{\boldsymbol{\beta}} - \frac{\widehat{\sigma}\widehat{\lambda}}{\sqrt{\widehat{\lambda}^2+1}}\mathbf{x}_j\mathcal{E}_{10j}(\widehat{\boldsymbol{\theta}})\right], \tag{3.2.2}$$

$$\widehat{E}_{2[i]} = \sum_{j\neq i}\left[\mathcal{E}_{02j}(\widehat{\boldsymbol{\theta}}) - 2\mathcal{E}_{01j}(\widehat{\boldsymbol{\theta}})\mathbf{x}_j^\top\widehat{\boldsymbol{\beta}} + \mathcal{E}_{00j}(\widehat{\boldsymbol{\theta}})(\mathbf{x}_j^\top\widehat{\boldsymbol{\beta}})^2\right], \tag{3.2.3}$$

and

$$\widehat{E}_{3[i]} = \sum_{j\neq i}\left[\mathcal{E}_{11j}(\widehat{\boldsymbol{\theta}}) - \mathcal{E}_{10j}(\widehat{\boldsymbol{\theta}})\mathbf{x}_j^\top\widehat{\boldsymbol{\beta}}\right]. \tag{3.2.4}$$

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

$$GD_i = \left(\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}}\right)^\top \left\{-\ddot{Q}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)\right\}\left(\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}}\right), \quad i = 1, \ldots, n. \tag{3.2.5}$$

Upon substituting (3.2.1) into (3.2.5), we obtain the following approximation of the *generalized Cook's distance*:

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

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

$$QD_i = 2\left\{Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q\left(\widehat{\boldsymbol{\theta}}_{[i]}|\widehat{\boldsymbol{\theta}}\right)\right\}. \tag{3.2.6}$$

We can compute an approximation of the likelihood displacement QD_i by substituting (3.2.1) into (3.2.6), resulting in the following approximation QD_i^1 of QD_i:

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

The approximated measures QD^1 and GD_i^1 have been satisfactorily applied in the context of censored regression models by Matos et al. (2013) and Massuia et al. (2014).

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

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

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

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

where

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

and

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

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

$$\ddot{Q}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) = \mathrm{J}\left(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\theta}}\right)\ddot{Q}\left(\widehat{\boldsymbol{\theta}}^*|\widehat{\boldsymbol{\theta}^*}\right)\mathrm{J}\left(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\theta}}\right)^\top, \tag{3.2.7}$$

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

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

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

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

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

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

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

Following the approach of Cook (1986) and Zhu 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

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

Then, it can be shown that

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

where $\ddot{Q}\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right)$ is as defined in (3.2.7).

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

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

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

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

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:

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

whose computation is quite simple and also has the property that $0 \leq B_{f_Q,\mathbf{d}} \leq 1$. Let $\mathbf{d}_l$ be a basic perturbation vector with l-th entry equal to 1 and all other entries equal to 0. Zhu 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.

3.4 Perturbation schemes

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

For each perturbation scheme, we have the partitioned form:

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

where

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

3.4.1 Case-weight perturbation

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

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

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

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

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

3.4.2 Scale perturbation

To study the effects of departures from the assumption regarding the scale parameter σ^2, we consider the perturbation $\sigma^2(\omega_i) = \omega_i^{-1}\sigma^2$, for $i = 1,\ldots,n$. Under this perturbation scheme, the non-perturbed model is obtained when $\boldsymbol{\omega}_0 = \mathbf{1}_n$. Moreover, the perturbed Q-function is as in (2.2.8), with $\sigma^2(\omega_i)$ and $\widehat{\boldsymbol{\theta}}$ replacing σ^2 and $\boldsymbol{\theta}^{(k)}$, respectively. The matrix $\boldsymbol{\nabla}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ has the following elements:

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

3.4.3 Response perturbation

A perturbation of the response variables V_i, $i = 1, \ldots, n$, can be introduced by replacing V_i by $V_i(\omega_i) = V_i + \omega_i \mathbf{S}_v$, where $\mathbf{S}_v$ is a scale factor that can represent the standard deviation of the censored response. Now substituting $V_i(\omega_i)$ into (2.2.3), we can write the perturbed model as:

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

where $c_i(\omega_i) = c_i - \omega_i \mathbf{S}_v$ and $Y_i(\omega_i) = Y_i - \omega_i \mathbf{S}_v$. Hence, the perturbed Q-function follows (2.2.8), with $\mathcal{E}_{rsi}(\widehat{\boldsymbol{\theta}}^{(j)}) = \mathrm{E}[U_i T_i^r Y_i^s | V_i, \rho_i, \widehat{\boldsymbol{\theta}}^{(j)}]$ replaced by $\mathcal{E}_{rsi}(\widehat{\boldsymbol{\theta}}^{(j)}, \omega_i) = \mathrm{E}[U_i T_i^r Y_i^s(\omega_i) | V_i(\omega_i), \rho_i, \widehat{\boldsymbol{\theta}}^{(j)}]$. Under this perturbation scheme, the vector $\boldsymbol{\omega}_0$, representing no perturbation, is given by $\boldsymbol{\omega}_0 = \mathbf{0}$ and $\boldsymbol{\nabla}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ has the following elements:

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

3.4.4 Explanatory variables perturbation

Here, we consider the influence that perturbation of the explanatory variables can produce on the parameter estimates. In this case, we are interested in perturbing a specific explanatory variable, thus we consider the perturbation $\mathbf{x}_{i\omega}^\top = \mathbf{x}_i^\top + \omega_i \mathbf{S}_t \mathbb{1}_t^\top$, $\mathbf{S}_t$ is a scale factor that can represent the standard deviation of the t-th explanatory variable and $\mathbb{1}_t^\top = (0, \ldots, 1, \ldots, 0)$ 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 (2.2.8), switching $\mathbf{x}_{i\omega}^\top$ with $\mathbf{x}_i^\top$ and the no perturbation case is obtained by taking $\boldsymbol{\omega}_0 = \mathbf{0}$. Under this perturbation scheme, $\boldsymbol{\nabla}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ has the following elements:

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

3.5 Application: Stellar abundances dataset

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

According to Feigelson and Babu (2012) in a supervised astronomical survey where a particular property of a previously defined sample of objects is sought, some objects in the sample may be too faint to detect. Thus, the dataset contains the full sample of interest, but some objects have upper limits and others have detections. In this dataset we have 12 left-censored data points (see Figure 3.1, panel b). The predictor variable (x) is effective stellar surface temperature (in Kelvin, $\mathrm{T}_{eff}/1000$). To verify the existence of skewness in the data, Figure 3.1 (panel a) presents the histogram of the response variable and shows an apparent non-normal pattern. This figure also presents the normal quantile-

quantile (Q-Q) plot for the residuals (panel c) obtained by fitting a Gaussian regression model using the **R** package **stats**. The Q-Q plot exhibits an asymmetrical heavy-tailed behavior, suggesting that the normality assumption for the errors might be inappropriate. In addition, Figure 3.1 (panel b) also indicates it is plausible to use a linear regression for the dataset.

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

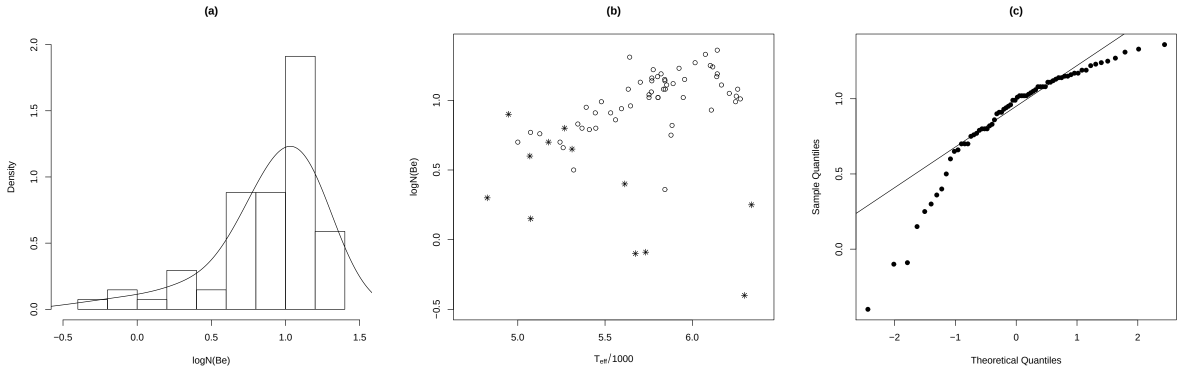

Figure 3.1: Stellar abundances dataset. (a) Histogram of the logN(Be). (b) Scatter-plot of the dataset. ($*$) represents the censored observations. (c) Normal Q-Q plot for model residuals obtained by using the **R** package **stats**.

3.5.1 ML estimates using SAEM algorithm

We fitted a regression model with an intercept parameter β_0 and applied the SAEM algorithm for censored data, as described in Subsection 2.2.1. We focus on the SN-CR, ST-CR, SCN-CR and SSL-CR distributions from the SMSN-CR class.

Table 3.1 contains the ML estimates for the parameters of the four models, together with the values of their corresponding standard errors (SE). The log-likelihood values (see column $\ell(\widehat{\boldsymbol{\theta}})$) indicate that the heavy-tailed SMSN distributions have a significantly better fit than the SN model. This finding is also confirmed by inspecting some model selection criteria, says the Akaike information criterion (AIC) (Akaike, 1974), the Bayesian information criterion (BIC) (Schwarz, 1978) and the efficient determination criterion (EDC) (Bai et al., 1989). Moreover, the SE values of the ST-CR, SCN-CR and SSL-CR models are smaller than that of the SN-CR model. These results indicate that the use of the SMSN-CR models with heavy tails produces more accurate estimates.

Table 3.1: Stellar abundances dataset. Parameter estimates of the SMSN-CR models. SE values in parentheses.

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

In order to identify atypical observations and/or model misspecification, we use the martingale-type residuals, r_{MT_i}, proposed by Barros et al. (2010) (see also Garay et al., 2015) for censored models. These residuals are defined in (2.5.1).

The normal probability plot of the MT residuals with generated envelopes is presented in Figure 3.2. We observe that the ST-CR, SCN-CR and SSL-CR models fit the data better than the SN-CR model, since, in that case, there are fewer observations which lie outside the envelopes.

In Figure 3.3, we present the Mahalanobis distance, given by $d_i^2 = \frac{(y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}})^2}{\widehat{\sigma}^2}$ vs, the estimated weights $u_i = \mathcal{E}_{00i}(\widehat{\boldsymbol{\theta}})$, for $i = 1, \ldots, 68$, considering the ST-CR, SSL-CR and SCN-CR models. We observe that when we use distributions with heavier tails than the SN one, the SAEM algorithm allows accommodating atypical observations by attributing small weights to them in the estimation procedure. The estimated weights for the SN-CR distribution are indicated as a continuous line. These results agree with similar considerations, presented for instance in Labra et al. (2012), where a nonlinear regression model under SMSN distributions is studied.

3.5.2 Diagnostic Analysis

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

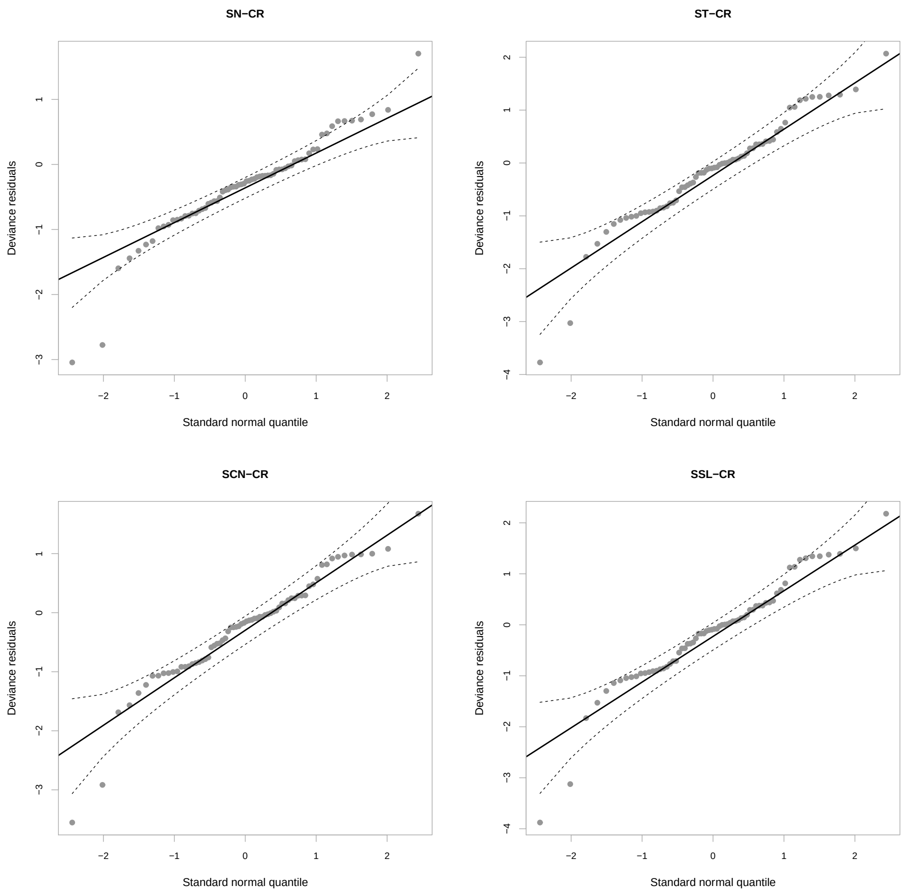

Figure 3.2: Stellar abundances dataset. Envelopes of the MT residuals for the SMSN-CR models

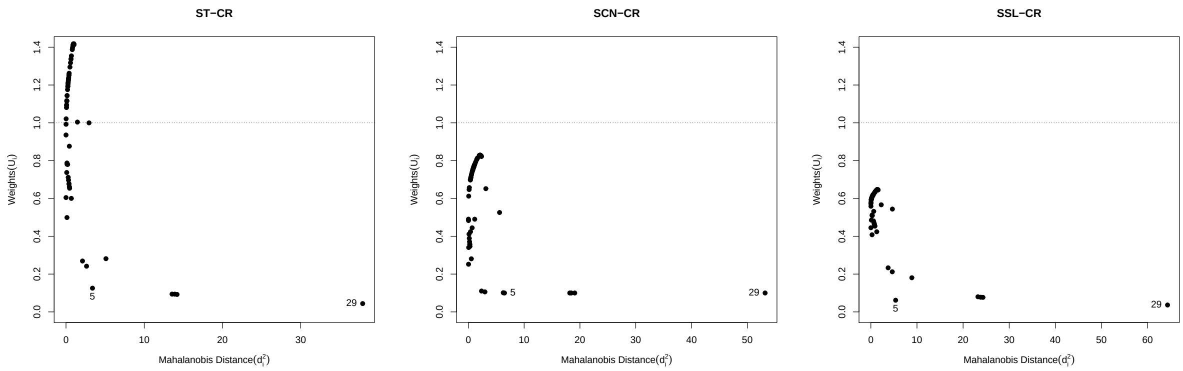

Figure 3.3: Stellar abundances dataset. Estimated u_i for the ST-CR, SCN-CR and SSL-CR models

Global Influence

In order to evaluate the effect on the ML estimates of the regression coefficients of the SMSN-CR models, when some observation is eliminated, we analyze the GD_i^1 and

QD_i^1 plot in Figures 3.4-3.5. As can be seen, case # 5 is the most influential in the estimation of the parameters, for the four SMSN-CR models considered.

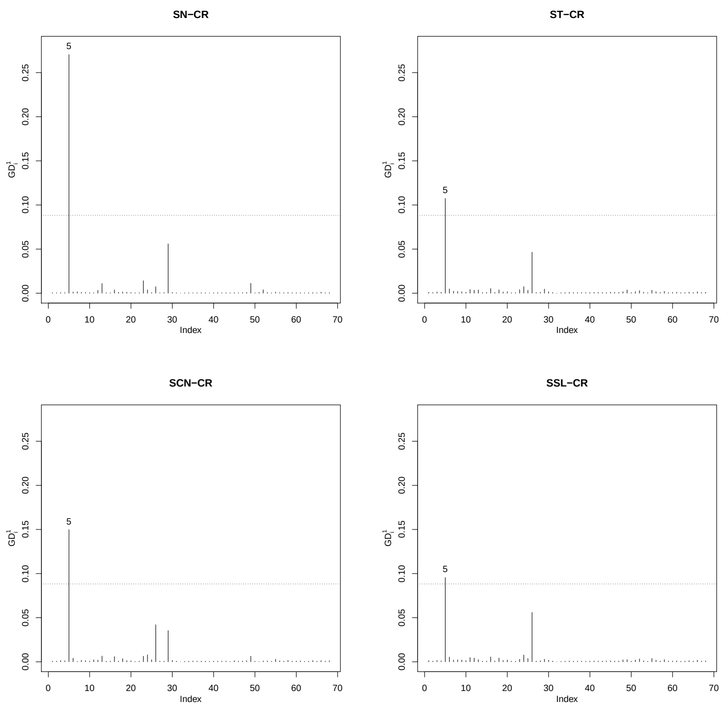

Figure 3.4: Stellar abundances dataset. Approximate generalized Cook's distance GD_i^1 for SMSN-CR models.

Local Influence

Next, we conduct a local influence study, with interest focused on $\boldsymbol{\theta}$, using the benchmark $M(0)$ from the conformal curvature $\mathbf{B}_{f_{Q,\mathbf{d}}}$, as described in Section 3.4 by considering the four different perturbation schemes. Thus, here we present a local influence analysis, using $c^* = 4$ to compute the benchmark $M(0)$.

From Figure 3.6, we observe that under the case-weight perturbation, case # 5 is identified as influential for the four SMSN-CR models considered. Under the scale perturbation, Figure 3.7, case # 29 appears influential in the ML estimates of $\boldsymbol{\theta}$ for the SN-CR, ST-CR and SCN-CR models. In addition, from Figure 3.8, this same observation # 29 is considered as influential for the SN-CR model, under the response variable perturbation. Finally, from Figure 3.9, we have that under explanatory variable perturbation no observations are considered potentially influential.

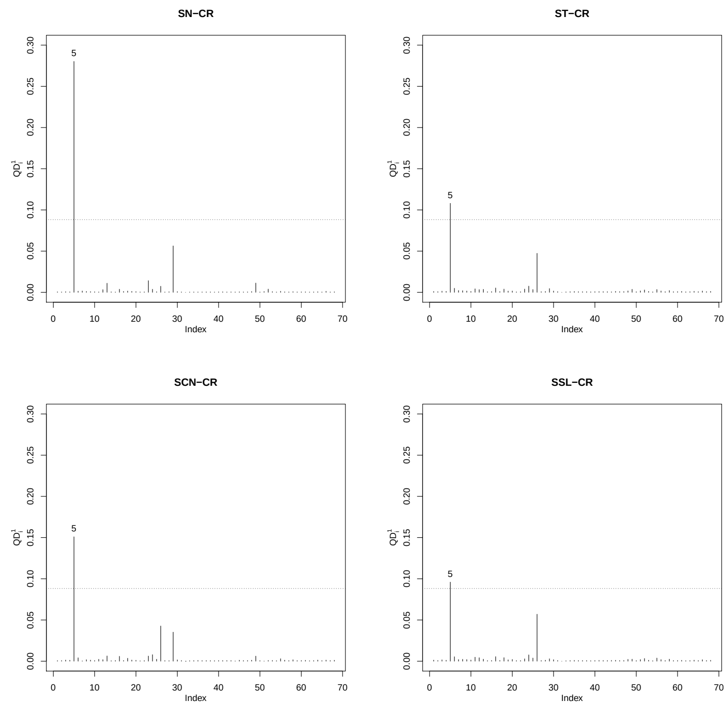

Figure 3.5: Stellar abundances dataset. Approximate likelihood displacement QD_i^1 for SMSN-CR models.

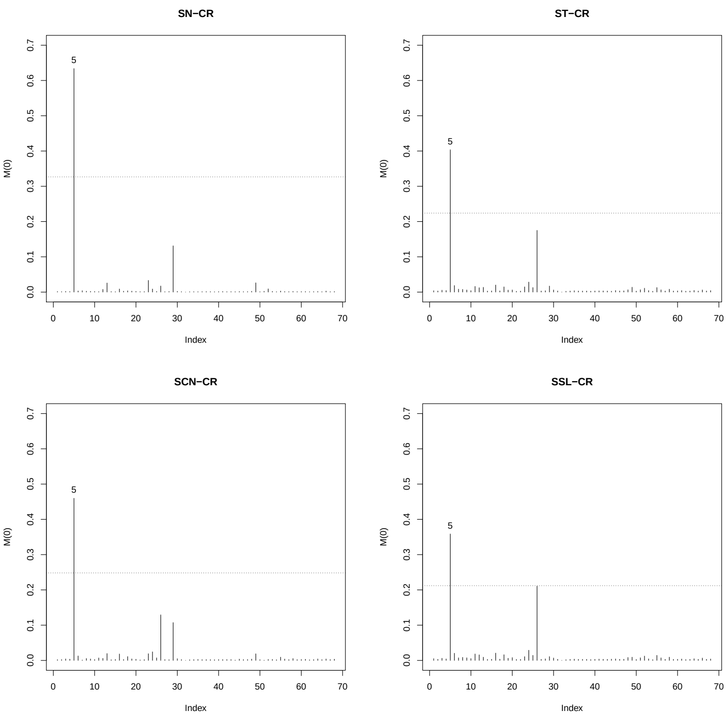

Figure 3.6: Stellar abundances dataset. Index plots of M(0) under the case-weight perturbation for SMSN-CR models.

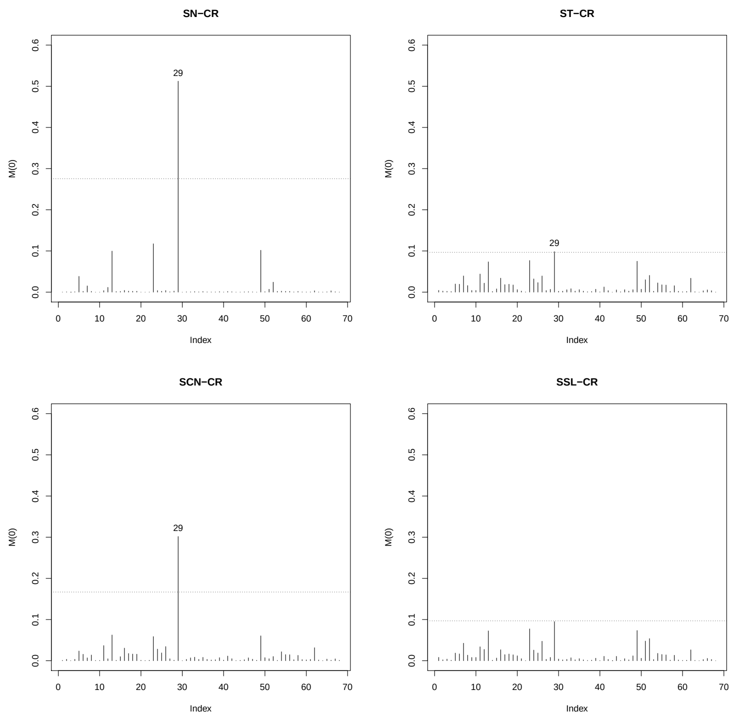

Figure 3.7: Stellar abundances dataset. Index plots of M(0) under scale perturbation for SMSN-CR models.

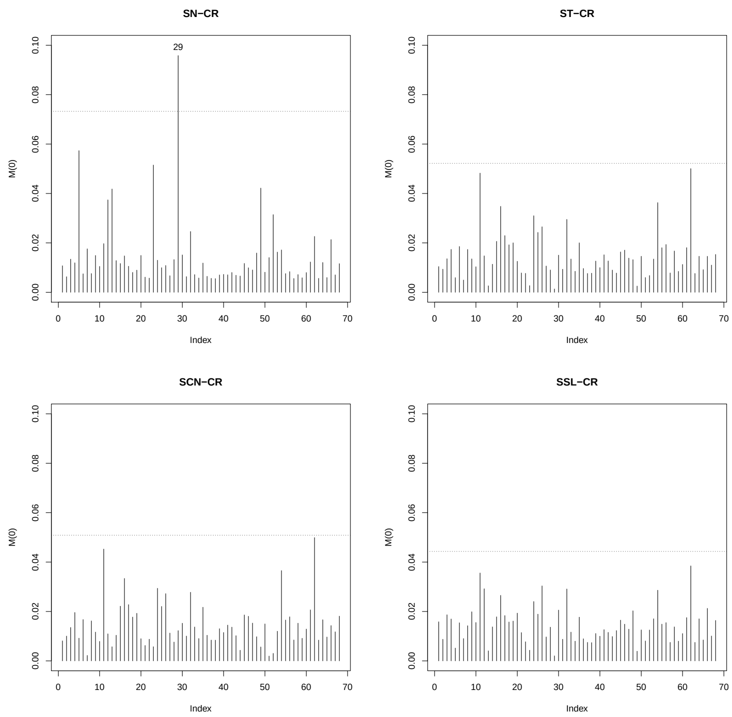

Figure 3.8: Stellar abundances dataset. Index plots of M(0) under response perturbation for SMSN-CR models.

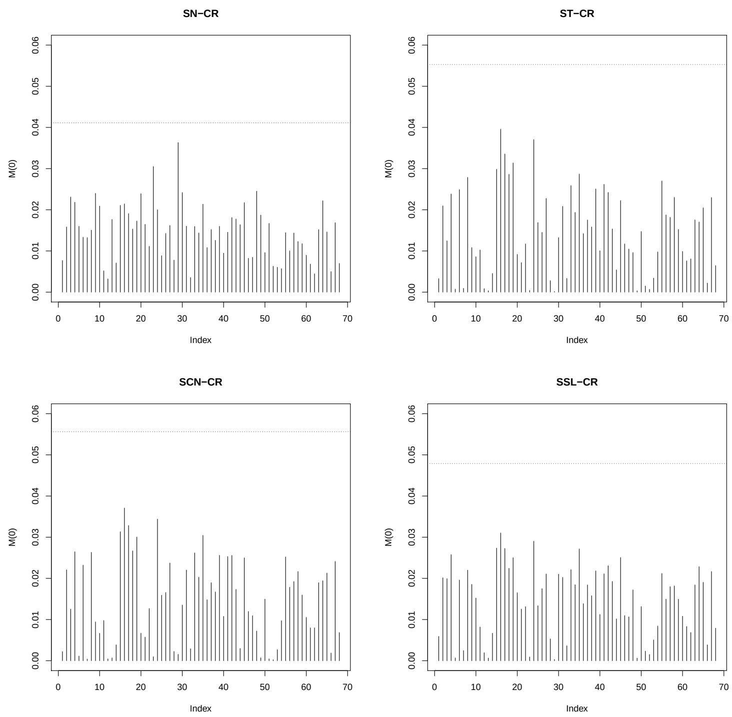

Figure 3.9: Stellar abundances dataset. Index plots of M(0) under explanatory variable perturbation for SMSN-CR models.

Impact of the detected influential observations

Table 3.2 shows that based in the global and local influence diagnostics, cases: # 5 and # 29 are detected as potentially influential observations under different perturbation schemes. In order to assess the impact of these possible influential observations on the ML estimates, we refitted the model dropping each one of these cases. Thus, in Table 3.3 we present the relative changes (RC) of these estimates, $\mathrm{RC}(\widehat{\boldsymbol{\theta}}) = \left|\frac{(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{[j]})}{\widehat{\boldsymbol{\theta}}}\right|$, where $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2, \lambda)$ and $\widehat{\boldsymbol{\theta}}_{[j]}$ denotes the SAEM estimate of $\boldsymbol{\theta}$ with the $j-$th observation removed. We observe that observation # 5, detected as influential under global and local influence diagnostics, causes a significant change in the parameters σ^2 and λ, and observation # 29 causes significant changes (in particular) to the parameter σ^2.

3.6 Simulation study

In order to examine the performance of the proposed methods, we present a simulation study to show the capacity of the method to detect atypical data. We consider the SMSN-CR model, defined by combining equations (2.2.1)-(2.2.3), where

Table 3.2: Stellar abundances dataset. Influential observations for SMSN-CR models

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

Table 3.3: Stellar abundances dataset. Relative change for the dataset.

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

$\boldsymbol{\beta}^\top = (\beta_0, \beta_1) = (3, -1)$, $\sigma^2 = 2$, $\lambda = 4$ and $\mathbf{x}_i^\top = (1, x_i)$. The values x_i, $i = 1, \ldots, 400$, were generated independently from a uniform distribution in the interval $(2, 5)$ and those values were kept constant throughout the experiment. The degree of freedom ($\boldsymbol{\nu}$) for the different cases of SMSN-CR models was fixed: $\nu = 3$ for the ST-CR and SSL-CR models and $(\nu, \gamma) = (0.1, 0.1)$ for the SCN-CR distribution. We generated 500 samples of size $n = 400$ from the SMSN-CR model, considering five censoring proportions, $p = \{0\%,\ 10\%,\ 20\%,\ 30\%,\ \text{and } 40\%\}$. To guarantee the presence of a perturbed observation, we chose case # 200 and we replaced its parameters $\boldsymbol{\beta}$ by $\{2.5\boldsymbol{\beta},\ 5\boldsymbol{\beta},\ 7.5\boldsymbol{\beta},\ 10\boldsymbol{\beta}\}$. Considering the criteria $M(0)_i > \overline{M(0)} + 3SM(0)$ and GD_i^1 for $i = 1, \ldots, 400$, to decide which point is influential or not.

Table 3.4 shows, in percentage, the number of times that the measure correctly identify the observation # 200 as the most influential. As expected, the percentage of correctly detecting atypical observations increases for increasing perturbation rates and we observe that in general, there is high sensitivity of the estimates in the presence of atypical data when the SN-CR model is considered.

Table 3.4: Simulation study. % of correctly identified influential observation using GD_i^1 and under case-weight perturbation.

Influence Measures	Perturbation	SN-CR Cens. Level 0%	10%	20%	30%	40%	ST-CR Cens. Level 0%	10%	20%	30%	40%
GD_i	2.5β	19.8	15.8	15.8	14.8	14.0	4.6	5.6	4.0	3.4	0.8
	5β	64.2	63.2	60.8	64.8	60.6	39.2	35.4	36.4	30.8	20.8
	7.5β	89.8	86.0	86.2	87.8	85.6	70.6	68.6	70.2	62.0	62.2
	10β	96.4	98.6	96.6	98.4	96.6	84.4	83.6	80.8	87.2	82.4
Case-Weight perturbation	2.5β	15.0	14.4	16.0	17.0	16.0	14.0	18.6	19.6	19.6	20.4
	5β	62.0	62.6	61.2	65.6	63.8	50.0	56.0	56.4	59.6	56.2
	7.5β	88.4	85.8	87.4	88.0	87.6	77.8	78.8	79.0	80.4	83.2
	10β	96.2	98.6	96.8	98.4	97.0	88.8	88.4	88.0	92.6	90.2

Influence Measures	Perturbation	SCN-CR Cens. Level 0%	10%	20%	30%	40%	SSL-CR Cens. Level 0%	10%	20%	30%	40%
GD_i	2.5β	15.4	11.0	11.0	9.0	4.8	14.2	12.0	13.0	7.2	7.2
	5β	36.2	34.4	37.2	40.2	38.0	53.6	50.0	53.4	46.4	46.8
	7.5β	54.6	44.4	39.6	31.8	42.2	79.8	78.6	80.2	77.6	73.2
	10β	80.2	73.4	63.2	16.8	35.8	93.0	94.4	100	91.8	91.8
Case-Weight perturbation	2.5β	14.4	12.2	15.0	13.0	9.6	11.8	13.0	15.2	12.6	10.8
	5β	34.4	40.6	47.2	48.8	44.4	52.0	50.6	55.6	52.6	50.4
	7.5β	52.8	52.4	55.8	48.6	53.0	78.8	79.2	81.8	79.0	76.4
	10β	78.6	77.4	70.8	26.6	48.0	92.6	94.4	100	92.8	92.6

3.7 Conclusion

In this chapter we presented a diagnostic analysis of linear regression models with censored responses and observational errors following a distribution belonging the class of SMSN distributions. The diagnostic analysis was based on the case-deletion and local influence techniques suggested by Zhu et al. (2001) and Zhu and Lee (2001), respectively, which are the counterparts for missing data models of the well-known ones proposed by Cook (1977) and Cook (1986). The structure of the complete-data likelihood function, obtained considering as if the missing data were in fact observed, is an essential element of the theory. Its simple form allows obtaining a tractable expression for the Q-function, which is essentially what is needed to provide an approximation of the maximum likelihood estimate of the parameters when an observation is excluded (for the case-deletion method). The same is true for the local influence method, in the case of the normal curvature expressions. Using the developed method, we analyzed a real dataset and carried out extensive simulation studies. We observe, through influence diagnostic procedures, that when we used distributions with heavier tails than the SN-CR model, some aspects of robustness of the SAEM estimators under heavy-tail SMSN distributions were noted.

Chapter 4

Concluding remarks

In this work we developed a full likelihood approach for linear regression models with censored responses based on scale mixtures of skew-normal distributions, denoted by SMSN-CR. By exploring statistical properties of the SMSN class, we discuss in detail the implementation of the SAEM algorithm for maximum likelihood estimation with the likelihood function, predictions of unobservable values of the response and the asymptotic standard errors as a byproducts. Next, influence techniques, such as case-deletion and local influence, are developed to show the robust aspect of the SMSN-CR models against outlying and influential observations.

In order to examine the performance of our proposed methods, we present various simulation studies and the methods are illustrated by analysis of two real dataset. The methods developed have been implemented in software `R` and are the codes are available upon request.

4.1 Future research

The are a number of possible extensions of the current work. For example, censored nonlinear regression models using SMSN distributions (SMSN-NLCR) as considered by Garay et al. (2011). The proposed methods can be extended to multivariate settings, such as the recent proposals of Matos et al. (2015) for censored mixed-effects models and Garay et al. (2015) for irregularly observed longitudinal data using multivariate SMSN distributions (Lachos et al., 2010). Due to the popularity of Markov chain Monte Carlo techniques, another potential work is to pursue a fully Bayesian treatment in this context for producing posterior inference. The method can also be extended to finite mixtures of regressions with skewed and heavy-tailed censored responses based on recent approaches by Caudill (2012) and Karlsson and Laitila (2014).

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Cont. 19*, 716–723.

Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B. 36*, 99–102.

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

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

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

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

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

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

Branco, M. D. and D. K. Dey (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis 79*, 99–113.

Cabral, C. R. B., V. H. Lachos, and M. R. Madruga (2012). Bayesian analysis of skew-normal independent linear mixed models with heterogeneity in the random-effects population. *Journal of Statistical Planning and Inference 142*, 181–200.

Cancho, V. G., D. K. Dey, V. H. Lachos, and M. G. Andrade (2011). Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. *Computational Statistics & Data Analysis 55*, 588–602.

Caudill, S. B. (2012). A partially adaptive estimator for the censored regression model based on a mixture of normal distributions. *Statistical Methods & Applications 21*(2), 121–137.

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

Couvreur, C. (1996). The EM algorithm: a guided tour. In *Proceedings of the 2d IIEE European Workshop on Computationaly Intensive Methos in Control and Signal Processing*.

Delyon, B., M. Lavielle, and E. Moulines (1999). Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics 27*(1), 94–128.

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

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

Galarza, C. E., D. Bandyopadhyay, and V. H. Lachos (2015). Quantile regression for linear mixed models: A stochastic approximation em approach. Technical Report 2, Universidade Estadual de Campinas.

Garay, A. M., L. M. Castro, J. Leskow, and V. H. Lachos (2015). Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Statistical Methods in Medical Research*, DOI: 10.1177/0962280214551191.

Garay, A. M., V. H. Lachos, and C. A. Abanto-Valle (2011). Nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of the Korean Statistical Society 40*, 115–124.

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

Greene, W. H. (2012). *Econometric analysis.* Prentice Hall, New York.

Ibacache-Pulgar, G. and G. A. Paula (2011). Local influence for Student-t partially linear models. *Computational Statistics & Data Analysis 55*, 1462–1478.

Jank, W. (2006). Implementing and diagnosing the stochastic approximation em algorithm. *Journal of Computational and Graphical Statistics 15*, 803–829.

Karlsson, M. and T. Laitila (2014). Finite mixture modeling of censored regression models. *Statistical papers 55*(3), 627–642.

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

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

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

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

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

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

Lesaffre, E. and G. Verbeke (1998). Local influence in linear mixed models. *Biometrics 54*, 570–582.

Lin, T. I. (2010). Robust mixture modeling using multivariate skew-t distributions. *Statistics and Computing 20*(3), 343–356.

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society, Series B 44*, 226–233.

Massuia, M. B., C. R. B. Cabral, L. A. Matos, and V. H. Lachos (2014). Influence diagnostics for Student-t censored linear regression models. *Statistics doi:10.1080/02331888.2014.958489*.

Massuia, M. B., A. M. Garay, V. H. Lachos, and C. R. B. Cabral (2015). Bayesian analysis of censored linear regression models with scale mixtures of skew-normal distributions. Technical Report 3, Universidade Estadual de Campinas.

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

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

Meilijson, I. (1989). A fast improvement to the em algorithm to its own terms. *J. R. Stat. Soc. Ser. B 51*, 127–138.

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

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

Nelson, F. D. (1977). Censored regression models with unobserved, stochastic censoring thresholds. *Journal of Econometrics 6*, 309–327.

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

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

Park, J. W., M. G. Genton, and S. K. Ghosh (2007). Censored time series analysis with autoregressive moving average models. *Canadian Journal of Statistics 35*, 151–168.

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

R Development Core Team (2015). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rubin, D. B. (1987). The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The sir algorithm. *Journal of the American Statistical Association*, 543–546.

Rubin, D. B. et al. (1988). Using the sir algorithm to simulate posterior distributions. *Bayesian statistics 3*(1), 395–402.

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

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

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

Thompson, M. L. and K. P. Nelson (2003). Linear regression with Type I interval and left-censored response data. *Environmental and Ecological Statistics 10*, 221–230.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society 26*, 24–36.

Vaida, F. (2005). Parameter convergence for EM and MM algorithms. *Statistica Sinica 15*(3), 831–840.

Vaida, F. and L. Liu (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics 18*, 797–817.

Wei, G. C. and M. A. Tanner (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association 85*(411), 699–704.

Wu, C. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics 11*(1), 95–103.

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

Ying-Zi, F., T. Nian-Sheng, and C. Xing (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., F. V. Labra, V. H. Lachos, and N. Balakrishnan (2010). Influence analyses of skew-normal/independent linear mixed models. *Computational Statistics & Data Analysis 54*, 1266–1280.

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

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

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

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

Appendix A

Additional results of Chapter 2

A.1 Complementary results of the simulation study 1

In this Section, we present the results of the simulation study 1 for different levels of censoring: $p = 0\%,\ 20\%$ and 35%.

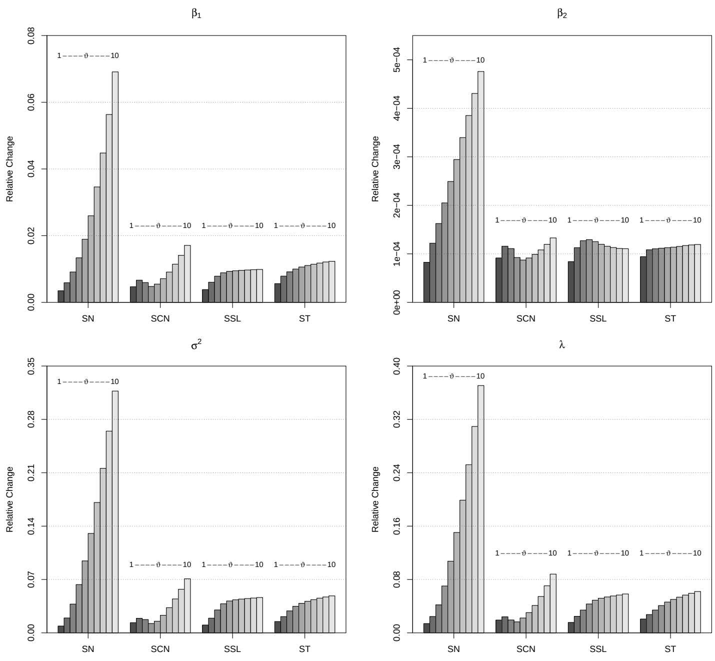

Figure A.1: Simulation study 1. Average relative changes on estimates for different perturbations ϑ and censoring level $p = 0\%$.

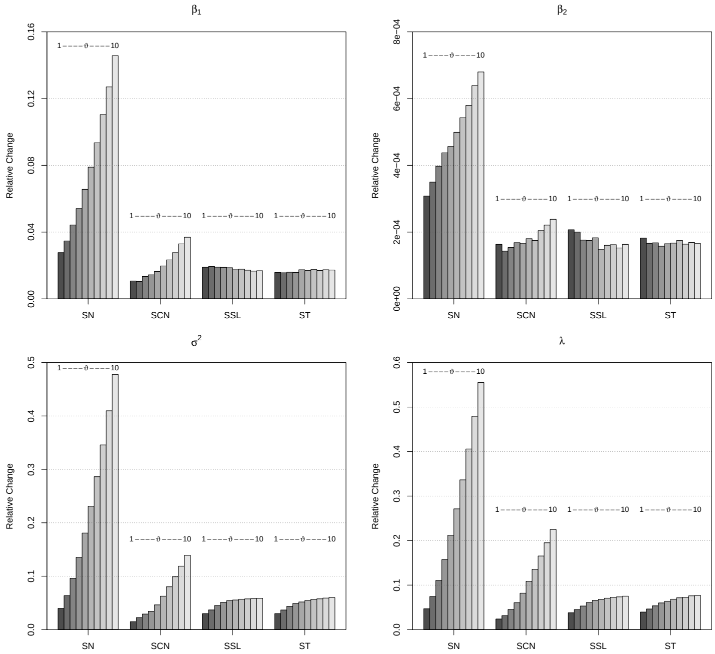

Figure A.2: Simulation study 1. Average relative changes on estimates for different perturbations ϑ and censoring level $p = 20\%$.

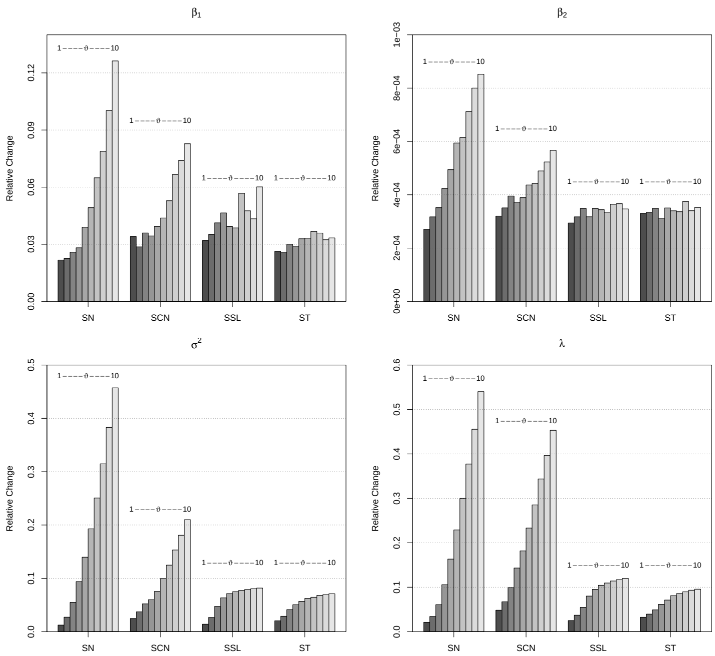

Figure A.3: Simulation study 1. Average relative changes on estimates for different perturbations ϑ and censoring level $p = 35\%$.

A.2 Complementary results of the simulation study 2

Here we show the Bias and MSE of parameters $\boldsymbol{\theta}$, for the levels of censoring $p = 20\%$ and 35%, respectively.

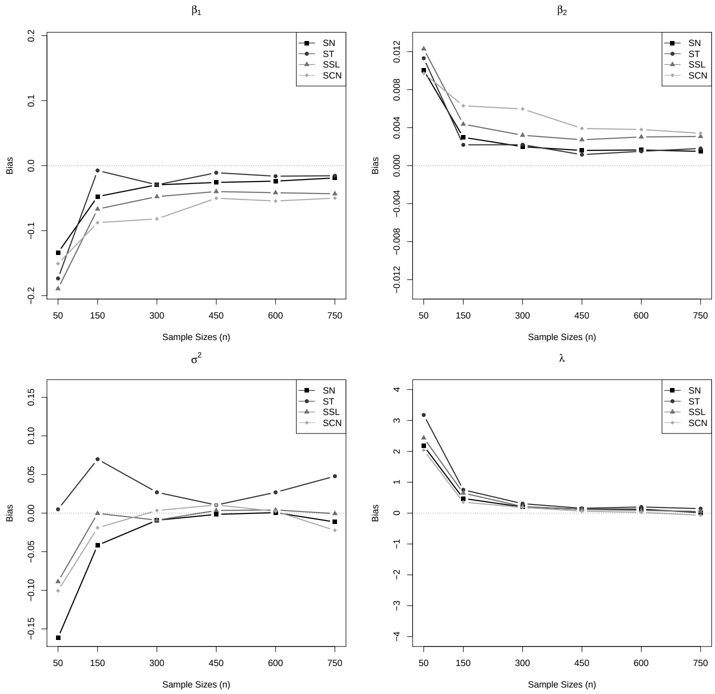

Figure A.4: Simulation study 2. Bias of parameters β_1, β_2, σ^2 and λ for SMSN-models with level of censoring $p = 20\%$.

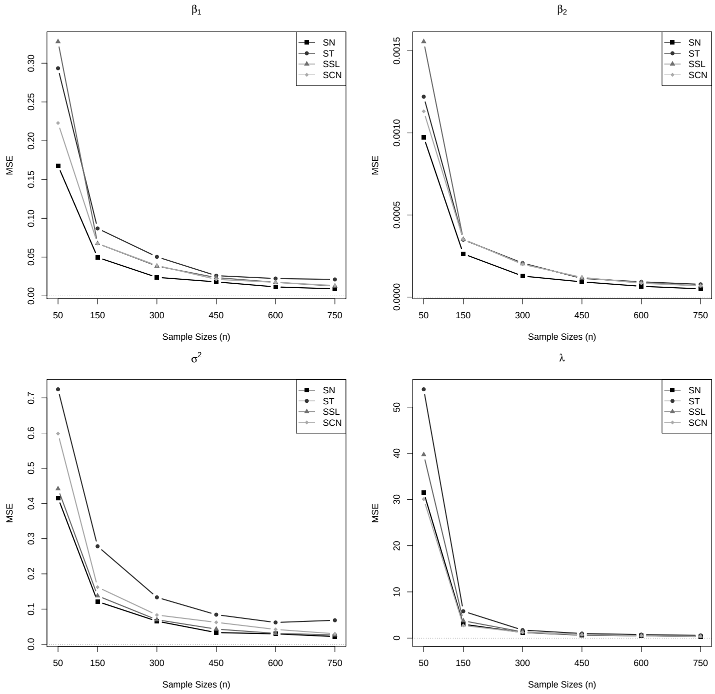

Figure A.5: Simulation study 2. MSE of parameters β_1, β_2, σ^2 and λ for SMSN-models with level of censoring $p = 20\%$.

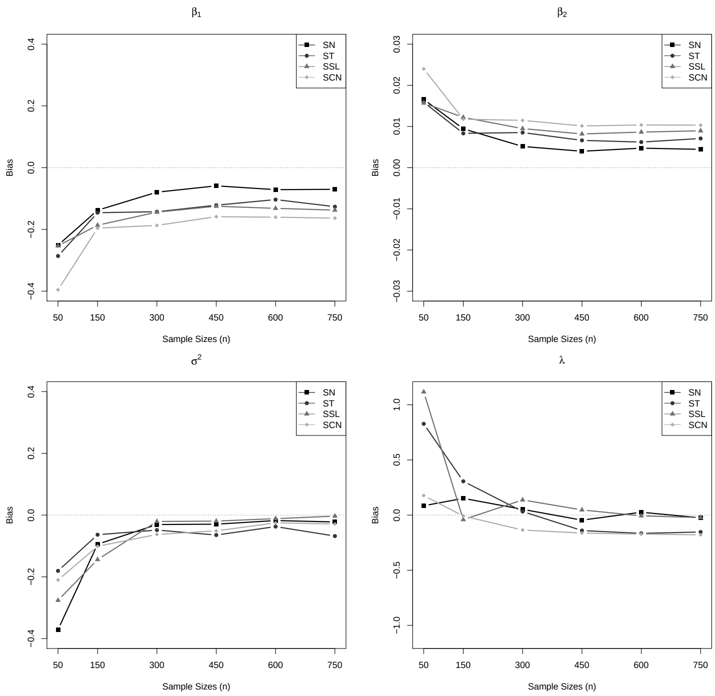

Figure A.6: Simulation study 2. Bias of parameters β_1, β_2, σ^2 and λ for SMSN-models with level of censoring $p = 35\%$.

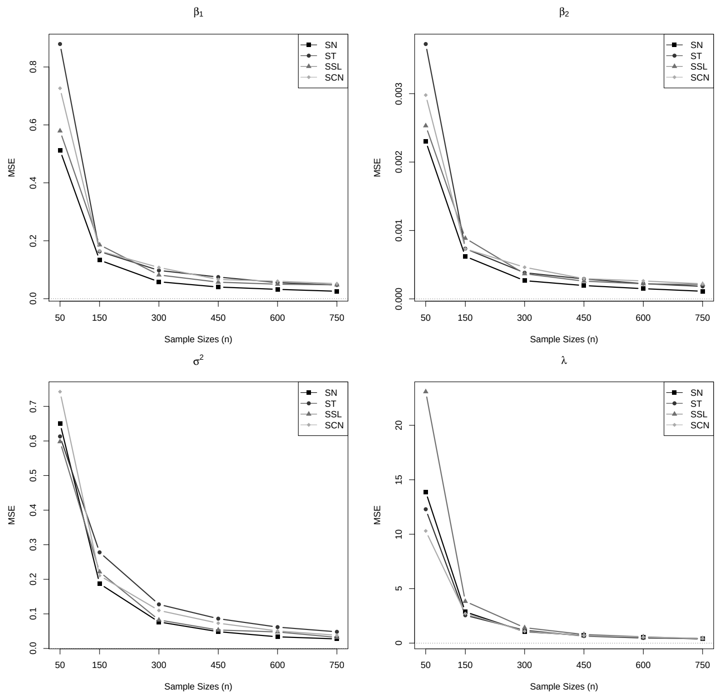

Figure A.7: Simulation study 2. MSE of parameters β_1, β_2, σ^2 and λ for SMSN-models with level of censoring $p = 35\%$.

A.3 *Wage rate* dataset under SMN-CR models

In this Section, we present the comparison between the SMN-CR models, considering the *wage rate* dataset.

Table A.1: Wage rate dataset. Values of some model selection criteria for SMN-CR models

Criteria	N-CR	T-CR	CN-CR	SL-CR
log-likelihood	-1481.6550	-1440.1450	-1432.0850	-1436.2860
AIC	2975.3110	2894.2910	2880.1710	2886.5730
BIC	3003.0550	2926.6590	2917.1630	2918.9410
EDC	2996.2400	2918.7080	2908.0760	2910.9900

A.4 Complementary results of the application

In this Section, we describe the summary of convergence for the parameters, $\boldsymbol{\beta}$, σ^2, λ, $\boldsymbol{\nu}$, for the SMSN-CR models. The vertical dashed line delimits the beginning of the almost sure convergence, as defined by the cutoff point, c.

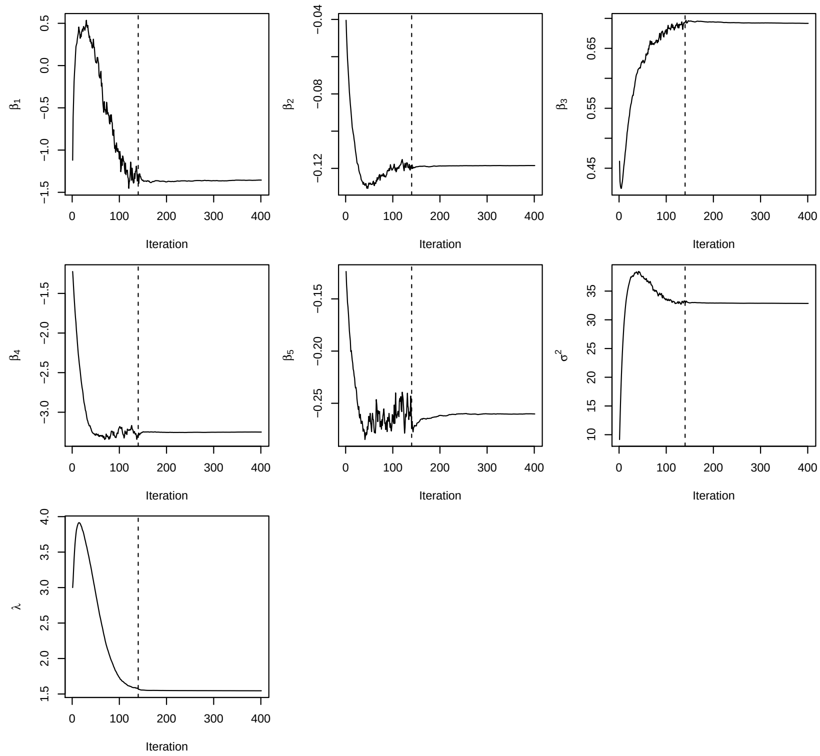

Figure A.8: Wage rate dataset. Graphical summary of convergence for the parameters from SN-CR model, $m = 20$, $c = 0.35$ and $S = 400$.

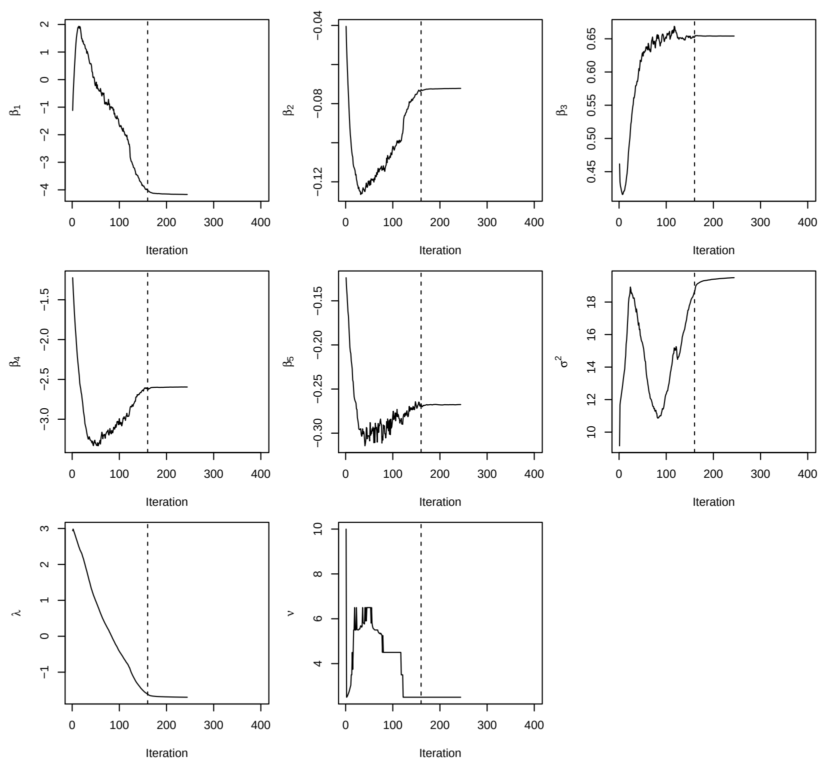

Figure A.9: Wage rate dataset. Graphical summary of convergence for the parameters from ST-CR model, $m = 20$, $c = 0.40$ and $S = 400$.

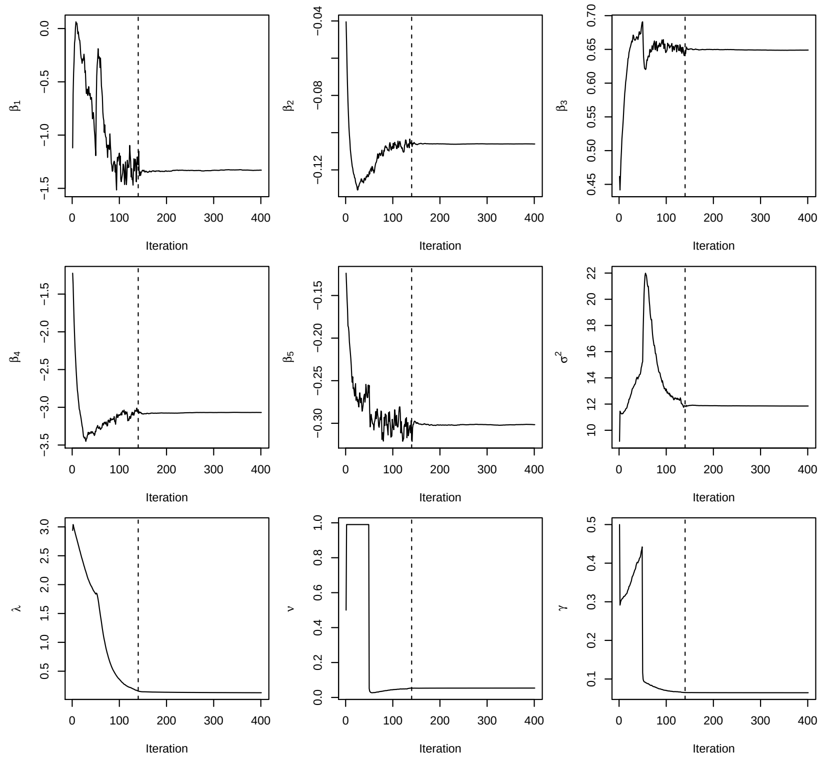

Figure A.10: Wage rate dataset. Graphical summary of convergence for the parameters from SCN-CR model, $m = 20$, $c = 0.35$ and $S = 400$.

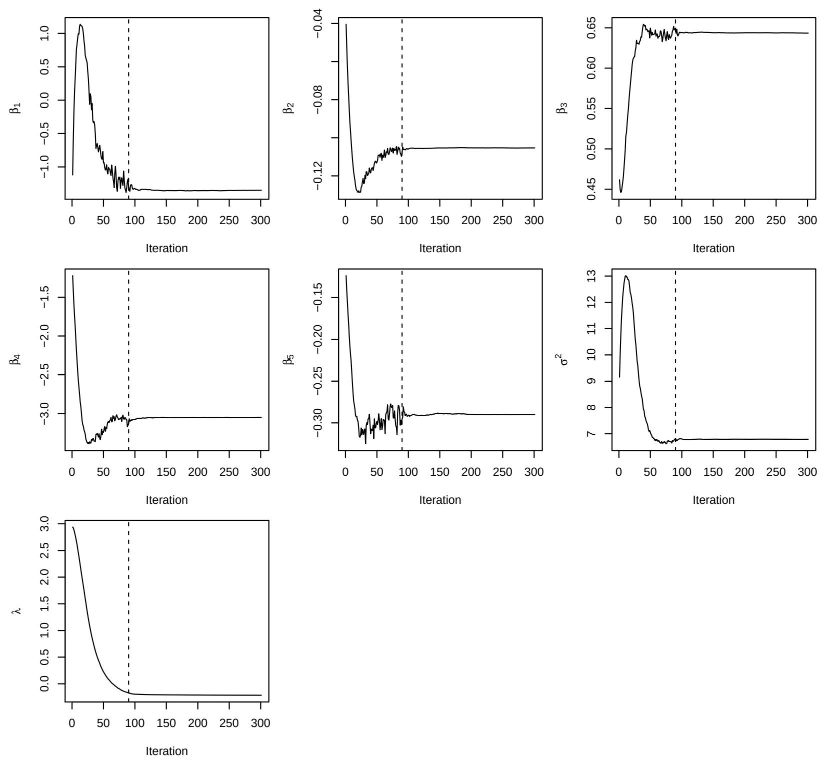

Figure A.11: Wage rate dataset. Graphical summary of convergence for the parameters from SSL-CR model, $m = 20$, $c = 0.30$ and $S = 300$.