



MATHEMATICAL SCIENCES

The Asymmetric Power-Student-t Model for Censored and Truncated Data

ROGER TOVAR-FALÓN, HELENO BOLFARINE & GUILLERMO MARTÍNEZ-FLÓREZ

Abstract: In this paper, we propose the power Student-t regression model for censored (limited) observations which extends the Student-t censored regression model. This extension is based on the asymmetric and heavy-tailed power Student-t distribution. The score functions and expected information matrix are given as well as the process for estimating the parameters in the model is discussed by using the likelihood approach. Two simulation studies are conducted to evaluate parameter recovery and properties of the model and finally, two applications to a real data set are reported to demonstrate the usefulness of this new methodology.

Key words: Censored regression model, Fisher information matrix, maximum likelihood estimation, power Student-t distribution.

INTRODUCTION

Regression models where the response variable is censored or limited are common in different fields: clinical essays, econometric analysis, social phenomena, engineering studies, among others. In clinical essays for example, in the first phases of development of the new vaccines, the determination of antibody concentration values often are left-censored due to detection limit by lack of sensitivity of the essay when the concentrations are near zero, see Moulton & Halsey (1995). In social phenomena, the study on extramarital behavior where the variable of interest is the number of extramarital affairs in the previous year, for example, it can result in a left-censored variable (Fair 1978). In econometrics analysis, the ordinary Tobit model (Tobin 1958) is commonly used to conduct studies of the labor force participation of married women. In this case, the observed response is the wage rate, which is typically considered as censored below zero, i.e., for working women, positive values for the wage rates are registered, whereas for the non-working women the observed wage rates are zero; see Mroz (1987).

In situations such as previously discussed, where censored regression (CR) models are proposed, it is common to assume a normal distribution for the error term, however, this assumption can not be suitable and it can be unrealistic due to the presence of atypical observations or high (or low) degree of skewness and kurtosis of the response variable, which the normal model is unable to capture, so considerable interest has centered on relaxing the assumption of normality of the errors in CR models. In this context, some authors have proposed a wide range of alternatives to the normal censored regression (NCR) model which is widely known in the literature as the Tobit model. Arellano-Valle et al.

(2012) for example, extend the classical Tobit model by introducing the Student- t censored regression (TCR) model that can be suitable when the response variable has heavy-tails and the kurtosis is greater than the usual normal model. Another extension of the Tobit model was proposed by Martínez-Flórez et al. (2013) by considering that random errors follow a power-normal (PN) distribution (Gupta & Gupta 2008). The novelty of this proposal is the incorporation of a shape parameter which gives flexibility to the assumption of the symmetric errors (normality assumption) and it allows to accommodating skewed forms to the left and the right for the error term in CR models. Recently Garay et al. (2017) proposed a family of censored regression models based on the family of symmetric distributions commonly known as the scale mixture of normal (SMN) distributions, which includes the TCR model proposed by Arellano-Valle et al. (2012). This family also includes Pearson type VII (PVII), slash (SL), power exponential (PE), contaminated normal (CN) and normal (N) distributions. In addition to being robust, these models have shown to be useful in detecting atypical observations in CR models.

Although some proposals that take into account the problem of atypical observations in censored regression models, most of them are based on the assumption of symmetry of the error and few studies that capture departure from symmetry in the distribution of errors as in Martínez-Flórez et al. (2013), for example, who support their work in the great virtues of the alpha-power models to fit data where distribution presents high or low asymmetry and/or kurtosis.

Within this class of alpha-power models Zhao & Kim (2016) proposed an extension of the Student- t model by defining the power-Student- t (PT) distribution as an alternative to the skew- t model by Azzalini & Capitanio (2003) for fitting skewed and heavy-tailed data. The PT model, which extends the power-normal model by Gupta & Gupta (2008) seems to be useful in situations where the data present higher degree of skewness and kurtosis than PN model in presence of atypical observations.

In this paper, we propose a censored regression model under the assumption that errors follow a PT distribution (hereafter we will call it the PTCR model). The assumption of PT distribution gives flexibility for accommodating skew forms to the left and the right, and kurtosis greater or smaller than the Student t -distribution can be also accommodated, hence, PTCR model extends the TCR model. The process of inference in the model is conducted by using the maximum likelihood (ML) approach and its large sample properties. Application is implemented to real data set where it is demonstrated that the proposed model can be very useful in fitting real data sets.

The rest of this paper is organized as follows: Section “The Power Student- t Distribution” presents a brief review of the main properties of the PT distribution. In Section “Power-Student- t Model for Censored and Truncated Data”, we introduce the censored and truncated PT models. Section “Censored Power-Student- t Regression Model” introduces the PTCR model. Here, ML equation and the observed and expected information matrices are given. Section “Simulation Study” presents the results of a simulation study which reveals the good performance of the estimation approach. The PTCR model is fitted to a data set of housewives wages in Section “Real Data Application”, revealing that the data set in question can be fitted by PTCR as well as by a CR model where the observational errors have a SMN distribution (SMNCR model).

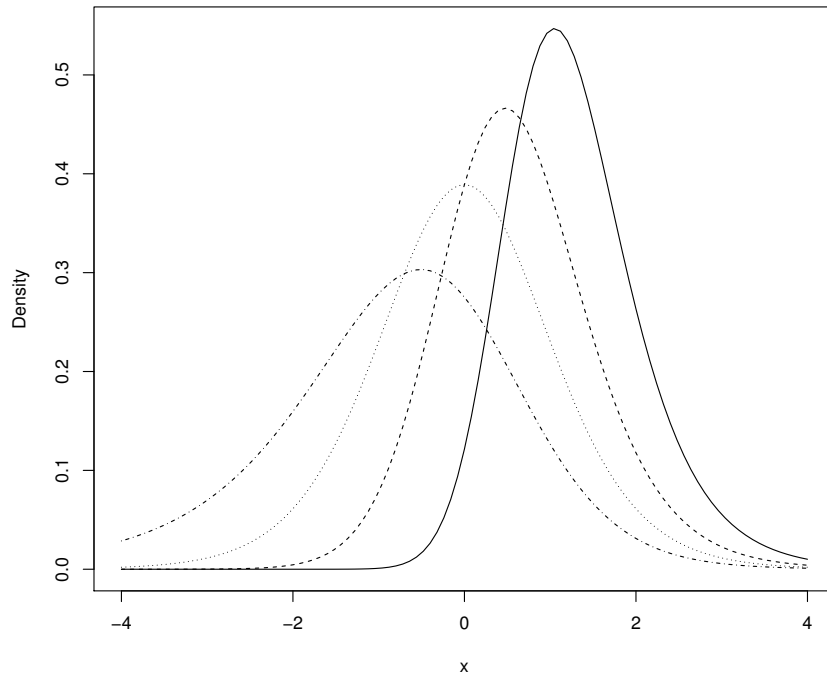


Figure 1. Density function of $PT(\alpha, 10)$ for $\alpha = 5$ (solid line), $\alpha = 2$ (dashed line), $\alpha = 1$ (dotted line), $\alpha = 0.5$ (dotted-dashed line).

THE POWER STUDENT-T DISTRIBUTION

In this section, we present the PT distribution and review some of its main characteristics and properties. The PT model was introduced by Zhao & Kim (2016) and it is an alternative to skew-t model for fitting data with high indices of asymmetry and kurtosis in addition to heavy tails.

Definition 1. The random variable X is said to have a PT distribution with parameter α and degree of freedom ν , if X has probability density function (PDF) given by

$$f_{PT}(x; \alpha, \nu) = \alpha f_T(x; \nu) [F_T(x; \nu)]^{\alpha-1}, \tag{1}$$

for $x \in \mathbb{R}$ and $\alpha > 0$. Functions $f_T(\cdot; \nu)$ and $F_T(\cdot; \nu)$ are the PDF and cumulative distribution function (CDF) of the standard Student-t distribution.

Random variable having $f_{PT}(x; \alpha, \nu)$ distribution is denoted shortly by $X \sim PT(\alpha, \nu)$. Figure 1 displays some forms of the PDF of the PN distribution for selected values of α . Note from figure that parameter α controls the skewness and kurtosis of the distribution. The CDF of the PT model is given by

$$F_{PT}(x; \alpha, \nu) = [F_T(x; \nu)]^\alpha, \tag{2}$$

for $x \in \mathbb{R}$. Some properties of the PT distribution can be proven as result of Definition 1.

Proposition 1. Let $X \sim PT(\alpha, \nu)$, then

- (i) if $\alpha = 1$, X follows Student-t distribution and we write $X \sim T(\nu)$,
- (ii) if $\nu \rightarrow \infty$, X converges to power-normal (PN) model with parameter α . The PDF is given by

$$f_{PN}(x; \alpha) = \alpha \varphi(x) [\Phi(x)]^{\alpha-1}, \quad x \in \mathbb{R}. \tag{3}$$

More details of PN distribution can be found in Gupta & Gupta (2008) and Pewsey et al. (2012).

- (iii) if $\alpha = 1$ and $\nu \rightarrow \infty$, X converges to standard normal distribution.

Proof. Proof of (i)-(iii) are directly obtained from definition of PT distribution □

Proposition 2. Let $X \sim PT(\alpha, \nu)$, then for $k \in \mathbb{N}$

$$E[X^k] = E[(F_T^{-1}(Y; \nu))^k], \tag{4}$$

where Y has a beta distribution, and $F_T^{-1}(\cdot; \nu)$ denotes the inverse of the function $F_T(\cdot; \nu)$.

Proof. We have by definition that

$$E[X^k] = \int_{\mathbb{R}} x^k \alpha f_T(x; \nu) [F_T(x; \nu)]^{\alpha-1} dx$$

thus, letting $y = F_T(x; \nu)$, then $x = F_T^{-1}(y; \nu)$, it follows that

$$E[X^k] = \int_0^1 (F_T^{-1}(y; \nu))^k \alpha y^{\alpha-1} dy$$

which is the expected value of the function $(F_T^{-1}(Y; \nu))^k$, where Y follows a beta distribution with parameter α and 1. □

The expected value, variance, indices of asymmetry and kurtosis of the PT model can be found by using the expressions

(i) $E[X] = \mu_1$

(ii) $V[X] = \mu_2 - \mu_1^2$

(iii) $\gamma_1 = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}}$

(iv) $\gamma_2 = \frac{\mu'_4}{(\mu'_2)^2}$

where $\mu_r = E[X^r]$ and $\mu'_r = E[X - E(X)]^r$. Table I presents the values of the asymmetry coefficient of the PT model for some values of the ν parameter and for values of α in the range of 0.1 to 100000.

Definition 2. Let $X \sim PT(\alpha, \nu)$. The PT density of location and scale is defined as the distribution of $Z = \xi + \eta X$, for $\xi \in \mathbb{R}$ and $\eta > 0$. The corresponding PDF is given by

$$f_{PT}(z; \xi, \eta, \alpha, \nu) = \alpha f_T\left(\frac{z - \xi}{\eta}; \nu\right) \left[F_T\left(\frac{z - \xi}{\eta}; \nu\right)\right]^{\alpha-1}, \tag{5}$$

for $z \in \mathbb{R}$. We denote this extension as $Z \sim PT(\xi, \eta, \alpha, \nu)$, and we have that $PT(0, 1, \alpha, \nu) \equiv PT(\alpha, \nu)$.

Table I. Skewness of the $PT(\alpha, \nu)$ model. Values of α ranging of 0.1 to 100000.

ν	min	max
4	-9.3423	5.5913
5	-7.5147	3.5101
6	-7.3825	2.7668
7	-4.6073	2.3721
8	-4.1468	2.1228
9	-3.8410	1.9490
10	-3.4826	1.8202
12	-2.2654	1.6409
15	-2.0131	1.4752
20	-1.8802	1.3208
25	-1.8245	1.2329
30	-1.6372	1.1760
50	-1.1158	1.0664
100	-1.0401	0.9864
500	-0.9390	0.9264

The k th moment of the random variable Z is given by

$$E[Z^k] = \sum_{j=0}^k \binom{k}{j} \xi^j \eta^{k-j} \mu_{k-j}, \tag{6}$$

where μ_j is the j th moment of a random variable $X \sim PT(\alpha, \nu)$. Zhao & Kim (2016) derived the information matrix for the location-scale version and showed that it is non-singular when $\alpha = 1$ for small values of the parameter ν (i.e., $\nu \ll 30$). When ν tends to $+\infty$, then PT distribution converges to PN model and here, we recall Pewsey et al. (2012) showed that PN model has non-singular information matrix. This result guarantees that regularity conditions are satisfied for the likelihood approach, hence, with PT model, symmetry can be tested by using ordinary large sample properties of the likelihood ratio statistics.

POWER-STUDENT-T MODEL FOR CENSORED AND TRUNCATED DATA

Based on the goodness of the PT distribution to fit data with high indices of asymmetry and kurtosis, in this section, we introduce the censored PT and the truncated PT models which we will be denoted by CPT and TPT , respectively.

Definition 3 (Censored PT Model). Suppose that random variable Y follows a $PT(\xi, \eta, \alpha, \nu)$ distribution. Let Y_1, \dots, Y_n a random sample of size n of Y , where only those values of Y_i greater than constant k ;

are recorded; and for values $Y_i \leq k_i$ only the value k_i is registered. The observed values Y_i^o can be written as

$$Y_i^o = \begin{cases} k_i, & \text{if } Y_i \leq k_i, \\ Y_i, & \text{if } Y_i > k_i, \end{cases}$$

for $i = 1, \dots, n$. The resulting sample is said to be a censored power-Student-t (CPT).

From Definition 3 it follows that $P(Y_i^o = k_i) = P(Y_i \leq k_i) = \left\{F_T\left(\frac{k_i - \xi}{\eta}; \nu\right)\right\}^\alpha$, and for the observations $Y_i^o = Y_i$ the distribution of Y_i^o is the same of Y_i , i.e., $Y_i^o \sim \text{PT}(\xi, \eta, \alpha, \nu)$. For convenience, we choose to work with the case of left-censored data, however, the followings results can be extended to other types of censorship.

Maximum Likelihood Estimation for CPT Model

Let Y_1^o, \dots, Y_n^o be a random sample of the censored $\text{PT}(\xi, \eta, \alpha, \nu)$ distribution (censored in k_i). To perform statistical inference for parameter vector $\vartheta = (\xi, \eta, \alpha, \nu)^\top$ by using the ML method, we use the reparameterization in Olsen (1978). Thus, let $\gamma = \sigma\xi$ and $\varsigma = 1/\eta$, the log-likelihood function for the new vector $\varphi = (\gamma, \sigma, \alpha, \nu)^\top$, given the observed sample \mathbf{Y}^o can be written as

$$\begin{aligned} \ell(\varphi; \mathbf{Y}^o) &= \alpha \sum_{i=1}^n (1 - d_i) \log\{F_T(c_i; \nu)\} \\ &\quad + \sum_{i=1}^n d_i \{\log \alpha + \log \sigma + \log f_T(z_i; \nu) + (\alpha - 1) \log F_T(z_i; \nu)\} \end{aligned} \quad (7)$$

where $c_i = \varsigma k_i - \gamma$, $z_i = \varsigma y_i - \gamma$, and d_i is an indicator variable defined as $d_i = 1$ if $Y_i^o = Y_i$; and $d_i = 0$ if $Y_i^o = k_i$. The components of the score function $\mathbf{U}(\varphi)$ are obtained by deriving partially the

log-likelihood function given in (7) with regard to components γ , ς , α and ν , we obtain the following equations

$$\begin{aligned}
 U(\gamma) &= -\alpha \sum_{i=1}^n (1 - d_i) r(c_i; \nu) \\
 &\quad + \sum_{i=1}^n d_i \left\{ \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} z_i - (\alpha - 1) r(z_i; \nu) \right\} \\
 U(\sigma) &= \alpha \sum_{i=1}^n (1 - d_i) r(c_i; \nu) k_i \\
 &\quad + \sum_{i=1}^n d_i \left\{ \frac{1}{\sigma} - \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} z_i y_i + (\alpha - 1) r(z_i; \nu) y_i \right\} \\
 U(\alpha) &= \sum_{i=1}^n (1 - d_i) \log\{F_T(c_i; \nu)\} + \sum_{i=1}^n d_i \left\{ \frac{1}{\alpha} + \log F_T(z_i; \nu) \right\} \\
 U(\nu) &= \frac{\alpha}{2} \sum_{i=1}^n (1 - d_i) \left\{ \psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - b_{01}(c_i; \nu) - \frac{c_i r(c_i; \nu)}{\nu} \right\} \\
 &\quad + \frac{1}{2} \sum_{i=1}^n d_i \left\{ \psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - \frac{1}{\nu} \right. \\
 &\quad \left. - \log \left(1 + \frac{z_i^2}{\nu} \right) + \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} \frac{z_i^2}{\nu} \right\} \\
 &\quad + \frac{\alpha - 1}{2} \sum_{i=1}^n d_i \left\{ \psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - b_{01}(z_i; \nu) - \frac{z_i r(z_i; \nu)}{\nu} \right\}
 \end{aligned}$$

where $r(x; \nu) = f_T(x; \nu)/F_T(x; \nu)$, $\psi(x) = \frac{d}{dx} \log \Gamma(x)$ is the digamma function and $b_{mn}(c_m; \nu + m)$ is the truncated moment defined as

$$b_{mn}(c_m; \nu + m) = \int_{-\infty}^{c_m} s^m \left\{ \log \left(1 + \frac{s^2}{\nu + m} \right) \right\}^n \frac{f_T(s; \nu + m)}{F_T(c_m; \nu + m)} ds, \tag{8}$$

with $c_m = \sqrt{\frac{\nu+m}{\nu}} c$ and $c_0 = c$. The moments b_{mn} in (8) are obtained by numerical integration, for example, by using *integrate* function of R Development Core Team (2018). The ML estimates of the parameters ξ , η , α and ν in censored PT model are obtained using iterative algorithms based on the score functions and by applying the inverse transformation $\xi = \gamma/\varsigma$ and $\eta = 1/\varsigma$. For obtaining the standard errors of the ML estimates one should compute the information matrix \mathbf{I}_φ . It is well known that the elements of \mathbf{I}_φ are given by

$$\mathbf{I}_\varphi(i, k) = -E \left[\frac{\partial^2 \ell(\varphi; \mathbf{Y}^o)}{\partial \varphi_i \partial \varphi_k} \right], \quad i, k = 1, \dots, 4$$

where $\varphi = (\gamma, \sigma, \alpha, \nu)^\top$. Since expectation over PT distribution and second-order derivatives are not straightforward, numerical methods should be performed to obtain the explicit form of the

information matrix \mathbf{I}_φ . Thus, we use the observed information matrix for calculating the standard errors in the rest of the paper. To recover the information matrix \mathbf{I}_ϑ of the original parameterization $\vartheta = (\xi, \eta, \alpha, \nu)^\top$, we use

$$\mathbf{I}_\vartheta = (\partial\varphi/\partial\vartheta)^\top \mathbf{I}_\varphi (\partial\varphi/\partial\vartheta),$$

where the Jacobian matrix is

$$\frac{\partial\varphi}{\partial\vartheta} = \begin{pmatrix} \frac{1}{\eta} & -\frac{1}{\eta^2} & 0 & 0 \\ 0 & -\frac{1}{\eta^2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{9}$$

Definition 4 (Truncated PT Model). Let X be a random variable with distribution $X \sim \text{PT}(\xi, \eta, \alpha, \nu)$. Let $a, b \in \mathbb{R}$ with $a < b$, such that $P(a < X < b) > 0$. It is said that random variable Y has a truncated power-Student- t (TPT) distribution in the interval (a, b) , if Y has the same distribution as $X \mid X \in (a, b)$. In this case, we write $Y \sim \text{TPT}_{(a,b)}(\xi, \eta, \alpha, \nu)$.

As a consequence of the Definition 4, the PDF of TPT distribution can be obtained as

$$f_{PT}(y \mid (a, b)) = \frac{\alpha}{\eta} f_T\left(\frac{y-\xi}{\eta}; \nu\right) \left[F_T\left(\frac{y-\xi}{\eta}; \nu\right)\right]^{\alpha-1} \times \left\{ \left[F_T\left(\frac{b-\xi}{\eta}\right)\right]^\alpha - \left[F_T\left(\frac{a-\xi}{\eta}\right)\right]^\alpha \right\}^{-1}$$

if $a < y < b$, and $f_{PT}(y \mid (a, b)) = 0$ in otherwise. Now, we consider that before the sample to be selected, the distribution of Y is truncated at the value k , so that we can only choose observations such that $Y \leq k$. Then, random variable Y has PDF given by

$$f_{PT}(y \mid y \leq k) = \frac{\alpha}{\eta} f_T\left(\frac{y-\xi}{\eta}; \nu\right) \left[F_T\left(\frac{y-\xi}{\eta}; \nu\right)\right]^{\alpha-1} \left[F_T\left(\frac{k-\xi}{\eta}; \nu\right)\right]^{-\alpha} \tag{10}$$

for $-\infty < y \leq k$.

Maximum Likelihood Estimation for TPT Model

Given a sample $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ of the TPT distribution in the value k , the log-likelihood function for vector $\varphi = (\gamma, \sigma, \alpha, \nu)^\top$, where $\gamma = \sigma\xi$ and $\varsigma = 1/\eta$, is given by

$$\begin{aligned} \ell(\varphi; \mathbf{Y}) &= -n_1 \alpha \log\{F_T(c; \nu)\} \\ &+ \sum_{y_i \leq k}^n \{\log \alpha + \log \sigma + \log f_T(z_i; \nu) + (\alpha - 1) \log F_T(z_i; \nu)\} \end{aligned} \tag{11}$$

where $c = \zeta k - \gamma$, $z_i = \zeta y_i - \gamma$, and n_1 is the number of observations in the sample such that $-\infty < y \leq k$. Deriving partially the log-likelihood function (11) with respect to the components of the vector φ the following elements of the score function are obtained

$$\begin{aligned}
 U(\gamma) &= n_1 \alpha r(c; \nu) + \sum_{y_i \leq k} \left\{ \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} z_i - (\alpha - 1) r(z_i; \nu) \right\} \\
 U(\sigma) &= -n_1 \alpha r(c; \nu) k + \sum_{y_i \leq k} \left\{ \frac{1}{\sigma} - \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} z_i + (\alpha - 1) r(z_i; \nu) \right\} \\
 U(\alpha) &= -n_1 \log\{F_T(c; \nu)\} + \sum_{y_i \leq k} \left\{ \frac{1}{\alpha} + \log F_T(z_i; \nu) \right\} \\
 U(\nu) &= -\frac{n_1 \alpha}{2} \left\{ \psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - b_{01}(c; \nu) - \frac{cr(c; \nu)}{\nu} \right\} \\
 &\quad + \frac{1}{2} \sum_{y_i \leq k} \left\{ \psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - \frac{1}{\nu} \right. \\
 &\quad \left. - \log \left(1 + \frac{z_i^2}{\nu} \right) + \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} \frac{z_i^2}{\nu} \right\} \\
 &\quad + \frac{\alpha - 1}{2} \sum_{y_i \leq k} \left\{ \psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - b_{01}(z_i; \nu) - \frac{z_i r(z_i; \nu)}{\nu} \right\}
 \end{aligned}$$

where $r(x; \nu) = f_T(x; \nu) / F_T(x; \nu)$, $\psi(x) = \frac{d}{dx} \log \Gamma(x)$ is the digamma function and $b_{mn}(c_m; \nu + m)$ is given by (8). To obtain the ML estimates of the parameters ξ , η , α and ν in the TPT model, we proceed in a similar way to the CPT model, and iterative methods based on the Newton-Rapshon algorithm are used with the score functions. We use the observed information matrix for calculating the standard errors and to recover the information matrix \mathbf{I}_ϑ of the original parameterization $\vartheta = (\xi, \eta, \alpha, \nu)^\top$, we use $\mathbf{I}_\vartheta = (\partial \varphi / \partial \vartheta)^\top \mathbf{I}_\varphi (\partial \varphi / \partial \vartheta)$, where $\partial \varphi / \partial \vartheta$ is given in (9).

CENSORED POWER-STUDENT-T REGRESSION MODEL

In this section, we introduce the censored power-Student- t regression model, which is denoted by PTCR. This model results from the consideration of the observed random variable $Y_i^o = D_i Y_i$, where $D_i = I_{(0, +\infty)}(Y_i)$ and $Y_i = \mathbf{x}_i^\top \beta + \varepsilon_i$, with $\varepsilon_i \stackrel{iid}{\sim} \text{PT}(0, \eta, \alpha, \nu)$, for $i = 1, \dots, n$; i.e.,

$$Y_i^o = \begin{cases} \mathbf{x}_i^\top \beta + \varepsilon_i, & \text{if } Y_i > 0, \\ 0, & \text{if } Y_i \leq 0, \end{cases} \tag{12}$$

where β is a vector of dimension p of unknown parameters, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$, for $i = 1, \dots, n$, are vectors of known covariates, and ε_i , for $i = 1, \dots, n$, are independent random variables with PT distribution with location parameter 0, scale parameter η , shape parameter α , and degrees of

freedom ν . This assumption is equivalent to considering that unobserved random variables Y_1, \dots, Y_n are independent with $Y_i \sim \text{PT}(\mathbf{x}_i^\top \beta, \eta, \alpha, \nu)$, that is, with PDF given by

$$g(y_i; \vartheta) = \frac{\alpha}{\eta} f_T \left(\frac{y_i - \mathbf{x}_i^\top \beta}{\eta}; \nu \right) \left[F_T \left(\frac{y_i - \mathbf{x}_i^\top \beta}{\eta}; \nu \right) \right]^{\alpha-1}$$

for $i = 1, \dots, n$, where $\vartheta = (\beta^\top, \eta, \alpha, \nu)^\top$. The contribution to likelihood function for observations $Y_i^o = 0$ is given by

$$P(Y_i^o = 0) = \left\{ F_T \left(-\frac{\mathbf{x}_i^\top \beta}{\eta}; \nu \right) \right\}^\alpha,$$

and for observations $Y_i^o \geq 0$, we have that $Y_i^o \sim \text{PT}(\mathbf{x}_i^\top \beta, \eta, \alpha, \nu)$. Therefore, the likelihood function of the PTCR model based on the observed sample $\mathbf{Y}^o = (Y_1^o, \dots, Y_n^o)^\top$ is given by

$$\begin{aligned} L(\vartheta; \mathbf{Y}^o) &= \prod_{i=1}^n \left\{ F_T \left(-\frac{\mathbf{x}_i^\top \beta}{\eta}; \nu \right) \right\}^{\alpha(1-d_i)} \left\{ \frac{\alpha}{\eta} f_T \left(\frac{y_i - \mathbf{x}_i^\top \beta}{\eta}; \nu \right) \right. \\ &\quad \left. \times \left[F_T \left(\frac{y_i - \mathbf{x}_i^\top \beta}{\eta}; \nu \right) \right]^{\alpha-1} \right\}^{d_i} \end{aligned} \tag{13}$$

where

$$d_i = \begin{cases} 1, & \text{if } Y_i^o > 0, \\ 0, & \text{if } Y_i^o = 0, \end{cases}$$

Model (12) can be extended to the situation where the value of the censorship associated with the observation i is replaced by the value k_i (a known value), i.e.,

$$Y_i^o = \begin{cases} \mathbf{x}_i^\top \beta + \varepsilon_i, & \text{if } Y_i > k_i, \\ k_i, & \text{if } Y_i \leq k_i, \end{cases} \tag{14}$$

for $i = 1, \dots, n$. Note that, by making $Y_i^{o*} = Y_i^o - k_i$, $\mathbf{x}_i^* = (\mathbf{x}_i^\top, k_i)^\top$ and $\beta^* = (\beta^\top, -1)^\top$, we have the previous model in (12), hence, the results of the inference based on the ML method can be used to fit the more general model in (14).

Proposition 3. Consider the model (12) with assumption $\varepsilon_i \stackrel{\text{iid}}{\sim} \text{PT}(0, \eta, \alpha, \nu)$, for $i = 1, \dots, n$, then

- (i) if $\alpha = 1$, model (12) is reduced to t-Student censored regression (TCR) model
- (ii) if $\nu \rightarrow \infty$, model (12) converges to power-normal censored regression (PNCR) model (see Martínez-Flórez et al. 2013)
- (iii) if $\alpha = 1$ and $\nu \rightarrow \infty$, model (12) converges to usual normal censored regression (NCR) model, that is, Tobit model.

Proof. Proof of (i)-(iii) are directly obtained from definition of PTCR model. □

Moments

Proposition 4. *The mean and variance for the i th observed response in PTCR model are given by*

$$E[Y_i^o] = \eta(c_i + m_1(c_i; \nu)) (1 - [F_T(-c_i; \nu)]^\alpha) \tag{15}$$

and

$$\begin{aligned} \text{Var}[Y_i^o] &= \eta^2(c_i^2 + 2c_i m_1(c_i; \nu)) ([F_T(-c_i; \nu)]^\alpha) (1 - [F_T(-c_i; \nu)]^\alpha) \\ &\quad + \eta^2(m_2(c_i; \nu) - [m_1(c_i; \nu)]^2 (1 - [F_T(-c_i; \nu)]^\alpha)) (1 - [F_T(-c_i; \nu)]^\alpha) \end{aligned} \tag{16}$$

respectively, where $c_i = \mu_i/\eta$ with $\mu_i = \mathbf{x}_i^\top \beta$ and

$$m_r(c; \nu) = \frac{1}{1 - \{F_T(-c; \nu)\}^\alpha} \int_{F_T(-c; \nu)}^1 [F_T^{-1}(u; \nu)]^r \alpha u^{\alpha-1} du, \quad r = 1, 2.$$

Note that $m_r(c; \nu)$ is the moment $E[(F_T^{-1}(U; \nu))^r]$ of a random variable U with distribution $\text{Beta}(\alpha, 1)$, truncated in the interval $(F_T(-c; \nu), 1)$.

Proof. The mean and variance for the i th observed response in PTCR model can be obtained by noting that $Y_i^o = D_i Y_i$, where $D_i = I_{(0, +\infty)}(Y_i)$ and $Y_i = \mu_i + \eta Z_i$, with $\mu_i = \mathbf{x}_i^\top \beta$ and $Z_i \stackrel{\text{iid}}{\sim} \text{PT}(\alpha, \nu)$, $i = 1, \dots, n$. We have

$$\begin{aligned} E[Y_i^o] &= E[Y_i \mid Y_i > 0] P(Y_i > 0) \\ &= E[\mu_i + \eta Z_i \mid \mu_i + \eta Z_i > 0] P(\mu_i + \eta Z_i > 0) \\ &= \left(\mu_i + \eta E\left[Z_i \mid Z_i > -\frac{\mu_i}{\eta}\right] \right) P\left(Z_i > -\frac{\mu_i}{\eta}\right) \\ &= \eta \left(\frac{\mu_i}{\eta} + E\left[Z_i \mid Z_i + \frac{\mu_i}{\eta} > 0\right] \right) \left(1 - P\left(Z_i \leq -\frac{\mu_i}{\eta}\right) \right) \\ &= \eta \left(c_i + E[Z_i \mid Z_i + c_i > 0] \right) \left(1 - [F_T(-c_i; \nu)]^\alpha \right) \end{aligned}$$

By Lemma 1 in Appendix A, we have

$$E[Y_i^o] = \eta(c_i + m_1(c_i; \nu)) (1 - [F_T(-c_i; \nu)]^\alpha)$$

To obtain the variance of Y_i^o , note that

$$\begin{aligned} E[(Y_i^o)^2] &= E[Y_i^2 \mid Y_i > 0] P(Y_i > 0) \\ &= E[(\mu_i + \eta Z_i)^2 \mid \mu_i + \eta Z_i > 0] P(\mu_i + \eta Z_i > 0) \\ &= \left(\mu_i^2 + 2\eta\mu_i E\left[Z_i \mid Z_i > -\frac{\mu_i}{\eta}\right] + \eta^2 E\left[Z_i^2 \mid Z_i > -\frac{\mu_i}{\eta}\right] \right) P\left(Z_i > -\frac{\mu_i}{\eta}\right) \\ &= \eta^2 \left(\frac{\mu_i^2}{\eta^2} + 2\frac{\mu_i}{\eta} E\left[Z_i \mid Z_i > -\frac{\mu_i}{\eta}\right] + E\left[Z_i^2 \mid Z_i > -\frac{\mu_i}{\eta}\right] \right) \left(1 - P\left(Z_i \leq -\frac{\mu_i}{\eta}\right) \right) \\ &= \eta^2 \left(c_i^2 + 2c_i E[Z_i \mid Z_i + c_i > 0] + E[Z_i^2 \mid Z_i + c_i > 0] \right) \left(1 - [F_T(-c_i; \nu)]^\alpha \right) \end{aligned}$$

Using Lemma 1 in Appendix, it follows that

$$E[(Y_i^o)^2] = \eta^2 \left(c_i^2 + 2c_i m_1(c_i; \nu) + m_2(c_i; \nu) \right) \left(1 - [F_T(-c_i; \nu)]^\alpha \right)$$

thus, by calculating $Var[Y_i^o] = E[(Y_i^o)^2] - (E[Y_i^o])^2$ and after some algebraic manipulations, we obtain

$$Var[Y_i^o] = \eta^2 \left(c_i^2 + 2c_i m_1(c_i; \nu) \right) \left([F_T(-c_i; \nu)]^\alpha \right) \left(1 - [F_T(-c_i; \nu)]^\alpha \right) + \eta^2 \left(m_2(c_i; \nu) - [m_1(c_i; \nu)]^2 (1 - [F_T(-c_i; \nu)]^\alpha) \right) \left(1 - [F_T(-c_i; \nu)]^\alpha \right)$$

□

It is important to note that, if ν tends to infinity, then (15) and (16) converge to the mean and variance of the PNCR model (Martínez-Flórez et al. 2013), i.e., when $\nu \rightarrow \infty$

$$E[Y_i^o] = \eta \left(c_i + m_1(c_i) \right) \left(1 - [\Phi(-c_i)]^\alpha \right) \\ Var[Y_i^o] = \eta^2 \left(c_i^2 + 2c_i m_1(c_i) \right) \left([\Phi(-c_i)]^\alpha \right) \left(1 - [\Phi(-c_i)]^\alpha \right) + \eta^2 \left(m_2(c_i) - [m_1(c_i)]^2 (1 - [\Phi(-c_i)]^\alpha) \right) \left(1 - [\Phi(-c_i)]^\alpha \right)$$

where

$$m_r(c) = \frac{1}{1 - \{\Phi(-c)\}^\alpha} \int_{\Phi(-c)}^1 [\Phi^{-1}(u)]^r \alpha u^{\alpha-1} du.$$

with $\Phi(\cdot)$ the CDF of the standard normal distribution, and $\Phi^{-1}(\cdot)$ the inverse function of $\Phi(\cdot)$. Also, worth noting that, if $\alpha = 1$ and $\nu \rightarrow \infty$, we have $m_1(c_i; \nu) \rightarrow \varphi(c_i)/\Phi(c_i)$ and $m_2(c_i; \nu) \rightarrow (\Phi(c_i) - c_i\varphi(c_i))/\Phi(c_i)$, thus

$$E[Y_i^o] = \eta \left(c_i \Phi(c_i) + \varphi(c_i) \right) \\ Var[Y_i^o] = \eta^2 \left(c_i^2 \Phi(c_i) + 2c_i \varphi(c_i) \right) \left(1 - \Phi(c_i) \right) + \eta^2 \left(\Phi(c_i) - \varphi(c_i)[c_i + \varphi(c_i)] \right)$$

which are the mean and variance, respectively, of the Tobit model (Tobin 1958).

Maximum Likelihood Estimation

The ML method is considered by using the reparameterization of Olsen (1978). Let $\gamma = \zeta\beta$, $\zeta = 1/\eta$ and

$$d_i = \begin{cases} 1, & \text{if } Y_i^o > k_i, \\ 0, & \text{if } Y_i^o = k_i, \end{cases}$$

the log-likelihood function for $\varphi = (\gamma^\top, \sigma, \alpha, \nu)^\top$ obtained from (13) under the new parameterization is given by

$$\ell(\varphi; \mathbf{Y}^o) = \alpha \sum_{i=1}^n (1 - d_i) \log\{1 - F_T(c_i; \nu)\} + \sum_{i=1}^n d_i \{ \log \alpha + \log \sigma + \log f_T(z_i; \nu) + (\alpha - 1) \log F_T(z_i; \nu) \} \tag{17}$$

where $z_i = c_i y_i - c_i$, with $c_i = \mathbf{x}_i^\top \boldsymbol{\gamma}$. The components of the score function $\mathbf{U}(\boldsymbol{\varphi})$ are obtained by deriving $\ell(\boldsymbol{\varphi}; \mathbf{Y}^0)$ partially in relation to the components $\boldsymbol{\gamma}$, η , α and ν . After some algebraic manipulations the following components of the score function are obtained

$$U(\boldsymbol{\gamma}) = -\alpha \sum_{i=1}^n (1 - d_i) r(-c_i; \nu) \mathbf{x}_i + \sum_{i=1}^n d_i \left\{ \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} z_i - (\alpha - 1) r(z_i; \nu) \right\} \mathbf{x}_i \tag{18}$$

$$U(\sigma) = \sum_{i=1}^n d_i \left\{ \frac{1}{\sigma} - \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} z_i y_i + (\alpha - 1) r(z_i; \nu) y_i \right\} \tag{19}$$

$$U(\alpha) = \sum_{i=1}^n (1 - d_i) \log\{1 - F_T(c_i; \nu)\} + \sum_{i=1}^n d_i \left\{ \frac{1}{\alpha} + \log F_T(z_i; \nu) \right\} \tag{20}$$

$$U(\nu) = -\frac{\alpha}{2} \sum_{i=1}^n (1 - d_i) \left\{ \psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - b_{01}(c_i; \nu) - \frac{c_i r(c_i; \nu)}{\nu} \right\} R(c_i; \nu) + \frac{1}{2} \sum_{i=1}^n d_i \left\{ \psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - \frac{1}{\nu} - \log \left(1 + \frac{z_i^2}{\nu} \right) \right\} + \frac{1}{2} \sum_{i=1}^n d_i \left\{ \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} \frac{z_i^2}{\nu} \right\} + \frac{\alpha - 1}{2} \sum_{i=1}^n d_i \left\{ \psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - b_{01}(z_i; \nu) - \frac{z_i r(z_i; \nu)}{\nu} \right\} \tag{21}$$

where $r(x; \nu) = f_T(x; \nu)/F_T(x; \nu)$, $R(x; \nu) = F_T(x; \nu)/(1 - F_T(x; \nu))$, $\psi(x) = \frac{d}{dx} \log \Gamma(x)$ is the digamma function and $b_{mn}(c_m; \nu + m)$ is the truncated moment defined by (8). Note that, if $\alpha = 1$ the equations (18)-(21) are reduced to the functions of the TCR model (Arellano-Valle et al. 2012), while, if $\alpha = 1$ and $\nu \rightarrow \infty$, then $r(c; \nu) \rightarrow r(c) = \varphi(c)/\Phi(c)$, $R(c; \nu) \rightarrow R(c) = \Phi(c)/(1 - \Phi(c))$, and $U(\nu) \rightarrow 0$, therefore, the equations (18) and (19) are reduced to score functions of the Tobit model.

The elements of the observed information matrix \mathbf{J}_φ for PTCR model, which are denoted by $j_{\varphi_i \varphi_j}$, can be obtained by calculating the second partial derivative of the log-likelihood function (17), i.e., $j_{\varphi_i \varphi_j} = -\partial^2 \ell(\boldsymbol{\varphi}; \mathbf{Y}^0) / \partial \varphi_i \partial \varphi_j$, while the expected information matrix is obtained as $\mathbf{I}_\varphi = E[\mathbf{J}_\varphi]$, which involves the calculation of truncated expected values that have no closed form and must be obtained numerically. The Appendix B presents the expressions for the elements of the matrices \mathbf{I}_φ and \mathbf{J}_φ . The

expected information matrix \mathbf{I}_ϑ of the original parameterization $\vartheta = (\beta, \eta, \alpha, \nu)^\top$ can be recovered by using $\mathbf{I}_\vartheta = (\partial\varphi/\partial\vartheta)^\top \mathbf{I}_\varphi (\partial\varphi/\partial\vartheta)$, where

$$\frac{\partial\varphi}{\partial\vartheta} = \begin{pmatrix} \frac{1}{\eta} \mathbf{I}_p & -\frac{1}{\eta^2} \beta & 0 & 0 \\ 0 & -\frac{1}{\eta^2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Finally, the ML estimates for $\vartheta = (\beta, \eta, \alpha, \nu)^\top$ can be obtained using iterative methods based on the Newton-Raphson algorithm from the score function (18) - (21) and applying the inverse transformation $\beta = \gamma/\varsigma$ and $\eta = 1/\varsigma$. Estimates of the variances of the estimator can be obtained by evaluating the inverse of the observed information matrix \mathbf{J}_φ^{-1} at the ML estimators $\hat{\varphi} = (\hat{\gamma}, \hat{\sigma}, \hat{\alpha}, \hat{\nu})^\top$ and by using the previous result.

Model Selection and Residual Analysis

In this section some criteria for the selection of the best-fitted model and a methodology for residual analysis are proposed.

Model Selection

Many model selection tools are generally used, such as the Akaike information criteria (AIC), (Akaike 1974), Bayesian information criterion (BIC) (Schwarz 1978), and the AIC corrected (AICc) (Sugiura 1978), which are defined by

$$-2\ell(\vartheta \mid \mathbf{y}) + c(k, n)$$

where the term $c(k, n)$ is a quantity that depends on the number of free parameters that are estimated in the model k , and the number of observations in the sample n . For the AIC one has $c(k, n) = 2k$, for BIC $c(k, n) = k \log(n)$ and for AICc, $c(k, n) = 2k(k + 1)/(n - (k + 1))$. To choose the best-fitted model, the criteria AIC, BIC and AICc are used.

Residual Analysis

The residual analysis has the purpose of detecting the presence of atypical observations and to evaluate the assumptions of the model, being able to include formal tests to detect departures from the assumptions of the considered model, as well as informal graphs to present general characteristics of the residuals.

Following Garay et al. (2017) and Arellano-Valle et al. (2012), in this work we considered the transformed martingal residuals r_{MTi} proposed by Barros et al. (2010) as diagnostic tool to evaluate deviations from the postulated model for the response variable, as well as to detect the presence of atypical observations. The residuals are defined as

$$r_{MTi} = \text{sign}(r_{Mi}) \sqrt{-2[r_{Mi} + \delta_i \log(\delta_i - r_{Mi})]}, \quad i = 1, \dots, n$$

where $r_{Mi} = \delta_i + \log S(y_i; \hat{\vartheta})$ is the martingal residual proposed by Ortega et al. (2003), where $\delta_i = 0, 1$ indicates whether the i th observation is censored or not, respectively, $\text{sign}(r_{Mi})$ denotes the sign of r_{Mi} and $S(y_i; \hat{\vartheta}) = P_{\hat{\vartheta}}(Y_i > y_i)$ represents the survival function evaluated at y_i , where $\hat{\vartheta}$ are the MLE for ϑ .

As suggested by Garay et al. (2017), this type of standardized residuals is used due to the fact that they are symmetrically distributed around zero, which facilitates the construction of the simulated envelopes with little computational effort and will be useful to detect an incorrect specification of the model, as well as the presence of observations atypical.

SIMULATIONS STUDIES

Simulation Study 1: Robustness of the Maximum Likelihood Estimates

In this section, we compare the performance of the estimates for PTCR model in the presence of outliers on the response variable. Following Garay et al. (2017) and Mattos et al. (2018) we performed a simulation study based on the NCR model. Specifically, we considered (12) with $\mathbf{x}_i^T = (1, x_i)$ and $\varepsilon_i \sim N(0, \eta^2)$ for $i = 1, \dots, n$. As in Garay et al. (2017) we generated 1000 artificial samples of size $n = 300$, considering $\beta = (\beta_1, \beta_2)^T = (2, 3.5)^T$, $\eta = 2$ and fixing the left censoring level at $p = 10, 20$ and 30% (that is, 10, 20 and 30% of the observations in each data set were left censored, respectively). We generated independently the values x_i , for $i = 1, \dots, n$, from a uniform distribution on the interval $(2, 20)$. These values were fixed throughout the simulations.

To assess how much the ML estimates are influenced by the presence of outliers, we replaced the observation y_{150} by $y_{150}(\delta) = y_{150} + \delta$, with $\delta = 1, 2, \dots, 10$. Let $\hat{\beta}_i(\delta)$ and $\hat{\beta}_i$ be the ML estimates of β_i with and without contamination, respectively, $i = 1, 2$. We are particularly interested in the relative changes

$$RC(\hat{\beta}_i(\delta)) = |(\hat{\beta}_i(\delta) - \hat{\beta}_i) / \hat{\beta}_i|$$

We define the relative changes for η analogously. For each replication we obtained the parameter estimates with and without outliers, under the PTCR model. Table II and Figure 2 depict the average values of the relative changes across all samples and different censoring levels.

We observe that influence increases dramatically when δ increases for $p = 10\%$, specially for the parameter. However, for the $p = 20$ and 30%, these measures vary little, which indicates that PTCR model is more robust in these cases in the presence of discrepant observations.

Simulation Study 2: Asymptotic properties

To study the performance of the ML estimator $\hat{\vartheta} = (\hat{\beta}^T, \hat{\eta}, \hat{\alpha}, \hat{\nu})^T$, a Monte Carlo simulation study with sample sizes $n = 150, 300, 750$, and 1000 is presented. We considered the PTCR model defined in Section "Censored Power-Student-t Regression Model" with $\mathbf{x}_i^T = (1, x_i)$ for $i = 1, \dots, n$. The true values of the parameters were taken as $\beta = (2, 1.5)^T$, $\eta = 1.5$, $\alpha = 0.4, 2.5$ and $\nu = 3.0$. We also consider levels of censorship equal to $p = 0, 10, 25$ and 45%. The covariate x_i was generated from a uniform distribution $U(0, 1, 20)$ as considered in Garay et al. (2017). For each combination of parameters, sample sizes and censorship levels, 2000 samples of the PTCR model were generated with errors $\varepsilon_i \sim PT(0, \eta, \alpha, \nu)$. To

Table II. Average relative changes on estimates for different contaminations δ and censoring level p .

Cens. Level	RC(\cdot)	δ									
		1	2	3	4	5	6	7	8	9	10
10%	RC(β_1)	0.00194	0.02610	0.01427	0.05336	0.08081	0.09474	0.12469	0.15554	0.15260	0.15643
	RC(β_2)	0.00010	0.00004	0.00009	0.00022	0.00024	0.00032	0.00045	0.00060	0.00064	0.00080
	RC(η)	0.000046	0.01137	0.00780	0.02271	0.03341	0.03780	0.04594	0.05558	0.05101	0.04596
20%	RC(β_1)	0.01337	0.01963	0.00945	0.06445	0.12935	0.10080	0.09735	0.07929	0.04515	0.03117
	RC(β_2)	0.00010	0.00008	0.00030	0.00031	0.00002	0.00018	0.00040	0.00075	0.00124	0.00157
	RC(η)	0.00219	0.00020	0.00160	0.00849	0.00984	0.00615	0.00417	0.00024	0.00910	0.01333
30%	RC(β_1)	0.02466	0.00670	0.10623	0.06087	0.09179	0.15043	0.05653	0.19465	0.03994	0.16353
	RC(β_2)	0.00062	0.00190	0.00016	0.00109	0.00006	0.00066	0.00019	0.00015	0.00012	0.00123
	RC(η)	0.00176	0.00860	0.01234	0.00391	0.00881	0.01215	0.00324	0.01675	0.00467	0.01425

Table III. Performance of the ML estimators of $\beta_0, \beta_1, \tau, \alpha$ and ν for the PTCR model.

n	$\hat{\theta}$	$\alpha = 0.40$												$\alpha = 2.5$											
		p = 0%		p = 10%		p = 25%		p = 45%		p = 0%		p = 10%		p = 25%		p = 45%									
		RB	MSE	RB	MSE	RB	MSE	RB	MSE	RB	MSE	RB	MSE	RB	MSE	RB	MSE								
150	$\hat{\beta}_1$	0.010	0.982	0.019	1.127	0.022	0.746	0.050	0.657	0.092	0.700	0.132	0.813	0.138	0.913	0.155	1.200								
	$\hat{\beta}_2$	0.002	0.083	0.002	0.070	0.002	0.047	0.001	0.041	0.002	0.014	0.002	0.026	0.003	0.034	0.005	0.055								
	$\hat{\eta}$	0.030	0.278	0.035	0.260	0.028	0.225	0.033	0.215	0.022	0.207	0.018	0.231	0.015	0.248	0.004	0.275								
	$\hat{\alpha}$	0.039	0.119	0.040	0.112	0.041	0.102	0.044	0.098	0.309	1.778	0.301	1.878	0.297	1.913	0.315	2.020								
	$\hat{\nu}$	0.309	2.673	0.296	2.428	0.255	2.125	0.223	1.980	0.188	1.735	0.196	1.840	0.211	2.057	0.216	2.228								
300	$\hat{\beta}_1$	0.010	0.839	0.025	0.787	0.014	0.504	0.025	0.441	0.061	0.552	0.086	0.601	0.094	0.689	0.122	0.922								
	$\hat{\beta}_2$	0.001	0.062	0.001	0.051	0.001	0.033	0.001	0.029	0.001	0.011	0.001	0.018	0.001	0.024	0.002	0.038								
	$\hat{\eta}$	0.020	0.203	0.020	0.184	0.014	0.155	0.016	0.147	0.018	0.150	0.018	0.168	0.016	0.178	0.016	0.205								
	$\hat{\alpha}$	0.035	0.091	0.031	0.080	0.026	0.072	0.026	0.068	0.195	1.223	0.184	1.343	0.206	1.472	0.262	1.715								
	$\hat{\nu}$	0.186	1.739	0.160	1.610	0.123	1.263	0.102	1.127	0.093	0.973	0.107	1.025	0.109	1.026	0.128	1.175								
750	$\hat{\beta}_1$	0.002	0.588	0.003	0.481	0.006	0.303	0.009	0.263	0.024	0.219	0.037	0.346	0.044	0.415	0.064	0.574								
	$\hat{\beta}_2$	0.001	0.046	0.001	0.031	0.001	0.021	0.001	0.018	0.001	0.009	0.001	0.012	0.001	0.014	0.001	0.024								
	$\hat{\eta}$	0.008	0.136	0.007	0.111	0.005	0.097	0.006	0.092	0.005	0.077	0.006	0.099	0.007	0.108	0.008	0.128								
	$\hat{\alpha}$	0.009	0.042	0.009	0.045	0.008	0.042	0.008	0.040	0.084	0.599	0.070	0.640	0.089	0.781	0.128	1.024								
	$\hat{\nu}$	0.052	0.715	0.041	0.602	0.034	0.534	0.029	0.502	0.020	0.512	0.035	0.498	0.040	0.531	0.048	0.603								
1000	$\hat{\beta}_1$	0.001	0.417	0.001	0.394	0.004	0.262	0.008	0.227	0.010	0.167	0.026	0.289	0.031	0.347	0.044	0.474								
	$\hat{\beta}_2$	0.001	0.031	0.001	0.025	0.001	0.017	0.001	0.015	0.001	0.009	0.001	0.010	0.001	0.012	0.001	0.020								
	$\hat{\eta}$	0.001	0.091	0.002	0.095	0.003	0.082	0.004	0.078	0.003	0.069	0.005	0.083	0.005	0.093	0.006	0.107								
	$\hat{\alpha}$	0.008	0.035	0.001	0.040	0.006	0.036	0.006	0.035	0.068	0.489	0.051	0.519	0.063	0.609	0.089	0.778								
	$\hat{\nu}$	0.035	0.605	0.036	0.513	0.027	0.444	0.004	0.036	0.013	0.293	0.028	0.401	0.031	0.435	0.038	0.487								

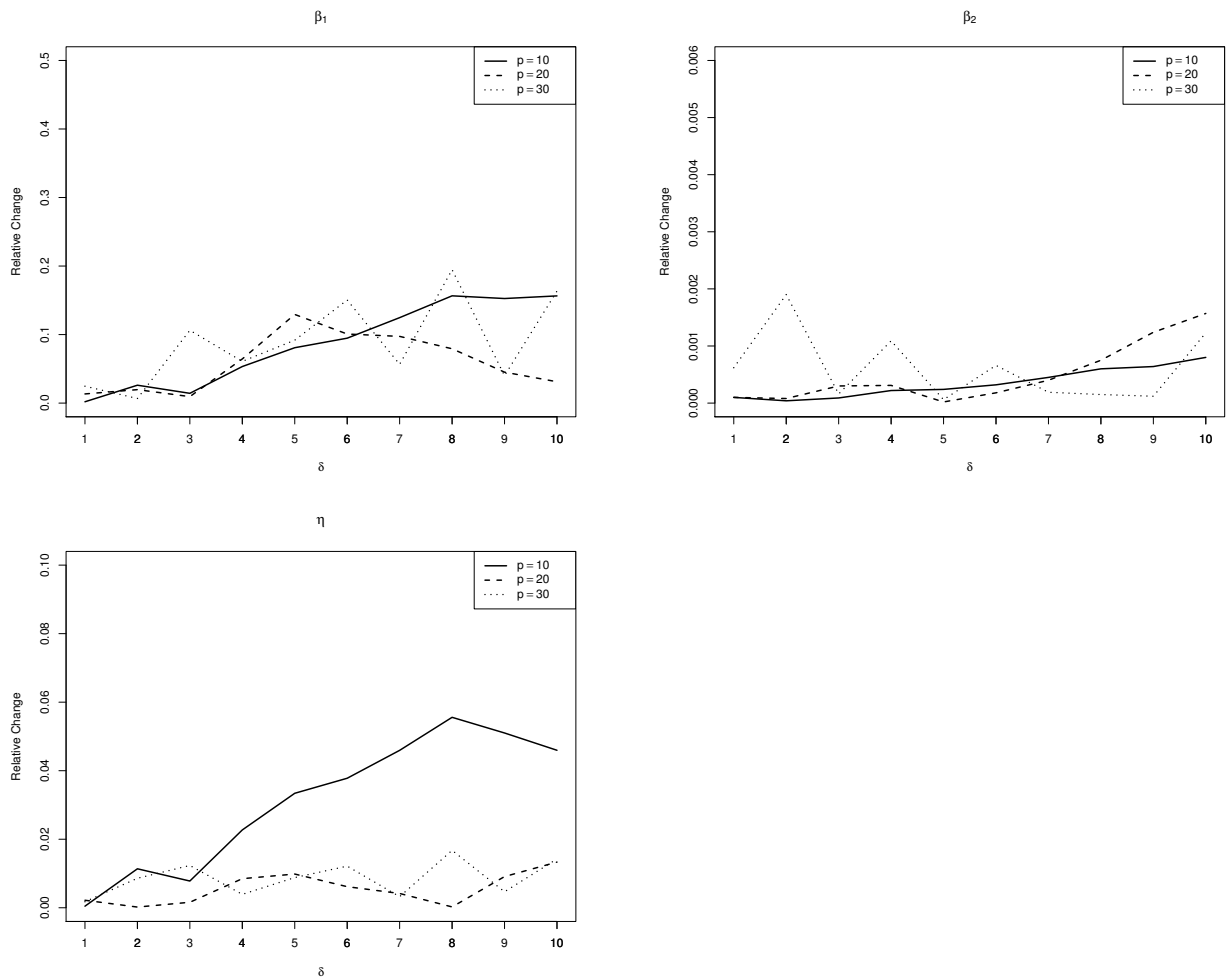


Figure 2. Simulation study 1. Average relative changes on estimates for different contaminations δ and censoring level.

evaluate the performance of the estimators, the absolute value of the relative bias (RB) and the mean square error (MSE) were considered, they are given by

$$RB(\hat{\vartheta}_i) = \frac{1}{2000} \sum_{j=1}^{2000} \left(\frac{\hat{\vartheta}_i^{(j)}}{\vartheta_i} - 1 \right), \quad MSE(\hat{\vartheta}_i) = \frac{1}{2000} \sum_{j=1}^{2000} \left(\hat{\vartheta}_i^{(j)} - \vartheta_i \right)^2,$$

respectively, where $\hat{\vartheta}_i^{(j)}$ is the estimator of ϑ_i for the j th sample, for $\vartheta_i \in \vartheta = (\beta^T, \eta, \alpha, \nu)^T$. The ML estimates of the parameters were calculated by using the *optim* function of R Development Core Team (2018). The optimization of the likelihood function was done using iterative methods based on the Newton-Rapshon algorithm by using the score functions.

It can be seen from the Table III that RB and MSE tend to decrease when the value of n increases, indicating that estimates based on the ML method have good asymptotic properties. That pattern is the same for the different levels of censorship of p under consideration. Note that, when the sample sizes is $n = 150$, the estimates for the β_0 parameter are unstable (in terms of MSE) because it is

affected by the bias of the asymmetry parameter α , however, when the sample size increases, the estimates become more stable. In general, this problem is very common in these types of models, see for example, Martínez-Flórez et al. (2013), so we recommend moderate and large sample sizes in these types of models

REAL DATA APPLICATIONS

Application 1: Wage rate

To illustrate the proposed model, we consider a data set described by Mroz (1987). The data set consists of 753 married white women with ages between 30 and 60 years old in 1975, with 428 women that worked at some point during that year. The response variable used in this application is the wage rate, which represents a measure of the wage of the housewife known as the average hourly earnings. In data set, we have that 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 considered covariates were the wife's age (X_1), years of schooling (X_2), the number of children younger than six years old in the household (X_3), and the number of children between six and nineteen years old (X_4). These data were analyzed previously by Arellano-Valle et al. (2012) using a TCR model and later by Garay et al. (2017) using the Scale Mixture of Normal Censored Regression (SMNCR) models. We analyzed the data set by fitting a PTCR model and we compare our proposal with SMNCR models by Garay et al. (2017): Student- t censored regression model (TCR) (Arellano-Valle et al. 2012), Slash censored regression model (SLCR), and normal censored regression model (NCR), that is, the usual tobit model. Table IV shows skewness and kurtosis index for complete data and also for uncensored observations. Notice that values for the skewness and kurtosis indexes justify using the PTCR.

Table IV. Statistical summary for wage rate data.

	n	Mean	SE	Skewness	Kurtosis
Complete Data	753	2.3746	3.2418	2.7778	15.7967
Non Censored Data	428	4.1777	3.3103	3.0801	16.8212

Table V presents parameter estimates together with their corresponding standard errors (SE) for the PTCR, TCR, SLCR and NCR models. To fit model of the PTCM, we use the R code in the Appendix C. Table VI presents some model selection criteria, together with the values of the log-likelihood. According to the AIC, BIC and AICc criteria, the PTCR model seems to yield a better fit to the Mroz's data than the SMNCR models (TCR and SLCR models) and the usual Tobit model (NCR model), supporting the contention of a departure from symmetry of the errors. Also, the SE of the PTCR model are smaller than SE of the SMNCR and NCR models.

A more emphatic indication that an asymmetric model should be considered comes from testing the hypothesis a TCR model against an asymmetric (PTCR model), that is,

$$H_0 : \alpha = 1 \quad \text{versus} \quad H_1 : \alpha \neq 1,$$

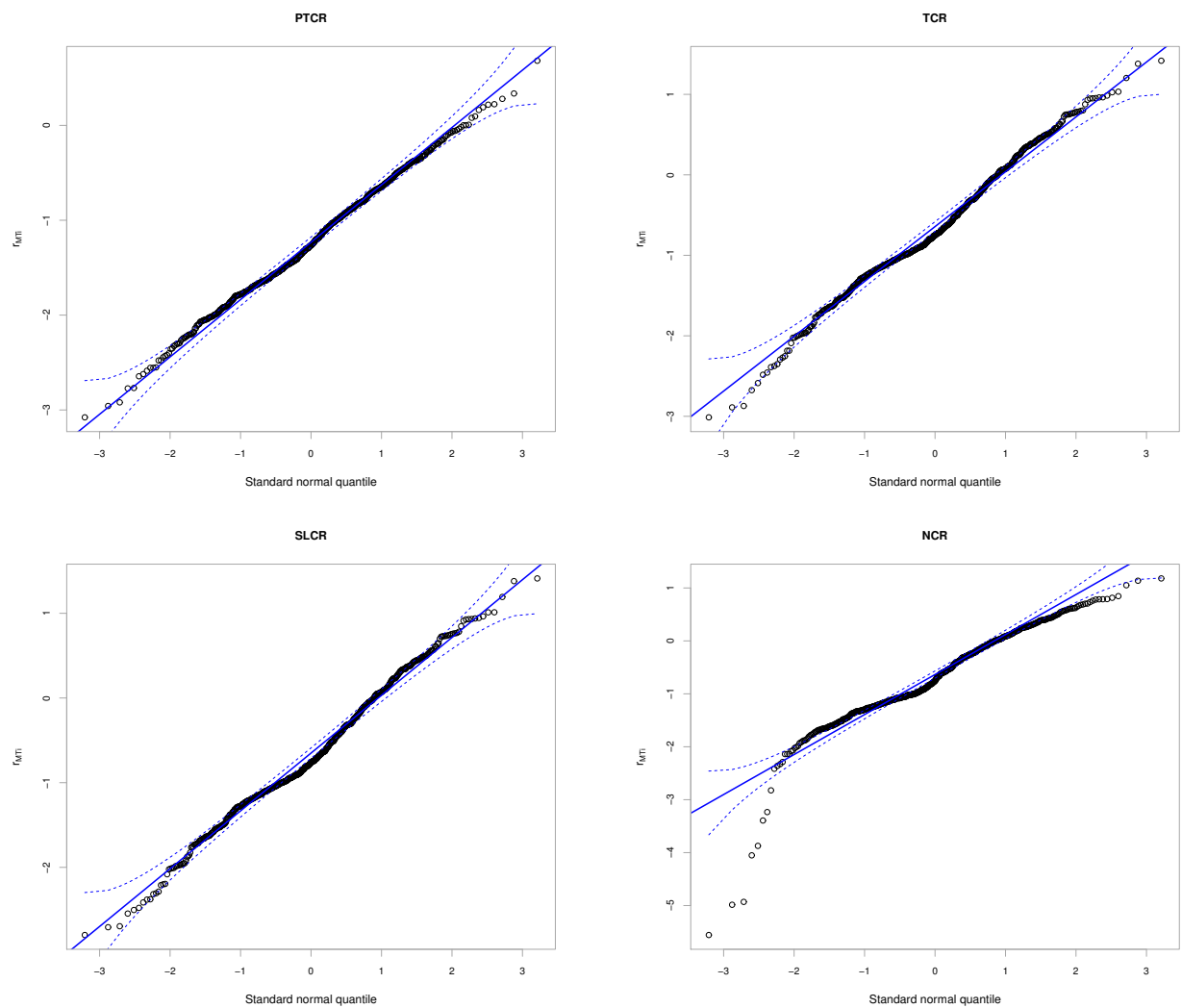


Figure 3. Wage rate data. Envelopes of transformed martingale residuals for PTCR, TCR, SLCR and NCR models.

by using the likelihood ratio (LR) statistics, $-2 \log(\Lambda) = -2(\ell_{TCR}(\hat{\vartheta}) - \ell_{PTCR}(\hat{\vartheta}))$, which for the data set under study, leads to $-2 \log(\Lambda) = 52.41$, which is greater than the critical 5% value with one degree of freedom which is given by $\chi^2_{1,95\%} = 3.84$. This is an indication that the PTCR model fits Mroz’s data better than the ordinary TCR model.

Finally, in order to verify if there is any incorrect specification in the assumptions of the fitted model, the simulated envelope graphs for the transformed martingale residuals are shown in Figure 3. This figure indicates that the PTCR model is, apparently, more suitable for the adjustment of this data than the SMNCR models. It can also be observed that the SMNCR models with heavy tails fit the data better than the NCR model, since there are few observations that are outside the envelopes.

Application 2: Stellar Abundances Data

The second censored dataset is described in Santos et al. (2002) and are available in the R package *astrodatR* (Feigelson 2014) under the name *Stellar abundances*. These data were analyzed Mattos et al.

Table V. Parameters and standard errors (SE) of the PTCR, TCR, SLCR and NCR models fitted to Wage rate data.

ϑ	PTCR		TCR		SLCR		NCR	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_0	-8.8412	0.3785	-1.0474	1.1549	-1.1445	1.4377	-2.7510	1.7334
β_1	-0.1331	0.0004	-0.1108	0.0204	-0.1084	0.0232	-0.1046	0.0276
β_2	0.4887	0.0050	0.6475	0.0658	0.6434	0.0635	0.7281	0.0831
β_3	-3.4010	0.1869	-3.1636	0.3878	-3.0958	0.3785	-3.0264	0.4406
β_4	-0.3053	0.0149	-0.2964	0.1247	-0.2946	0.1210	-0.2143	0.1527
η	4.1538	0.1320	3.2616	0.1731	2.5958	0.3071	4.5760	0.1697
ν	6.5899	0.9014	4.2000	0.8068	1.4061	0.4361	-	-
α	12.8760	0.9871	-	-	-	-	-	-

Table VI. Wage rate data. Model selection criteria.

Criterion	PTCR	TCR	SLCR	NCR
$\ell(\vartheta)$	-1413.942	-1440.145	-1436.286	-1481.655
AIC	2843.883	2894.291	2886.573	2975.311
AICc	2828.077	2880.441	2872.723	2963.424
BIC	2880.876	2926.659	2918.941	3003.055

(2018) by using the Scale Mixture of Skew Normal Censored Regression (SMSNCR) models. We analyzed the data set by fitting a PTCR model and again, we compare our proposal with SMNCR models by Garay et al. (2017).

The dataset consists of measurements for 68 solar-type stars and for our analysis we followed Mattos et al. (2018) and consider:

- $\log N(Be)$ as the response variable, which represents the log of the abundance of beryllium scaled to Sun’s abundance (i.e. the Sun has $\log N(Be) = 0.0$)
- $T_{\text{eff}}/1000$ as the explanatory variable, which represents the effective stellar surface temperature (in kelvin).

In astronomical research, a previously identified sample of objects (stars, galaxies, quasars, X-ray sources, etc.) is observed at some new wavebands. According to Feigelson (2014), due to limited sensitivities, some objects may be undetected, leading to upper limits in their derived luminosities. For this dataset we have 12 left-censored data points, i.e. 12 undetected beryllium measurement, that represents 19.35% of observations.

Table VII presents the ML estimates for the parameters of the four models, i.e. PTCR, TCR, SLCR and NCR models, together with their corresponding standard errors. Table VIII compares the fit of the four models using the model selection criteria (AIC, AIC_c and BIC). Note that again the PTCR model

Table VII. Parameters and standard errors (SE) of the PTCR, TCR, SLCR and NCR models fitted to stellar abundances data.

ϑ	PTCR		TCR		SLCR		NCR	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_0	-1.9692	0.0658	-1.8923	0.1482	-0.5441	0.8733	-1.6478	0.5941
β_1	0.5353	0.0021	0.5091	0.0046	0.2719	0.0262	0.4372	0.0180
η	0.0608	0.0002	0.0873	0.0004	0.1987	0.0254	0.3843	0.0014
ν	2.0891	0.4530	0.8335	0.0419	1.0567	0.1652	–	–
α	0.3265	0.0084	–	–	–	–	–	–

with heavy tails have better fit than the TCR, SLCR and NCR models. The QQ-plots and envelopes for the martingale residuals are shown in Figure 4. This figure clearly indicates that the PTCR, TCR and SLCR models are more suitable for modeling the current data than the NCR model, since there are not observations falling outside the envelope.

Table VIII. Stellar abundances data. Model selection criteria.

Criterion	PTCR	TCR	SLCR	NCR
$\ell(\vartheta)$	3.688	-6.635	-25.096	-38.281
AIC	2.624	21.269	58.192	82.562
AICc	-6.408	13.905	50.827	76.937
BIC	13.722	30.148	67.069	89.220

CONCLUSIONS

In this paper, an asymmetric alternative for the Student-*t* censored regression model by Arellano-Valle et al. (2012) and SMNCR by Garay et al. (2017) has been developed. It is based on the new family of asymmetric and heavy-tailed power-*t* distribution (Zhao & Kim 2016). Moreover, it follows that the ordinary tobit model and the Student-*t* censored regression models are special cases. The observed and expected information matrix is analytically obtained, allowing for the direct implementation of the inference on this type of models. The problem of estimating the parameters in the model is dealt by using the maximum likelihood approach which is also used for developing large sample properties for the estimators. The likelihood ratio statistics can be used for testing the PTCR null hypothesis since the TCR model is special case of the model entertained. Applications to Wage rate data and Stellar Abundances Data indicate that the PTCR model can be a useful alternative to the TCR and SMNCR models.

The proposed PT distribution can be considered in the statistical models based on the scale mixtures of normal family to improve the fit the models such as Maleki & Nematollahi (2017).

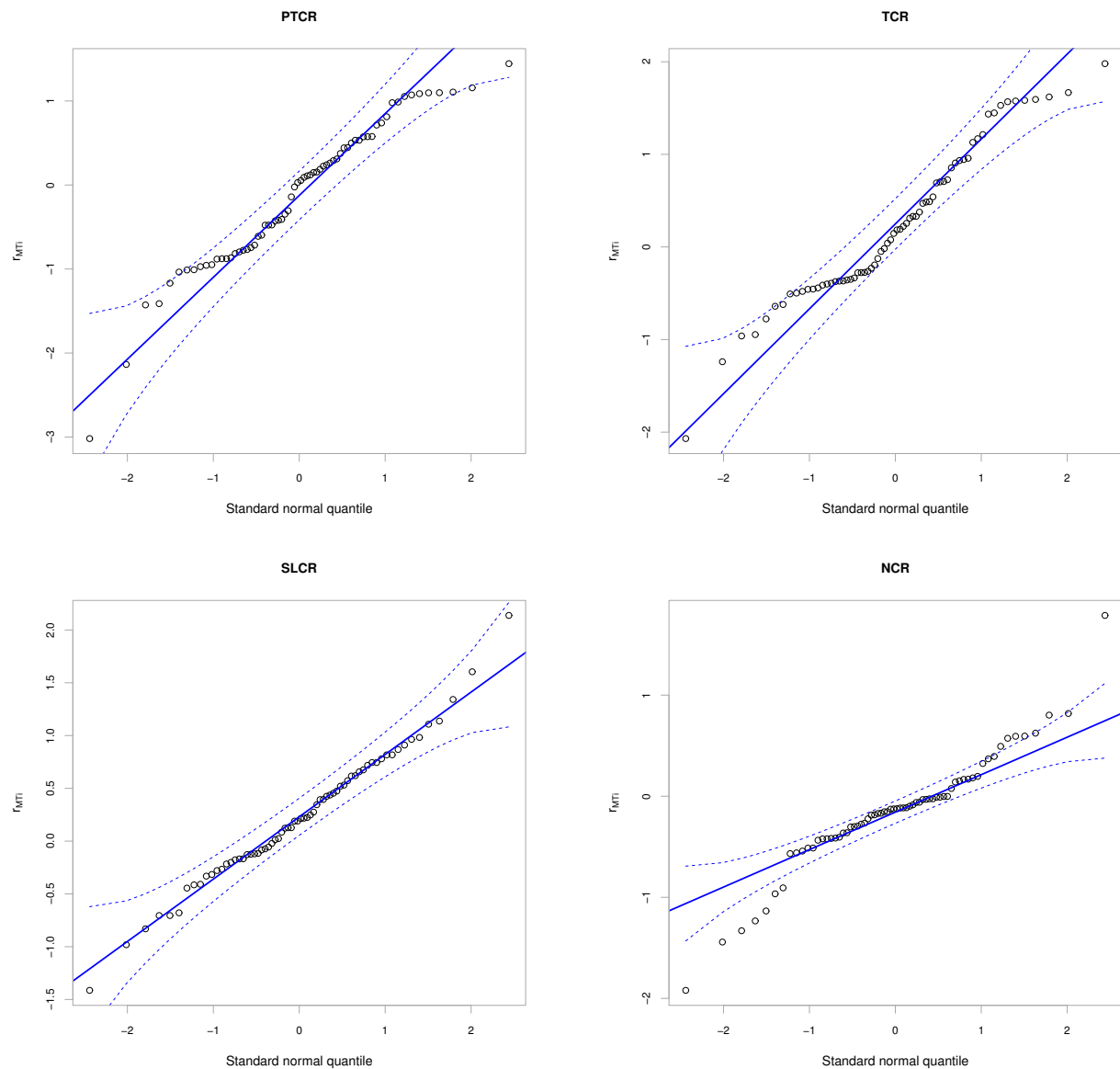


Figure 4. Stellar abundances data. Envelopes of transformed martingale residuals for PTCR, TCR, SLCR and NCR models.

Also, the methodology of constructing the asymmetric distribution on the symmetric version of the Skew-Reflected-Gompertz distribution which recently introduced by Hosseinzadeh et al. (2019), can be considered as a future work for researchers.

REFERENCES

AKAIKE H. 1974. A new look at statistical model identification. *IEEE Trans Automat Contr* AU-19(4): 716-722.
 ARELLANO-VALLE RB, CASTRO LM, GONZÁLEZ-FARÍAS G & MUNÓZ GAJARDO KA. 2012. Student-t censored regression

model: properties and inference. *Stat Methods Appt* 21(4): 453-473. doi:10.1007/s10260-012-0199-y.

AZZALINI A & CAPITANIO A. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J R Stat Soc Series B Stat Methodol* 65(2): 367-389.

- BARROS M, GALEA M, GONZÁLEZ M & LEIVA V. 2010. Influence diagnostics in the tobit censored response model. *Stat Methods Appt* 19(3): 379-397. doi:10.1007/s10260-010-0135-y.
- FAIR RC. 1978. A theory of extramarital affairs. *J Polit Econ* 86(1): 45-61.
- FEIGELSON ED. 2014. *astrodatR: Astronomical Data*. Available at. URL <https://cran.r-project.org/web/packages/astrodatR/>. R package v. 0.1.
- GARAY AM, LACHOS VH, BOLFARINE H & CABRAL CRB. 2017. Linear censored regression models with scale mixtures of normal distributions. *Stat Pap* 58(1): 247-278. doi:10.1007/s00362-015-0696-9.
- GUPTA RD & GUPTA RC. 2008. Analyzing skewed data by power-normal model. *Test* 17: 197-210.
- HOSSEINZADEH A, MALEKI M, KHODADADI Z & CONTRERAS-REYES JE. 2019. The Skew-Reflected-Gompertz distribution for analyzing the symmetric and asymmetric data. *J Comput Appl Math* 349: 132-141.
- MALEKI M & NEMATOLLAHI AR. 2017. Autoregressive Models with Mixture of Scale Mixtures of Gaussian innovations. *Iran J Sci Technol Trans A Sci* 41(4): 1099-1107.
- MARTÍNEZ-FLÓREZ G, BOLFARINE H & GÓMEZ HW. 2013. The alpha-power tobit model. *Commun Stat Theory Methods* 42(4): 633-643.
- MATTOS T, GARAY AM & LACHOS VH. 2018. Likelihood-based inference for censored linear regression models with scale mixtures of skew-normal distributions. *J Appl Stat* 45(11): 2039-2066.
- MOULTON LH & HALSEY NA. 1995. A Mixture Model With Detection Limits for Regression Analyses of Antibody Response to Vaccine. *Biometrics* 51: 1570-1578.
- MROZ TA. 1987. The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* 55(4): 765-799. doi:10.2307/1911029.
- OLSEN RJ. 1978. Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model. *Econometrica* 46(5): 1211-1215.
- ORTEGA EM, BOLFARINE H & PAULA GA. 2003. Influence diagnostics in generalized log-gamma regression models. *Comput Stat Data Anal* 42: 165-186.
- PEWSEY A, GÓMEZ HW & BOLFARINE H. 2012. Likelihood-based inference for power distributions. *Test* 21(4): 775-789.
- R DEVELOPMENT CORE TEAM. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- SANTOS N, LÓPEZ RG, ISRAELIAN G, MAYOR M, REBOLO R, GARCÍA-GIL A, DE TAORO MP & RANDICH S. 2002. Beryllium abundances in stars hosting giant planets. *Astron Astrophys* 386: 1028-1038.
- SCHWARZ G. 1978. Estimating the dimension of a model. *Ann Stat* 6(2): 461-464.
- SUGIURA N. 1978. Further analysis of the data by akaike's information criterion and the finite corrections. *Commun Stat Theory Methods* 7(1): 13-26. doi:10.1080/03610927808827599.
- TOBIN J. 1958. Estimation of relationship for limited dependent variables. *Econometrica* 26(1): 24-36.
- ZHAO J & KIM HM. 2016. Power t distribution. *Commun Stat Appl Methods* 23(4): 321-334.

How to cite

TOVAR-FALÓN R, BOLFARINE H & MARTÍNEZ-FLÓREZ G. 2021. The Asymmetric Power-Student-t Model for Censored and Truncated Data. *An Acad Bras Cienc* 93: e20190920. DOI 10.1590/0001-3765202120190920.

*Manuscript received on August 11, 2019;
accepted for publication on October 22, 2019*

ROGER TOVAR-FALÓN¹

<https://orcid.org/0000-0001-5649-532X>

HELENO BOLFARINE²

<https://orcid.org/0000-0001-9195-3672>

GUILLERMO MARTÍNEZ-FLÓREZ¹

<https://orcid.org/0000-0001-6441-5377>

¹Departamento de Matemáticas y Estadística, Facultad de Ciencias Básicas, Universidad de Córdoba, Cra. 6ª, No. 77-306, Montería, Colombia

²Departamento de Estatística, IME, Universidade de São Paulo, Cidade Universitária, Rua da Matão, 1010, 05508-090 São Paulo, SP, Brazil

Correspondence to: **Roger Tovar-Falón**

E-mail: rjtovar@correo.unicordoba.edu.co

Author contributions

The findings in this paper are part of the first author's (Roger Tovar-Falón) doctoral thesis, which was supervised by the second author (Heleno Bolfarine) and the third author (Guillermo Martínez-Flórez). The participation of the authors in the production of the manuscript is as follows: Roger Tovar-Falón - conceptualization, characterization of the new model, mathematical properties and implementation of computational routines. Guillermo Martínez-Flórez - application, simulation studies, writing the original draft, and computational routines. Heleno Bolfarine - review and general correction of the paper.



APPENDIX A: LEMMAS

Lemma 1. Let $Z \sim PT(\alpha, \nu)$, then $E[Z^k \mid Z + c > 0] = m_k(c; \nu)$, with

$$m_k(c; \nu) = \frac{1}{1 - \{F_T(-c; \nu)\}^\alpha} \int_{F_T(-c; \nu)}^1 [F_T^{-1}(u; \nu)]^k \alpha u^{\alpha-1} du.$$

where $F_T^{-1}(\cdot; \nu)$ is the inverse of $F_T(\cdot; \nu)$.

Lemma 2. Let $Z \sim PT(\alpha, \nu)$, and define $r(Z; \nu) = f_T(Z; \nu) / F_T(Z; \nu)$. Then

(i) $E \left\{ [r(Z; \nu)]^k \left(1 + \frac{Z^2}{\nu}\right)^{-m/2} Z^n \mid Z + c > 0 \right\} = \frac{[f_T(0; \nu)]^k}{1 - \{F_T(-c; \nu)\}^\alpha} a_{kmn}(-c; \nu)$, where

$$a_{kmn}(-c; \nu) = \int_{F_T(-c; \nu)}^1 [F_T^{-1}(u; \nu)]^n \left\{ 1 + \frac{[F_T^{-1}(u; \nu)]^2}{\nu} \right\}^{-\frac{k(\nu+1)+m}{2}} \alpha u^{\alpha-k-1} du$$

(ii) $E \left\{ [r(Z; \nu)]^k \log \left(1 + \frac{Z^2}{\nu}\right) Z^n \mid Z + c > 0 \right\} = \frac{[f_T(0; \nu)]^k}{1 - \{F_T(-c; \nu)\}^\alpha} a_{kn}(-c; \nu)$, where

$$a_{kn}(-c; \nu) = \int_{F_T(-c; \nu)}^1 \log \left\{ 1 + \frac{[F_T^{-1}(u; \nu)]^2}{\nu} \right\} [F_T^{-1}(u; \nu)]^n \alpha u^{\alpha-k-1} du$$

(iii) $E \left\{ [r(Z; \nu)]^k [b_{01}(Z; \nu)]^m Z^n \mid Z + c > 0 \right\} = \frac{[f_T(0; \nu)]^k}{1 - \{F_T(-c; \nu)\}^\alpha} b_{kmn}(-c; \nu)$, where

$$b_{kmn}(-c; \nu) = \int_{F_T(-c; \nu)}^1 [b_{01}(F_T^{-1}(u; \nu))]^m [F_T^{-1}(u; \nu)]^n \alpha u^{\alpha-k-1} du$$

con $b_{01}(c; \nu) = \int_{-\infty}^c \left\{ \log \left(1 + \frac{s^2}{\nu}\right) \right\} \frac{f_T(s; \nu)}{F_T(c; \nu)} ds$.

The proof of the Lemmas 1 and 2 are straightforward and they follow directly for the definition of expected value.

APPENDIX B: INFORMATION MATRICES FOR PTCR MODEL

Observed Information Matrix

The elements of the observed information matrix $\mathbf{J}(\boldsymbol{\varphi})$ for the PTCR model are given by

$$\begin{aligned}
 \mathbf{j}_{\Upsilon\Upsilon}^\top &= \alpha \sum_{i=1}^n (1 - d_i) \left\{ r(-c_i; \nu) - \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{c_i^2}{\nu} \right)^{-1} c_i \right\} r(-c_i; \nu) \mathbf{x}_i \mathbf{x}_i^\top \\
 &\quad + \left(\frac{\nu + 1}{\nu} \right) \sum_{i=1}^n d_i \left\{ \left(1 + \frac{z_i^2}{\nu} \right)^{-1} - \frac{2}{\nu} \left(1 + \frac{z_i^2}{\nu} \right)^{-2} z_i^2 \right\} \mathbf{x}_i \mathbf{x}_i^\top \\
 &\quad + (\alpha - 1) \sum_{i=1}^n d_i \left\{ r(z_i; \nu) + \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} z_i \right\} r(z_i; \nu) \mathbf{x}_i \mathbf{x}_i^\top \\
 \mathbf{j}_{\Upsilon\sigma} &= - \left(\frac{\nu + 1}{\nu} \right) \sum_{i=1}^n d_i \left\{ \left(1 + \frac{z_i^2}{\nu} \right)^{-1} - \frac{2}{\nu} \left(1 + \frac{z_i^2}{\nu} \right)^{-2} z_i^2 \right\} y_i \mathbf{x}_i \\
 &\quad - (\alpha - 1) \sum_{i=1}^n d_i \left\{ r(z_i; \nu) + \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} z_i \right\} \\
 &\quad \times r(z_i; \nu) y_i \mathbf{x}_i \\
 \mathbf{j}_{\Upsilon\nu} &= \frac{\alpha}{2} \sum_{i=1}^n (1 - d_i) \left\{ \left(\psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) \right) (1 + R(c_i; \nu)) \right. \\
 &\quad \left. - \frac{1}{\nu} (1 + c_i r(c_i; \nu) R(c_i; \nu)) - R(c_i; \nu) b_{01}(c_i; \nu) - \log \left(1 + \frac{c_i^2}{\nu} \right) \right. \\
 &\quad \left. + \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{c_i^2}{\nu} \right)^{-1} \frac{c_i^2}{\nu} \right\} r(-c_i; \nu) \mathbf{x}_i \\
 &\quad + \frac{1}{\nu^2} \sum_{i=1}^n d_i \left\{ \left(1 + \frac{z_i^2}{\nu} \right)^{-1} z_i - \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-2} z_i^3 \right\} \mathbf{x}_i \\
 &\quad + \frac{\alpha - 1}{2} \sum_{i=1}^n d_i \left\{ \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} \frac{z_i^2}{\nu} + b_{01}(z_i; \nu) \right. \\
 &\quad \left. - \log \left(1 + \frac{z_i^2}{\nu} \right) - \frac{1}{\nu} (1 - z_i r(z_i; \nu)) \right\} r(z_i; \nu) \mathbf{x}_i
 \end{aligned}$$

$$\begin{aligned}
\mathbf{j}_{\gamma\alpha} &= \sum_{i=1}^n (1-d_i)r(-c_i; \nu)\mathbf{x}_i + \sum_{i=1}^n d_i r(z_i; \nu)\mathbf{x}_i \\
j_{\sigma\sigma} &= \sum_{i=1}^n d_i \left\{ \frac{1}{\sigma^2} + \left(\frac{\nu+1}{\nu}\right) \left[\left(1 + \frac{z_i^2}{\nu}\right)^{-1} - \frac{2}{\nu} \left(1 + \frac{z_i^2}{\nu}\right)^{-2} z_i^2 \right] y_i^2 \right\} \\
&\quad + (\alpha-1) \sum_{i=1}^n l_i \left\{ \left(\frac{\nu+1}{\nu}\right) \left(1 + \frac{z_i^2}{\nu}\right)^{-1} z_i + r(z_i; \nu) \right\} r(z_i; \nu) y_i^2 \\
j_{\sigma\nu} &= -\frac{1}{\nu^2} \sum_{i=1}^n d_i \left\{ \left(1 + \frac{z_i^2}{\nu}\right)^{-1} z_i - \left(\frac{\nu+1}{\nu}\right) \left(1 + \frac{z_i^2}{\nu}\right)^{-2} z_i^3 \right\} y_i \\
&\quad - \frac{\alpha-1}{2} \sum_{i=1}^n d_i \left\{ \left(\frac{\nu+1}{\nu}\right) \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \frac{z_i^2}{\nu} + b_{01}(z_i; \nu) \right. \\
&\quad \left. - \log \left(1 + \frac{z_i^2}{\nu}\right) - \frac{1}{\nu} (1 - z_i r(z_i; \nu)) \right\} r(z_i; \nu) y_i \\
j_{\sigma\alpha} &= -\sum_{i=1}^n d_i r(z_i; \nu) y_i \\
j_{\nu\alpha} &= -\frac{1}{2} \sum_{i=1}^n (1-d_i) \left\{ \psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}(c_i; \nu) - \frac{c_i}{\nu} r(c_i; \nu) \right\} \\
&\quad \times R(c_i; \nu) + \frac{1}{2} \sum_{i=1}^n d_i \left\{ \psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}(z_i; \nu) \right. \\
&\quad \left. - \frac{z_i}{\nu} r(z_i; \nu) \right\} \\
j_{\alpha\alpha} &= -\sum_{i=1}^n \frac{d_i}{\alpha^2}
\end{aligned}$$

$$\begin{aligned}
 j_{\nu\nu} = & + \frac{\alpha}{4} \sum_{i=1}^n (1 - d_i) \left\{ \psi \left(\frac{\nu + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - b_{01}(c_i; \nu) \right. \\
 & \left. - \frac{c_i}{\nu} r(c_i | \nu) \right\}^2 R(c_i; \nu) (1 + R(c_i; \nu)) \\
 & - \frac{\alpha}{4} \sum_{i=1}^n (1 - d_i) \left\{ \psi_1 \left(\frac{\nu}{2} \right) - \psi_1 \left(\frac{\nu + 1}{2} \right) - \frac{2}{\nu(\nu + 1)} + b_{01}^2(c_i; \nu) \right. \\
 & \left. - \frac{1}{\nu} b_{01}(c_i; \nu) + \frac{1}{\nu} \left(2b_{01}(c_i; \nu) - \frac{\nu + 3}{\nu(\nu + 1)} \right) c_i r(c_i; \nu) \right. \\
 & \left. + \frac{1}{\nu + 2} \frac{F_T(c_{2j}; \nu + 2)}{F_T(c_j; \nu)} b_{21}(c_{2j}; \nu + 2) - b_{02}(c_i; \nu) \right. \\
 & \left. + \frac{1}{\nu} \left[\left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{c_i^2}{\nu} \right)^{-1} \frac{c_i^2}{\nu} - \log \left(1 + \frac{c_i^2}{\nu} \right) + \frac{c_i}{\nu} r(c_i; \nu) \right] \right. \\
 & \left. \times c_i r(c_i; \nu) \right\} R(c_i; \nu) \\
 & - \frac{1}{4} \sum_{i=1}^n d_i \left\{ \psi_1 \left(\frac{\nu + 1}{2} \right) - \psi_1 \left(\frac{\nu}{2} \right) + \frac{2}{\nu^2} - \frac{4}{\nu^3} \left(1 + \frac{z_i^2}{\nu} \right)^{-1} z_i^2 \right. \\
 & \left. + \frac{2}{\nu^3} \left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-2} z_i^4 \right\} \\
 & + \frac{\alpha - 1}{4} \sum_{i=1}^n d_i \left\{ \psi_1 \left(\frac{\nu}{2} \right) - \psi_1 \left(\frac{\nu + 1}{2} \right) - \frac{2}{\nu(\nu + 1)} + b_{01}^2(z_i; \nu) \right. \\
 & \left. - \frac{1}{\nu} b_{01}(z_i; \nu) + \frac{1}{\nu} \left(2b_{01}(z_i; \nu) - \frac{\nu + 3}{\nu(\nu + 1)} \right) z_i r(z_i; \nu) \right. \\
 & \left. + \frac{1}{\nu + 2} \frac{F_T(z_{2j}; \nu + 2)}{F_T(z_j; \nu)} b_{21}(z_{2j}; \nu + 2) - b_{02}(z_i; \nu) \right. \\
 & \left. + \frac{1}{\nu} \left[\left(\frac{\nu + 1}{\nu} \right) \left(1 + \frac{z_i^2}{\nu} \right)^{-1} \frac{z_i^2}{\nu} \right. \right. \\
 & \left. \left. - \log \left(1 + \frac{z_i^2}{\nu} \right) + \frac{z_i}{\nu} r(z_i; \nu) \right] z_i r(z_i; \nu) \right\}
 \end{aligned}$$

Expected Information Matrix

The expected information matrix \mathbf{I}_φ is obtained by taking $\mathbf{I}_\varphi = E[\mathbf{J}_\varphi]$, which involves the calculation of truncated moments that have no closed form and must be obtained numerically. The elements $I_{\varphi_i\varphi_j}$ of the expected information matrix are given by

$$\begin{aligned}
 I_{\gamma\gamma} &= \sum_{i=1}^n \left\{ r(-c_i; \nu) - \left(\frac{\nu+1}{\nu} \right) \left(1 + \frac{c_i}{\nu} \right)^{-1} c_i \right\} f_{PT}(-c_i; \alpha, \nu) \mathbf{x}_i \mathbf{x}_i^\top \\
 &\quad + \left(\frac{\nu+1}{\nu} \right) \sum_{i=1}^n \left\{ a_{020}(-c_i; \nu) - \frac{2}{\nu} a_{042}(-c_i; \nu) \right\} \mathbf{x}_i \mathbf{x}_i^\top \\
 &\quad + (\alpha-1) f_T(0; \nu) \sum_{i=1}^n \left\{ \left(\frac{\nu+1}{\nu} \right) a_{121}(-c_i; \nu) + f_T(0; \nu) a_{200}(-c_i; \nu) \right\} \mathbf{x}_i \mathbf{x}_i^\top \\
 I_{\gamma\sigma} &= -\frac{1}{\sigma} \left(\frac{\nu+1}{\nu} \right) \sum_{i=1}^n \{ a_{021}(-c_i; \nu) + c_i a_{020}(-c_i; \nu) \} \mathbf{x}_i \\
 &\quad + \frac{2}{\sigma} \left(\frac{\nu+1}{\nu^2} \right) \sum_{i=1}^n \{ a_{043}(-c_i; \nu) + c_i a_{042}(-c_i; \nu) \} \mathbf{x}_i \\
 &\quad - (f_T(0; \nu))^2 \left(\frac{\alpha-1}{\sigma} \right) \sum_{i=1}^n \{ a_{201}(-c_i; \nu) + c_i a_{200}(-c_i; \nu) \} \mathbf{x}_i \\
 &\quad - f_T(0; \nu) \left(\frac{\alpha-1}{\sigma} \right) \left(\frac{\nu+1}{\nu} \right) \sum_{i=1}^n \{ a_{122}(-c_i; \nu) + c_i a_{121}(-c_i; \nu) \} \mathbf{x}_i \\
 I_{\gamma\nu} &= \frac{\alpha}{2} \sum_{i=1}^n \left\{ \left(\psi \left(\frac{\nu+1}{2} \right) - \psi \left(\frac{\nu}{2} \right) \right) (1 + R(c_i; \nu)) \right. \\
 &\quad \left. - \frac{1}{\nu} (1 + c_i r(c_i; \nu) R(c_i; \nu)) - R(c_i; \nu) b_{01}(c_i; \nu) - \log \left(1 + \frac{c_i^2}{\nu} \right) \right. \\
 &\quad \left. + \left(\frac{\nu+1}{\nu} \right) \left(1 + \frac{c_i^2}{\nu} \right)^{-1} \frac{c_i^2}{\nu} \right\} f_{PT}(-c_i; \alpha, \nu) \mathbf{x}_i \\
 &\quad + \frac{1}{\sqrt{2}} \sum_{i=1}^n \left\{ a_{021}(-c_i; \nu) - \left(\frac{\nu+1}{\nu} \right) a_{043}(-c_i; \nu) \right\} \mathbf{x}_i \\
 &\quad + f_T(0; \nu) \left(\frac{\alpha-1}{2} \right) \sum_{i=1}^n \left\{ \left(\frac{\nu+1}{\nu} \right) a_{122}(-c_i; \nu) - a_{10}(-c_i; \nu) \right. \\
 &\quad \left. + b_{110}(-c_i; \nu) - \frac{1}{\nu} (a_{100}(-c_i; \nu) - f_T(0; \nu) a_{201}(-c_i; \nu)) \right\} \mathbf{x}_i
 \end{aligned}$$

$$\begin{aligned}
 l_{\gamma\alpha} &= \frac{1}{\alpha} \sum_{i=1}^n f_{PT}(-c_i; \alpha, \nu) \mathbf{x}_i + f_T(0; \nu) \sum_{i=1}^n a_{100}(-c_i; \nu) \mathbf{x}_i \\
 l_{\sigma\sigma} &= \frac{1}{\sigma^2} \sum_{i=1}^n \left\{ 1 + \left(\frac{\nu+1}{\nu} \right) [a_{022}(-c_i; \nu) + 2c_i a_{021}(-c_i; \nu) \right. \\
 &\quad \left. + c_i^2 a_{020}(-c_i; \nu) - \frac{2}{\nu} (a_{044}(-c_i; \nu) + 2c_i a_{043}(-c_i; \nu) \right. \\
 &\quad \left. + c_i^2 a_{042}(-c_i; \nu))] \right\} \\
 &\quad + f_T(0; \nu) \left(\frac{\alpha-1}{\sigma^2} \right) \sum_{i=1}^n \left\{ \left(\frac{\nu+1}{\nu} \right) [a_{123}(-c_i; \nu) + 2c_i a_{122}(-c_i; \nu) \right. \\
 &\quad \left. + c_i^2 a_{121}(-c_i; \nu)] + f_T(0; d\nu) [a_{202}(-c_i; \nu) + 2c_i a_{201}(-c_i; \nu) \right. \\
 &\quad \left. + c_i^2 a_{200}(-c_i; \nu)] \right\} \\
 l_{\sigma\alpha} &= -\frac{f_T(0; \nu)}{\sigma} \sum_{i=1}^n \{a_{101}(-c_i; \nu) + c_i a_{100}(-c_i; \nu)\} \\
 l_{\alpha\alpha} &= \frac{1}{\alpha^2} \sum_{i=1}^n (1 - \{F_T(-c_i; \nu)\}^\alpha) \\
 l_{\sigma\nu} &= -\frac{1}{\sigma\nu^2} \sum_{i=1}^n \{a_{022}(-c_i; \nu) + c_i a_{021}(-c_i; \nu) \\
 &\quad - \left(\frac{\nu+1}{\nu} \right) [a_{044}(-c_i; \nu) + c_i a_{043}(-c_i; \nu)] \} \{F_T(-c_i; \nu)\}^\alpha \\
 &\quad - f_T(0; \nu) \left(\frac{\alpha-1}{2\sigma} \right) \sum_{i=1}^n \left\{ \left(\frac{\nu+1}{\nu^2} \right) [a_{123}(-c_i; \nu) + c_i a_{122}(-c_i; \nu)] \right. \\
 &\quad \left. - a_{11}(-c_i; \nu) - c_i a_{10}(-c_i; \nu) + b_{111}(-c_i; \nu) + c_i b_{110}(-c_i; \nu) \right. \\
 &\quad \left. - \frac{1}{\nu} [a_{101}(-c_i; \nu) + c_i a_{100}(-c_i; \nu)] \right. \\
 &\quad \left. + \frac{1}{\nu} [f_T(0; \nu) a_{202}(-c_i; \nu) + c_i f_T(0; \nu) a_{201}(-c_i; \nu)] \right\} \\
 l_{\nu\alpha} &= \frac{1}{2\alpha} \sum_{i=1}^n \left\{ \psi \left(\frac{\nu+1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - b_{01}(c_i; \nu) - \frac{c_i}{\nu} r(c_i; \nu) \right\} \\
 &\quad \times (r(-c_i; \nu))^{-1} f_{PT}(-c_i; \alpha, \nu) \\
 &\quad - \frac{1}{2} \sum_{i=1}^n \left\{ \psi \left(\frac{\nu+1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - \frac{b_{010}(-c_i; \nu)}{1 - \{F_T(-c_i; \nu)\}^\alpha} \right. \\
 &\quad \left. - \frac{f_T(0; \nu) a_{101}(-c_i; \nu)}{\nu(1 - \{F_T(-c_i; \nu)\}^\alpha)} \right\} (1 - \{F_T(-c_i; \nu)\}^\alpha)
 \end{aligned}$$

$$\begin{aligned}
 l_{\nu\nu} = & \frac{\alpha}{4} \sum_{i=1}^n \left\{ \psi \left(\frac{\nu+1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - b_{01}(c_i; \nu) - \frac{c_i}{\nu} r(c_i; \nu) \right\}^2 \\
 & \times R(c_i; \nu)(1 + R(c_i; \nu)) \{F_T(-c_i; \nu)\}^\alpha \\
 & - \frac{\alpha}{4} \sum_{i=1}^n \left\{ \psi_1 \left(\frac{\nu}{2} \right) - \psi_1 \left(\frac{\nu+1}{2} \right) - \frac{2}{\nu(\nu+1)} + b_{01}^2(c_i; \nu) \right. \\
 & - \frac{1}{\nu} b_{01}(c_i; \nu) + \frac{1}{\nu} \left(2b_{01}(c_i; \nu) - \frac{\nu+3}{\nu(\nu+1)} \right) c_i r(c_i; \nu) \\
 & - b_{02}(c_i; \nu) + \frac{F_T(c_{2i}; \nu+2) b_{21}(c_{2i}; \nu+2)}{(\nu+2) F_T(c_{2i}; \nu)} \\
 & + \frac{1}{\nu} \left[\left(\frac{\nu+1}{\nu} \right) \left(1 + \frac{c_i^2}{\nu} \right)^{-1} \frac{c_i^2}{\nu} \right. \\
 & \left. - \log \left(1 + \frac{c_i^2}{\nu} \right) + \frac{c_i}{\nu} r(c_i; \nu) \right] c_i r(c_i; \nu) \left. \right\} R(c_i; \nu) \{F_T(-c_i; \nu)\}^\alpha \\
 & - \frac{1}{4} \sum_{i=1}^n \left\{ \psi_1 \left(\frac{\nu+1}{2} \right) - \psi_1 \left(\frac{\nu}{2} \right) + \frac{2}{\nu^2} - \frac{4a_{022}(-c_i; \nu)}{\nu^3(1 - \{F_T(-c_i; \nu)\}^\alpha)} \right. \\
 & \left. + \frac{2}{\nu^3} \left(\frac{\nu+1}{\nu} \right) \frac{a_{044}(-c_i; \nu)}{(1 - \{F_T(-c_i; \nu)\}^\alpha)} \right\} (1 - \{F_T(-c_i; \nu)\}^\alpha) \\
 & + \frac{\alpha-1}{4} \sum_{i=1}^n \left\{ \psi_1 \left(\frac{\nu}{2} \right) - \psi_1 \left(\frac{\nu+1}{2} \right) - \frac{2}{\nu(\nu+1)} \right. \\
 & + \frac{b_{020}(-c_i; \nu)}{1 - \{F_T(-c_i; \nu)\}^\alpha} - \frac{b_{010}(-c_i; \nu)}{\nu(1 - \{F_T(-c_i; \nu)\}^\alpha)} \\
 & + \frac{1}{\nu} \left[\frac{2f_T(0; \nu) b_{111}(-c_i; \nu)}{1 - \{F_T(-c_i; \nu)\}^\alpha} - \frac{\nu+3}{\nu(\nu+1)} \frac{f_T(0; \nu) a_{101}(-c_i; \nu)}{(1 - \{F_T(-c_i; \nu)\}^\alpha)} \right] \\
 & - E[b_{02}(z_i; \nu)] + \left(\frac{1}{\nu+2} \right) E \left[\frac{F_T(z_{2i}; \nu+2)}{F_T(z_i; \nu)} b_{21}(z_{2i}; \nu+2) \right] \\
 & + \frac{f_T(0; \nu)}{\nu(1 - \{F_T(-c_i; \nu)\}^\alpha)} \left[\left(\frac{\nu+1}{\nu^2} \right) a_{123}(-c_i; \nu) - a_{11}(-c_i; \nu) \right. \\
 & \left. + \frac{1}{\nu} f_T(0; \nu) a_{202}(-c_i; \nu) \right] \left. \right\} (1 - \{F_T(-c_i; \nu)\}^\alpha)
 \end{aligned}$$

where b_{01} , b_{02} and b_{21} are given by (8) and must be obtained numerically. The components a_{kmn} , a_{kn} and b_{kmn} , are given in the Lemman 2.

APPENDIX C: R CODE TO FIT THE PTCR MODEL

```

data( "Mroz87", package = "sampleSelection" )
require(censReg) # This function fit the NCR model
estResult <- censReg( wage ~ age + educ + kids5 + kids618, data = Mroz87 )
summary(estResult)

reg <- lm(wage ~ age + educ + kids5 + kids618, data = Mroz87)
X <- model.matrix(reg)
Y <- Mroz87$wage

# This is the log-likelihood function for PTCR model
loglik.pt <- function(param,Y,X){
p <- ncol(X)
beta<- param[1:p]
eta <- param[p+1]
alp <- param[p+2]
nu <- param[p+3]
x.beta <- X%*%beta
z <- (Y-x.beta)/eta
ll <- vector(mode = "numeric", length = length(z))
di <- vector(mode = "numeric", length = length(z))
for (i in 1:length(z))
{
di[i] <- ifelse(Y[i]==0,0,1)
ll[i] <- (1-di[i]) * alp* pt(z[i],df=nu,log.p=TRUE) +
di[i]*(log(alp) - log(eta) + dt(z[i],df=nu,log=TRUE) + (alp-1)*pt(z[i],df=nu,log.p=TRUE))
}
return(-sum(ll))
}

# Initial values
betas <- as.vector(coef(estResult)[1:ncol(X)],mode="numeric")
sig <- sqrt(sum((Y-X%*%betas)^2)/(nrow(X)-(ncol(X)-1)))
asy = sum((Y-mean(Y))^3/length(Y))/(sum((Y-mean(Y))^2/length(Y))^(3/2))
kur = sum((Y-mean(Y))^4/length(Y))/(sum((Y-mean(Y))^2/length(Y))^(4/2))

# Optimisation of log-likelihood function
param=c(betas,sig,asy,kur)
fit.pt = optim(param, loglik.pt, Y=Y,X=X, hessian=TRUE,method='BFGS')
coef <- c(fit.pt$par)
sds <- diag(solve(fit.pt$hessian))
Est <- round(cbind(coef,sds),4)
colnames(Est)= c('Estimates','SE')
rownames(Est)= c('b0','b1','b2','b3','b4','eta','alp','nu')
Est
-fit.pt$value

```