Stat Methods Appl (2012) 21:453–473
DOI 10.1007/s10260-012-0199-y

Student-*t* censored regression model: properties and inference

Reinaldo B. Arellano-Valle · Luis M. Castro · Graciela González-Farías · Karla A. Muñoz-Gajardo

Accepted: 10 June 2012 / Published online: 27 June 2012

Abstract In statistical analysis, particularly in econometrics, it is usual to consider regression models where the dependent variable is censored (limited). In particular, a censoring scheme to the left of zero is considered here. In this article, an extension of the classical normal censored model is developed by considering independent disturbances with identical Student-*t* distribution. In the context of maximum likelihood estimation, an expression for the expected information matrix is provided, and an efficient EM-type algorithm for the estimation of the model parameters is developed. In order to know what type of variables affect the income of housewives, the results and methods are applied to a real data set. A brief review on the normal censored regression model or Tobit model is also presented.

Keywords ECM algorithm · Limited dependent variable · Maximum likelihood estimation · Student-*t* distribution · Tobit model

R. B. Arellano-Valle · K. A. Muñoz-Gajardo
Departamento de Estadística, Facultad de Matemática, Pontificia Universidad Católica de Chile, Santiago 22, Chile
e-mail: reivalle@mat.puc.cl

K. A. Muñoz-Gajardo
e-mail: kmunoz@mat.puc.cl

L. M. Castro (✉)
Departamento de Estadística, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Concepción, Chile
e-mail: luiscastroc@udec.cl

G. González-Farías
Research Center in Mathematics, CIDICS, Campus de la Salud, UANL Ave. Carlos Canseco s/n con Ave. Gonzalitos, 64460 Col Mitras Centro Monterrey, NL, México
e-mail: farias@cimat.mx

1 Introduction

The problem of estimation for a regression model where the dependent (or endogenous) variable is limited, has been studied in different fields: econometric analysis, clinical essays, wide range of political phenomena, among others. Most of the results on the normal regression model with a censored response variable are based on the development of the so called Tobit model, see Tobin (1958); where the response variable of theoretical interest, Y, is censored. Instead, we may observe a dependent variable Y^{o} given by

$$Y^{\mathrm{o}} = Y\, I(Y > a), \tag{1}$$

for some constant a, where $I(\cdot)$ is the indicator function. Specifically, the *censoring* where the threshold a equals zero receives consideration, hence our Tobit model accepts only positive observations.

When the data are censored, the appropriate distribution for the sample is a mixture of discrete and continuous distributions. In order to analyze this situation, the observed response Y^{o} is considered to be related to the original, but censored, Y throughout (1) with $a = 0$. Hence, the Tobit model corresponds to the censored linear regression model defined by

$$Y_i^{\mathrm{o}} = D_i Y_i \quad \text{and} \quad Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{2}$$

$i = 1, \ldots, n$, where $D_i = I(Y_i > 0)$, $\boldsymbol{\beta}$ is a k-dimensional column vector of unknown parameters, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})^T$, $i = 1, \ldots, n$, are known covariable vectors, and ϵ_i, $i = 1, \ldots, n$, are the model disturbances assumed to be independent and normally distributed, with mean zero and a common variance parameter σ^2. Thus, since $Y_i \overset{ind.}{\sim} \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$, $i = 1, \ldots, n$, where $\mathcal{N}(\mu, \sigma^2)$ represents the normal distribution with mean μ and variance σ^2, we have $P(Y_i^{\mathrm{o}} = 0) = P(Y_i \leq 0) = 1 - \Phi(\mathbf{x}_i^T \boldsymbol{\beta}/\sigma)$, and for the non-nulls Y_i^{o}'s we have that they are distributed as the respective Y_i's. Here, $\Phi(\cdot)$ represents the cumulative distribution function (*cdf*) of the $\mathcal{N}(0, 1)$ distribution. Under this model, Tobin (1958) focused on the estimation of the parameters $\boldsymbol{\beta}$ and σ^2, on the basis of $n = n_0 + n_1$ observations $(x_1, d_1 y_1), \ldots, (x_n, d_n y_n)$, where n_0 and n_1 are the number of observations on the sets $N_0 = \{i : d_i = 0\} = \{i : y_i = 0\}$ and $N_1 = \{i : d_i = 1\} = \{i : y_i > 0\}$, respectively. From the relations mentioned above, the likelihood function for the Tobit model is

$$L_N(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} \left[1 - \Phi\left(\frac{1}{\sigma}\, \mathbf{x}_i^T \boldsymbol{\beta}\right)\right]^{1-d_i} \left[\frac{1}{\sigma}\, \phi\left(\frac{1}{\sigma}\,(y_i - \mathbf{x}_i^T \boldsymbol{\beta})\right)\right]^{d_i}, \tag{3}$$

where $\phi(\cdot)$ denotes the probability density function (*pdf*) of the $\mathcal{N}(0, 1)$ distribution.

Let be, $\mathbf{y} = (y_1, \ldots, y_n)^T$ and $\mathbf{d} = (d_1, \ldots, d_n)^T$, the experimental latent response and observed indicating vectors, respectively, and let be $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ the $n \times k$ corresponding design matrix. We define also the $n \times n$ diagonal matrix $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$, and we denote the $n \times n$ identity matrix by $\boldsymbol{I}_n$. From this notation, we can write the observed and missing/unobserved parts of $\mathbf{y}$ as $\mathbf{y}^{\mathrm{o}} = \mathbf{D}\,\mathbf{y}$ and $\mathbf{y}^{\mathrm{m}} = (\boldsymbol{I}_n - \mathbf{D})\,\mathbf{y}$, respectively. Moreover, the likelihood equations resulting from (3) can be expressed as

$$\sigma^2 = \frac{1}{n_1}\mathbf{y}^T\mathbf{D}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \tag{4}$$

$$\mathbf{X}^T(\boldsymbol{I}_n - \mathbf{D})\boldsymbol{\eta} = \mathbf{X}^T\mathbf{D}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \tag{5}$$

where $\boldsymbol{\eta} = (\sigma\, r(-\mathbf{x}_1^T\boldsymbol{\beta}/\sigma), \ldots, \sigma\, r(-\mathbf{x}_n^T\boldsymbol{\beta}/\sigma))^T$, with $r(z) = \phi(z)/\Phi(z)$.

The two first order conditions defined by (4) and (5) are clearly non-linear equations, which means that there is not a straightforward solution. Thus, iterative procedures, such as the Newton-Raphson or Fisher-scoring methods have been considered in the literature, in order to obtain the maximum likelihood estimates of $\boldsymbol{\beta}$ and σ^2. The observed and expected information matrices were derived by Amemiya (1973), who also studied the asymptotic properties of the maximum likelihood estimators. For a review of the normal theory about the Tobit model, including the maximum likelihood estimations, the bias produced by the censoring and further asymptotic results see Amemiya (1984) and Maddala (1983). Most of these results are developed considering the one-to-one re-parameterization $\boldsymbol{\varphi}_{\mathrm{N}} = (\boldsymbol{\gamma}^T, \tau)^T$ of $\boldsymbol{\theta}_N = (\boldsymbol{\beta}^T, \sigma^2)$, where $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$ and $\tau = 1/\sigma$. Olsen (1978) showed that for the foregoing log-likelihood function, the matrix $\partial^2\log L_N/\partial\boldsymbol{\varphi}_{\mathrm{N}}\partial\boldsymbol{\varphi}_{\mathrm{N}}^T$ is negative semi-definite, with block-entries given by

$$\frac{\partial^2\log L_N}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^T} = -\sum_{i=1}^{n}(1 - d_i)r(-c_i)\,(r(-c_i) - c_i)\,\mathbf{x}_i\mathbf{x}_i^T - \sum_{i=1}^{n} d_i\mathbf{x}_i\mathbf{x}_i^T, \tag{6}$$

$$\frac{\partial^2\log L_N}{\partial\boldsymbol{\gamma}\partial\tau} = \sum_{i=1}^{n} d_i\, y_i\mathbf{x}_i, \tag{7}$$

$$\frac{\partial^2\log L_N}{\partial\tau^2} = -\frac{n_1}{\tau^2} - \sum_{i=1}^{n} d_i\, y_i^2, \tag{8}$$

where $c_i = \mathbf{x}_i^T\boldsymbol{\gamma}$. It is important to stress that, from (6)–(8), the computation of the expected information matrix is straightforward since $E(D_i) = P(Y_i > 0) = \Phi(c_i)$ and, by Lemma 5 of the "Appendix 2", $E(D_iY_i) = E(D_i)E(Y_i \mid Y_i > 0) = (1/\tau)(c_i\Phi(c_i) + \phi(c_i))$ and, similarly, $E(D_iY_i^2) = (1/\tau^2)\{(1 + c_i)\Phi(c_i) + c_i^2\phi(c_i)\}$.

There are several extensions of Tobit model in the literature. For example, Blundell and Meghir (1987) discussed some generalisations of the Tobit model that allow for distinct processes determining the censoring rule and the continuous observations. Alternatively, semi-parametric censored models such as the binary response model, the ordered response model, the grouped dependent model, the multinomial response model among many others can be found in Powell (1994). Hutton and Stanghellini (2011) proposed a censored regression model assuming a skew-normal distribution in order to study health care interventions.

As was mentioned above, most of the theory on censored regression models is related to the described Tobit model, which is constructed in terms of the normal assumption. However, as is well-known, the inferences based in the normal model can be strongly affected by any perturbation of the normality in data. In particular, normal inferences are quite vulnerable to the presence of outliers. Furthermore, many models treat this problem editing the data that represents outlying observations; this last procedure may cause the fact that the uncertainty could not be well reflected when the inference occurs. For that reason, this work is focused on the study of alternative censored regression models, whose construction is based on a more realistic assumption than the normal one. In this sense, natural extensions for the Tobit model can be obtained by assuming e.g., that the distribution of the perturbations belongs to the scale mixtures of normal family of distributions; see Andrews and Mallows (1974), from which the normal model can be obtained as a special (or limit) case. One of the most important and popular members of that class is the Student-t distribution, which can be reduced to the Cauchy or normal distributions depending on whether the degrees of freedom parameter is equal to one or it goes to infinity, respectively. The importance of choosing the Student-t distribution is based on the robustness that it posses; see He et al. (2000), He et al. (2004). In fact, as it was noticed by Lange et al. (1989), this distribution provides a useful extension of the normal one for statistical modeling of data sets involving errors with heavier tails than the normal distribution. The degrees of freedom parameter of the Student-t distribution provides a convenient dimension for achieving robust statistical inference, with moderate increases in computational complexity for many models. Basically, this work describes the censored model using an approach different than that of the normal model, which is quite vulnerable to the presence of outliers.

In this work, an extension of the normal censored regression model (2) to the case where the error terms are independent and have a Student-t distribution is proposed, and a description of the the maximum likelihood approach considering an EM-type algorithm to find the estimators for the model parameters is provided. In addition, the respective observed and expected information matrices are obtained in order to compute the standard errors for the parameter estimates.

The paper is organized as follows. Section 2 introduces the Student-t censored regression model. Maximum likelihood equations and the observed and expected information matrices are given in Sect. 2.2. Section 3 presents a convenient EM algorithm to find the maximum likelihood estimates. An application with a real data set of housewives wages is presented in Sect. 4. Finally, some conclusions are noted in Sect. 5.

2 The Student-*t* censored regression model

This section presents the Student-t censored regression model that results from replacing in (2) the normal assumption for the disturbances with that of them being independent and Student-t distributed, namely $\epsilon_i \overset{ind.}{\sim} t(0, \sigma^2, \nu), i = 1, \ldots, n$, where $t(\mu, \sigma^2, \nu)$ denotes the Student-t distribution with location parameter μ, scale parameter σ^2 and ν degrees of freedom. This is equivalent to considering that the unobserved random variables $Y_1, \ldots, Y_n$ are independent, with $Y_i \sim t(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2, \nu)$, i.e., with *pdf* given by

$$f(y_i \mid \boldsymbol{\theta}) = \frac{1}{\sigma} t\left(\frac{1}{\sigma}(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \mid \nu\right), \tag{9}$$

$i = 1, \ldots, n$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \nu)^T$ and $t(z \mid \nu) = c(\nu)\{1 + z^2/\nu\}^{-(\nu+1)/2}$, with $c(\nu) = \Gamma((\nu+1)/2)/\sqrt{\pi\nu}\,\Gamma(\nu/2)$, is the $t(0, 1, \nu)$-*pdf*. In the rest of the paper, the notation $T(z \mid \nu)$ will represent the $t(0, 1, \nu)$-*cdf*. Thus, one has $P(Y_i^{\mathrm{o}} = 0) = 1 - T(\frac{1}{\sigma}\mathbf{x}_i^T \boldsymbol{\beta} \mid \nu)$, for $i \in N_0$, and $Y_i^{\mathrm{o}} \sim t(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2, \nu)$, for $i \in N_1$. Hence, the likelihood function of the Student-t censored regression model is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left[1 - T\left(\frac{1}{\sigma}\mathbf{x}_i^T \boldsymbol{\beta} \mid \nu\right)\right]^{1-d_i} \left[\frac{1}{\sigma} t\left(\frac{1}{\sigma}(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \mid \nu\right)\right]^{d_i}. \tag{10}$$

Since (10) reduces to the Cauchy censored regression likelihood function when $\nu = 1$ and also since it converges to the Tobit likelihood function in (3) as $\nu \to \infty$, the Student-t censored regression model provides a robust generalization of the Tobit model.

2.1 The mean and variance of a censored Student-t response

In the Student-t censored regression model, the ith observed response is $Y_i^{\mathrm{o}} = D_i Y_i$, where $D_i = I(Y_i > 0)$ and $Y_i = \mu_i + \sigma Z_i$, with $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ and $Z_i \overset{ind.}{\sim} t(0, 1, \nu)$, $i = 1, \ldots, n$. Hence, since $E(Y_i^{\mathrm{o}}) = E(Y_i \mid Y_i > 0)P(Y_i > 0)$ and $Var(Y_i^{\mathrm{o}}) = E(Y_i^2 \mid Y_i > 0)P(Y_i > 0) - [E(Y_i^{\mathrm{o}})]^2$, we have

$$\begin{aligned} E(Y_i^{\mathrm{o}}) &= \{\mu_i + \sigma E(Z_i \mid Z_i + c_i > 0)\}\, T(c_i \mid \nu), \\ Var(Y_i^{\mathrm{o}}) &= \{\mu_i^2 + 2\sigma\mu_i E(Z_i \mid Z_i + c_i > 0)\}\, T(c_i \mid \nu)\{1 - T(c_i \mid \nu)\} \\ &\quad + \sigma^2 \left\{E(Z^2 \mid Z_i + c_i > 0)\, T(c_i \mid \nu) - [E(Z_i \mid Z_i + c_i > 0)\, T(c_i \mid \nu)]^2\right\}, \end{aligned}$$

where $c_i = \mu_i/\sigma$ and, by Lemma 3 of the "Appendix 1",

$$\begin{aligned} E(Z_i \mid Z_i + c_i > 0) &= \left(\frac{\nu}{\nu - 1}\right)\left(1 + \frac{c_i^2}{\nu}\right)\frac{t(c_i \mid \nu)}{T(c_i \mid \nu)}, \quad \nu > 1, \\ E(Z_i^2 \mid Z_i + c_i > 0) &= \left(\frac{\nu}{\nu - 2}\right)\frac{T\,(c_{-2i} \mid \nu - 2)}{T(c_i \mid \nu)} \\ &\quad - c_i E(Z_i \mid Z_i + c_i > 0), \quad \nu > 2, \end{aligned}$$

where $c_{-ki} = \sqrt{\frac{\nu-k}{\nu}}\, c_i$, $\nu > k$. Note also that for $\nu > 2$, the above first truncated mean can be rewritten as $E(Z_i \mid Z_i + c_i > 0) = \sqrt{\frac{\nu}{\nu-2}}\,\frac{t(c_{-2i} \mid \nu-2)}{T(c_i \mid \nu)}$. Hence, the mean and variance of the Student-t censored model are given by

$$E(Y_i^{\mathrm{o}}) = \sigma \left\{ c_i T(c_i \mid \nu) + \left(\frac{\nu}{\nu-1}\right)\left(1+\frac{c_i^2}{\nu}\right) t(c_i \mid \nu) \right\}, \;\; \nu > 1, \tag{11}$$

$$\begin{aligned} Var(Y_i^{\mathrm{o}}) = \sigma^2 &\left\{ c_i^2 T(c_i \mid \nu) + 2c_i \left(\frac{\nu}{\nu-1}\right)\left(1+\frac{c_i^2}{\nu}\right) t(c_i \mid \nu) \right\} \{1 - T(c_i \mid \nu)\} \\ &+ \sigma^2 \left\{ T\,(c_{-2i} \mid \nu-2) - \left[c_{-2i} - t\,(c_{-2i} \mid \nu-2)\right] t\,(c_{-2i} \mid \nu-2) \right\}, \end{aligned} \tag{12}$$

$\nu > 2$. It is important to stress that, when $\nu \to \infty$, (11) and (12) converge to the normal censored mean and variance, respectively, i.e.,

$$\begin{aligned} E(Y_i^{\mathrm{o}}) &\to \sigma\,\{c_i \Phi(c_i) + \phi(c_i)\}, \\ Var(Y_i^{\mathrm{o}}) &\to \sigma^2 \left\{ c_i^2 \Phi(c_i) + c_i \phi\,(c_i) \right\} \{1 - \Phi(c_i)\} \\ &\quad + \sigma^2 \{\Phi\,(c_i) - [c_i - \phi\,(c_i)]\,\phi\,(c_i)\}. \end{aligned}$$

2.2 Maximum likelihood estimation

By convenience, in this section the Olsen (1978)'s reparameterization is considered $\boldsymbol{\gamma} = \tau\boldsymbol{\beta}$ and $\tau = 1/\sigma$. Under this new parameterization, the log-likelihood function for $\boldsymbol{\varphi} = (\boldsymbol{\gamma}^T, \tau, \nu)^T$ obtained from (10) is

$$\begin{aligned} \log L(\boldsymbol{\varphi}) = &\sum_{i=1}^{n} (1-d_i) \log\{1 - T\,(c_i \mid \nu)\} \\ &+ \sum_{i=1}^{n} d_i\,\{\log \tau + \log t\,(z_i \mid \nu)\}, \end{aligned} \tag{13}$$

where $z_i = \tau y_i - c_i$, with $c_i = \mathbf{x}_i^T \boldsymbol{\gamma}$, and

$$\begin{aligned} \log t(z \mid \nu) = &\log \Gamma\left(\frac{\nu+1}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) \\ &- \frac{1}{2}\log(\pi\nu) - \frac{\nu+1}{2}\log\left(1+\frac{z^2}{\nu}\right). \end{aligned} \tag{14}$$

In order to derive the scores components $S_{\boldsymbol{\gamma}} = \partial \log L/\partial \boldsymbol{\gamma}$, $S_\tau = \partial \log L/\partial \tau$ and $S_\nu = \partial \log L/\partial \nu$, the following partial derivatives are considered first:

$$\begin{aligned} &\frac{\partial c_i}{\partial \boldsymbol{\gamma}} = \mathbf{x}_i, \quad \frac{\partial z_i}{\partial \boldsymbol{\gamma}} = -\frac{\partial c_i}{\partial \boldsymbol{\gamma}}, \quad \frac{\partial z_i}{\partial \tau} = y_i, \quad \frac{\partial T\,(c_i \mid \nu)}{\partial \boldsymbol{\gamma}} = t\,(c_i \mid \nu)\,\mathbf{x}_i, \\ &\frac{\partial t(z \mid \nu)}{\partial z} = -\left(\frac{\nu+1}{\nu}\right)\left(1+\frac{z^2}{\nu}\right)^{-1} t(z \mid \nu)\, z, \end{aligned}$$

$$\frac{\partial \log t(z \mid \nu)}{\partial \nu} = \frac{1}{2}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu} - \log\left(1+\frac{z^2}{\nu}\right) \right.$$
$$\left. + \left(\frac{\nu+1}{\nu}\right)\left(1+\frac{z^2}{\nu}\right)^{-1}\frac{z^2}{\nu}\right\},$$
$$\frac{\partial T(c \mid \nu)}{\partial \nu} = \frac{1}{2}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}(c \mid \nu) - \frac{1}{\nu} c\, r(c \mid \nu)\right\} T(c \mid \nu),$$

where $\psi(x) = \mathrm{d}\Gamma(x)/\mathrm{d}x$ is the digamma function and $b_{km}(c_k \mid \nu+k)$ is the truncated moment defined by

$$b_{km}(c_k \mid \nu+k) = \int_{-\infty}^{c_k} z^k \left\{\log\left(1+\frac{z^2}{\nu+k}\right)\right\}^m \frac{t(z \mid \nu+k)}{T(c_k \mid \nu+k)} dz, \tag{15}$$

with $c_k = \sqrt{\frac{\nu+k}{\nu}}\, c$, and $r(z \mid \nu) = t(z \mid \nu)/T(z \mid \nu)$. To calculate the above partial derivative of $T(z \mid \nu)$ with respect to ν, the Lemmas 1–4 from the "Appendix 1" have been used.

Thus, after some simple algebraic manipulations we obtain, from (13) and (14), the following score functions:

$$S_{\boldsymbol{\gamma}} = -\sum_{i=1}^{n}(1-d_i) r(-c_i \mid \nu)\,\mathbf{x}_i + \left(\frac{\nu+1}{\nu}\right)\sum_{i=1}^{n}\left(1+\frac{z_i^2}{\nu}\right)^{-1} z_i\, d_i\, \mathbf{x}_i, \tag{16}$$

$$S_\tau = \sum_{i=1}^{n}\left\{\frac{1}{\tau} - \left(\frac{\nu+1}{\nu}\right)\left(1+\frac{z_i^2}{\nu}\right)^{-1} z_i\, y_i\right\} d_i, \tag{17}$$

$$S_\nu = -\frac{1}{2}\sum_{i=1}^{n}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}(c_i \mid \nu) - \frac{1}{\nu} c_i r(c_i \mid \nu)\right\}$$
$$\times R(c_i \mid \nu)\,(1-d_i) + \frac{1}{2}\sum_{i=1}^{n}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu}\right.$$
$$\left. + \left(\frac{\nu+1}{\nu}\right)\left(1+\frac{z_i^2}{\nu}\right)^{-1}\frac{z_i^2}{\nu} - \log\left(1+\frac{z_i^2}{\nu}\right)\right\} d_i, \tag{18}$$

where $R(c \mid \nu) = T(c \mid \nu)/(1-T(c \mid \nu))$. It is important to stress that if $\nu \to \infty$, then $r(c \mid \nu) \to r(c) = \phi(z)/\Phi(z)$ and $R(c \mid \nu) \to R(c) = \Phi(c)/(1-\Phi(c))$. Therefore, $S_\nu \to 0$, while the Eqs. (16)–(17) are reduced to the following Tobit's score functions:

$$S_{\boldsymbol{\gamma}} = -\sum_{i=1}^{n}(1-d_i) r(-c_i)\mathbf{x}_i + \sum_{i=1}^{n} d_i z_i \mathbf{x}_i, \tag{19}$$

$$S_\tau = \sum_{i=1}^{n} \left\{ \frac{1}{\tau} - z_i y_i \right\} d_i. \tag{20}$$

Finally, note that the expressions (19) and (20) can be reformulated in terms of the original parameterization $\boldsymbol{\theta}_N$ through the Tobit's likelihood equations given by (4) and (5).

2.3 The information matrix

We compute now the information matrix for the Student-t censored regression model. Let $S_{\alpha\lambda}$ be the second partial derivatives of the likelihood function (13) with respect to the components α and λ of $\boldsymbol{\varphi}$. The computation of these elements is simplified by using the following additional results:

$$\begin{aligned}
\frac{\partial r(-z \mid \nu)}{\partial z} &= r(-z \mid \nu) \left\{ r(-z \mid \nu) - \left(\frac{\nu+1}{\nu}\right) \left(1 + \frac{z^2}{\nu}\right)^{-1} z \right\}, \\
\frac{\partial r(-c \mid \nu)}{\partial \nu} &= \frac{1}{2} r(-c \mid \nu) \left\{ \left(\frac{\nu+1}{\nu}\right) \left(1 + \frac{c^2}{\nu}\right)^{-1} \frac{c^2}{\nu} \right. \\
&\quad -\log\left(1 + \frac{c^2}{\nu}\right) - R(c \mid \nu) b_{01}(c \mid \nu) + (1 + R(c \mid \nu)) \\
&\quad \left. \left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right) - \frac{1}{\nu} (1 + c\, R(c \mid \nu) r(c \mid \nu)) \right\}, \\
\frac{\partial r(c \mid \nu)}{\partial \nu} &= \frac{1}{2} r(c \mid \nu) \left\{ \left(\frac{\nu+1}{\nu}\right) \left(1 + \frac{c^2}{\nu}\right)^{-1} \frac{c^2}{\nu} - \log\left(1 + \frac{c^2}{\nu}\right) + b_{01}(c \mid \nu) \right. \\
&\quad \left. - \frac{1}{\nu} (1 - c r(c \mid \nu)) \right\}, \\
\frac{\partial R(c \mid \nu)}{\partial \nu} &= \frac{1}{2} R(c \mid \nu)(1 + R(c \mid \nu)) \left\{ \psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) \right. \\
&\quad \left. - b_{01}(c \mid \nu) - \frac{1}{\nu} c\, r(c \mid \nu) \right\}, \\
\frac{\partial b_{01}(c \mid \nu)}{\partial \nu} &= \frac{1}{2} \left\{ b_{01}^2(c \mid \nu) - b_{02}(c \mid \nu) - \frac{1}{\nu} (1 - c r(c \mid \nu)) \left(b_{01}(c \mid \nu) + \frac{2}{\nu+1} \right) \right. \\
&\quad \left. + \frac{1}{\nu+2} \frac{T(c_2 \mid \nu+2)}{T(c \mid \nu)} b_{21}(c_2 \mid \nu+2) \right\}.
\end{aligned}$$

This last partial derivative was obtained using again the Lemmas 1–4 of the "Appendix 1". Thus, after some extensive algebra, we obtain from (16)–(17) that

$$S_{\gamma\gamma} = -\sum_{i=1}^{n} \left\{ r(-c_i \mid \nu) - \left(\frac{\nu+1}{\nu}\right) c_i \left(1 + \frac{c_i^2}{\nu}\right) \right\} r(-c_i \mid \nu)(1 - d_i) \mathbf{x}_i \mathbf{x}_i^T$$

$$-\left(\frac{\nu+1}{\nu}\right)\sum_{i=1}^{n}\left\{\left(1+\frac{z_i^2}{\nu}\right)^{-1}-\frac{2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-2}z_i^2\right\}d_i\mathbf{x}_i\mathbf{x}_i^T,$$

$$S_{\gamma\tau}=\left(\frac{\nu+1}{\nu}\right)\sum_{i=1}^{n}\left\{\left(1+\frac{z_i^2}{\nu}\right)^{-1}-\frac{2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-2}z_i^2\right\}d_i y_i\mathbf{x}_i,$$

$$S_{\gamma\nu}=-\frac{1}{2}\sum_{i=1}^{n}\left\{\left(\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}\right)\right)(1+R(c_i\mid\nu))\right.$$
$$-\frac{1}{\nu}(1+c_iR(c_i\mid\nu)r(c_i\mid\nu))-R(c_i\mid\nu)b_{01}(c_i\mid\nu)$$
$$\left.-\log\left(1+\frac{c_i^2}{\nu}\right)+\left(\frac{\nu+1}{\nu}\right)\left(1+\frac{c_i^2}{\nu}\right)^{-1}\frac{c_i^2}{\nu}\right\}r(-c_i\mid\nu)(1-d_i)\mathbf{x}_i$$
$$-\frac{1}{\nu^2}\sum_{i=1}^{n}\left\{\left(1+\frac{z_i^2}{\nu}\right)^{-1}z_i-\left(\frac{\nu+1}{\nu}\right)\left(1+\frac{z_i^2}{\nu}\right)^{-2}z_i^3\right\}d_i\mathbf{x}_i,$$

$$S_{\tau\tau}=-\sum_{i=1}^{n}\left\{\frac{1}{\tau^2}+\left(\frac{\nu+1}{\nu}\right)\left[\left(1+\frac{z_i^2}{\nu}\right)^{-1}-\frac{2}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-2}z_i^2\right]y_i^2\right\}d_i,$$

$$S_{\tau\nu}=\frac{1}{\nu^2}\sum_{i=1}^{n}\left\{\left(1+\frac{z_i^2}{\nu}\right)^{-1}z_i-\left(\frac{\nu+1}{\nu}\right)\left(1+\frac{z_i^2}{\nu}\right)^{-2}z_i^3\right\}d_i y_i,$$

$$S_{\nu\nu}=-\frac{1}{4}\sum_{i=1}^{n}\left\{\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}\right)-b_{01}\left(c_i\mid\nu\right)-\frac{1}{\nu}c_ir(c_i\mid\nu)\right\}^2$$
$$\times R\left(c_i\mid\nu\right)\{1+R(c_i\mid\nu)\}(1-d_i)+\frac{1}{4}\sum_{i=1}^{n}\left\{\psi'\left(\frac{\nu}{2}\right)-\psi'\left(\frac{\nu+1}{2}\right)\right.$$
$$-\frac{2}{\nu(\nu+1)}+b_{01}^2(c_i\mid\nu)+\frac{1}{\nu}\left(b_{01}(c_i\mid\nu)-\frac{\nu+3}{\nu(\nu+1)}\right)c_ir(c_i\mid\nu)$$
$$-b_{02}(c_i\mid\nu)+\frac{1}{\nu+2}\frac{T(c_{2i}\mid\nu+2)}{T(c_i\mid\nu)}b_{21}(c_{2i}\mid\nu+2)$$
$$+\frac{1}{\nu}\left[\left(\frac{\nu+1}{\nu}\right)\left(1+\frac{c_i^2}{\nu}\right)^{-1}\frac{c_i^2}{\nu}-\log\left(1+\frac{c_i^2}{\nu}\right)\right.$$
$$\left.\left.+\frac{1}{\nu}c_ir(c_i\mid\nu)\right]c_ir(c_i\mid\nu)\right\}R(c_i\mid\nu)(1-d_i)$$

$$+\frac{1}{4}\sum_{i=1}^{n}\left\{\psi'\left(\frac{\nu+1}{2}\right)-\psi'\left(\frac{\nu}{2}\right)+\frac{2}{\nu^2}\right.$$
$$\left.-\frac{4}{\nu^3}\left(1+\frac{z_i^2}{\nu}\right)^{-1}z_i^2-\frac{2}{\nu}\left(\frac{\nu+1}{\nu}\right)\left(1+\frac{z_i^2}{\nu}\right)^{-2}z_i^2\right\}d_i.$$

In order to obtain the expected information matrix $I(\boldsymbol{\theta})=-E\{\partial^2\log L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T\}$, truncated expectations of the form $E\left\{z^q\left(1+\frac{z^2}{\nu}\right)^{-s}|z+c>0\right\}$, where $z\sim t(0,1,\nu)$ are required. This expectations are obtained from the results given in the "Appendix 1". Also, the truncated expectations b_{01}, b_{02} and b_{21} defined by (15) are needed, which must be computed numerically. Hence, using those results appropriately, after some extensive but straightforward algebraic manipulations it is obtained that $I(\theta)$ has (block-matrix) elements $I_{\alpha\lambda}=-E\{S_{\alpha\lambda}\}$ given by

$$I_{\gamma\gamma}=\sum_{i=1}^{n}\left\{r(-c_i\mid\nu)-\left(\frac{\nu+1}{\nu}\right)\left(1+\frac{c_i^2}{\nu}\right)c_i\right\}t(c_i\mid\nu)\mathbf{x}_i\mathbf{x}_i^T$$
$$+\sum_{i=1}^{n}\left\{\left(\frac{\nu+1}{\nu+3}\right)T(c_{2i}\mid\nu+2)+\left(\frac{2}{\nu+3}\right)c_{2i}\,t(c_{2i}\mid\nu+2)\right\}\mathbf{x}_i\mathbf{x}_i^T,$$
$$I_{\gamma\tau}=-\frac{1}{\tau}\sum_{i=1}^{n}\left\{\left(\frac{\nu+1}{\nu+3}\right)c_i\,T(c_{2i}\mid\nu+2)+\left(\frac{\nu-1}{\nu+3}\right)t(c_i\mid\nu)\right\}\mathbf{x}_i,$$
$$I_{\gamma\nu}=\frac{1}{2}\sum_{i=1}^{n}\left\{\left[\left(\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}\right)\right)(1+R(c_i\mid\nu))\right.\right.$$
$$-\frac{1}{\nu}(1+c_iR(c_i\mid\nu)r(c_i\mid\nu))-R(c_i\mid\nu)b_{01}(c_i\mid\nu)$$
$$\left.-\log\left(1+\frac{c_i^2}{\nu}\right)-\frac{2}{\nu(\nu+1)}\left(\frac{\nu-1}{\nu+3}\right)\right]t(c_i\mid\nu)$$
$$\left.+\frac{1}{\nu}\left(\frac{\nu+5}{\nu+3}\right)\left(\frac{\nu}{\nu+2}\right)^{1/2}c_{2i}^2\,t(c_{2i}\mid\nu+2)\right\}\mathbf{x}_i,$$
$$I_{\tau\tau}=\frac{1}{\tau^2}\sum_{i=1}^{n}\left\{\left(\frac{5\nu+9}{\nu+3}\right)T(c_i\mid\nu)+\left(\frac{\nu}{\nu+2}\right)\left(\frac{\nu+1}{\nu+3}\right)c_{2i}^2\,T(c_{2i}\mid\nu+2)\right.$$
$$\left.+\left(\frac{\nu+1}{\nu+3}\right)c_i\,t(c_i\mid\nu)\right\},$$
$$I_{\tau\nu}=-\frac{1}{\tau}\frac{1}{\nu+3}\sum_{i=1}^{n}\left\{\left(\frac{2}{\nu+1}\right)T(c_i\mid\nu)+\left(\frac{2}{\nu+2}\right)c_{2i}^3\,t(c_{2i}\mid\nu+2)\right.$$
$$\left.-\frac{1}{\nu}\left(\frac{3\nu-1}{\nu+1}\right)c_i\,t(c_i\mid\nu)\right\},$$

$$
\begin{aligned}
I_{\nu\nu} = \frac{1}{4}\sum_{i=1}^{n}\left\{\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - b_{01}\left(c_i \mid \nu\right) - \frac{1}{\nu}c_i r(c_i \mid \nu)\right\}^2 R\left(c_i \mid \nu\right) \\
-\frac{1}{4}\sum_{i=1}^{n}\left\{\left[b_{01}^2(c_i \mid \nu) - b_{02}(c_i \mid \nu) - \frac{1}{\nu}b_{01}(c_i \mid \nu)\left(1 - 2c_i r(c_i \mid \nu)\right)\right] T(c_i \mid \nu)\right. \\
+\left[\frac{1}{\nu+2}b_{21}(c_{2i} \mid \nu+2) - \frac{2}{\nu+3}\left(1 - c_{2i}\, r(c_{2i} \mid \nu+2)\right)\right] T(c_{2i} \mid \nu+2) \\
+\frac{1}{\nu(\nu+2)}c_{2i}^3\, t(c_{2i} \mid \nu) + \frac{1}{\nu}\left[-\log\left(1+\frac{c_i^2}{\nu}\right) + \frac{1}{\nu}\left(1 - c_i\, r(c_i \mid \nu)\right)\right] \\
\left. c_i\, t(c_i \mid \nu)\right\}.
\end{aligned}
$$

To recover the expected information matrix $J(\boldsymbol{\theta})$ of the original parametrization $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \nu)^T$, recall that $J(\boldsymbol{\theta}) = (\partial\boldsymbol{\varphi}/\partial\boldsymbol{\theta})^T I(\boldsymbol{\varphi})(\partial\boldsymbol{\varphi}/\partial\boldsymbol{\theta})$, where the Jacobian matrix is

$$
\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\psi}} = \begin{pmatrix} \sigma I_k & -\frac{1}{2}\sigma^3\beta & 0 \\ 0 & -\frac{1}{2\sigma^3} & 0 \\ 0 & 0 & 1 \end{pmatrix}.
$$

Finally, note that if $\nu \to \infty$, then

$$
I(\boldsymbol{\varphi}) \to \begin{pmatrix} \sum_{i=1}^n \left\{[r(-c_i) - c_i]\,\phi(c_i) + \Phi(c_i)\right\}\mathbf{x}_i\mathbf{x}_i^T & -\frac{1}{\tau}\sum_{i=1}^n \left\{c_i\Phi(c_i) + \phi(c_i)\right\}\mathbf{x}_i & 0 \\ -\frac{1}{\tau}\sum_{i=1}^n \left\{c_i\Phi(c_i) + \phi(c_i)\right\}\mathbf{x}_i^T & \frac{1}{\tau^2}\sum_{i=1}^n \left\{(1+c_i^2)\Phi(c_i) + c_i\phi(c_i)\right\} & 0 \\ 0 & 0 & 0 \end{pmatrix}.
$$

All the above expressions have been checked against the outcome of numerical differentiation/integration, which has confirmed that they are correct.

3 Parameter estimation via an EM-type algorithm

The EM algorithm Dempster et al. (1977) is a very popular iterative optimization strategy commonly used to obtain maximum likelihood estimates when the model has missing data. The incomplete data is often referred to as a latent variable or unobservable data. In censored regression models, the consideration of this approach is clearly justified since the original responses $Y_1, \ldots, Y_n$ are latent variables, where only those responses that were not censored are fully observed.

In this work, we propose the Student-t censored regression model. Consequently, there is a second reason to consider the EM algorithm to obtain the maximum likelihood estimates of this model. In fact, as is well-known, the Student-t distribution can be represented as a scale-mixture of normal distributions. This mean that if $Y_1, \ldots, Y_n$ are independent random variables, with $Y_i \sim t(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2, \nu)$, $i = 1, \ldots, n$, then there are random variables $V_1, \ldots, V_n$, which in our case are unobservable, such that

$$Y_i \mid V_i = v_i \overset{ind.}{\sim} \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, v_i^{-1}\sigma^2), \tag{21}$$

$$V_i \overset{iid}{\sim} \mathcal{G}a(\nu/2, \nu/2), \tag{22}$$

$i = 1, \ldots, n$, where $\mathcal{G}a(\alpha, \lambda)$ denotes the gamma distribution with shape and scale parameters α and λ, respectively. This representation is commonly used in the statistical modelling of the Student-t distribution in both, classical; see Liu and Rubin (1995), and Bayesian; see Lin et al. (2004), approaches.

Considering the hierarchical representation (21)–(22) for the complete Student-t regression model and the observed data $(x_i, d_i y_i)$, $i = 1, \ldots, n$, we implement next the EM algorithm to find the maximum likelihood estimators. The complete-data is given by (x_i, d_i, Y_i, V_i), $i = 1, \ldots, n$. Therefore, the complete-data log-likelihood function is

$$\log L_c(\boldsymbol{\theta}) = -\frac{1}{2} n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} V_i (Y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2$$
$$+ \frac{n\nu}{2} \log \frac{\nu}{2} - n \log \Gamma\left(\frac{\nu}{2}\right) + \frac{\nu}{2} \sum_{i=1}^{n} (\log V_i - V_i).$$

Let be $\widehat{\boldsymbol{\theta}}^{(t)} = (\boldsymbol{\beta}^{(t)T}, \sigma^{2(t)}, \nu^{(t)})^T$ the estimate of $\boldsymbol{\theta}$ at the t-th iteration, and denote by $Q\left(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(t)}\right)$ the conditional expectation of $\log L_c(\boldsymbol{\theta})$ given the observed data and $\widehat{\boldsymbol{\theta}}^{(t)}$. With these notations, we have

$$Q\left(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(t)}\right) = -\frac{1}{2} n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (1 - d_i) E\left(V_i (Y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 \mid Y_i \le 0, \widehat{\boldsymbol{\theta}}^{(t)}\right)$$
$$- \frac{1}{2\sigma^2} \sum_{i=1}^{n} d_i E\left(V_i \mid Y_i > 0, \widehat{\boldsymbol{\theta}}^{(t)}\right) (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 + n \frac{\nu}{2} \log \frac{\nu}{2}$$
$$- n \log \Gamma\left(\frac{\nu}{2}\right) + \frac{\nu}{2} \sum_{i=1}^{n} \left\{ (1 - d_i) E\left[(\log V_i - V_i) \mid Y_i \le 0, \widehat{\boldsymbol{\theta}}^{(t)}\right] \right.$$
$$\left. + d_i E[(\log V_i - V_i) \mid Y_i > 0, \widehat{\boldsymbol{\theta}}^{(t)}] \right\}. \tag{23}$$

To compute the left and right truncated expectations involved in (23), the auxiliary results given in the Lemmas 5–7 of the "Appendix 2" are needed. By using those lemmas, the following truncated expectations are obtained at the iteration t:

$$E\left(V_i \mid Y_i > 0, \widehat{\boldsymbol{\theta}}^{(t)}\right) = \frac{T\left(\widehat{c}_{2i}^{(t)} \mid \widehat{\nu}^{(t)} + 2\right)}{T\left(\widehat{c}_i^{(t)} \mid \widehat{\nu}^{(t)}\right)}, \tag{24}$$

$$E\left(\log V_i \mid Y_i > 0, \widehat{\boldsymbol{\theta}}^{(t)}\right) = \psi\left(\frac{\widehat{\nu}^{(t)} + 1}{2}\right) - \log\left(\frac{\widehat{\nu}^{(t)}}{2}\right) - b_{01}\left(\widehat{c}_i^{(t)}\right), \tag{25}$$

and

$$E\left(V_i \mid Y_i \le 0, \widehat{\boldsymbol{\theta}}^{(t)}\right) = \frac{T\left(-\widehat{c}_{2i}^{(t)} \mid \widehat{\nu}^{(t)} + 2\right)}{T\left(-\widehat{c}_i^{(t)} \mid \widehat{\nu}^{(t)}\right)}, \tag{26}$$

$$E\left(\log V_i \mid Y_i \le 0, \widehat{\boldsymbol{\theta}}^{(t)}\right) = \psi\left(\frac{\widehat{\nu}^{(t)} + 1}{2}\right) - \log\left(\frac{\widehat{\nu}^{(t)}}{2}\right) - b_{01}\left(-\widehat{c}_i^{(t)}\right), \tag{27}$$

$$E\left(V_i Y_i \mid Y_i \le 0, \widehat{\boldsymbol{\theta}}^{(t)}\right) = \widehat{\sigma}^{(t)}\widehat{c}_i^{(t)} \frac{T\left(-\widehat{c}_{2i}^{(t)} \mid \widehat{\nu}^{(t)} + 2\right)}{T\left(-\widehat{c}_i^{(t)} \mid \widehat{\nu}^{(t)}\right)} - \sigma^{(t)} r\left(-\widehat{c}_i^{(t)} \mid \widehat{\nu}^{(t)}\right), \tag{28}$$

$$E\left(V_i Y_i^2 \mid Y_i \le 0, \widehat{\boldsymbol{\theta}}^{(t)}\right) = \widehat{\sigma}^{2(t)} + \left(\widehat{\sigma}^{(t)}\widehat{c}_i^{(t)}\right)^2 \frac{T\left(-\widehat{c}_{2i}^{(t)} \mid \widehat{\nu}^{(t)} + 2\right)}{T\left(-\widehat{c}_i^{(t)} \mid \widehat{\nu}^{(t)}\right)} - \widehat{\sigma}^{2(t)}\widehat{c}_i^{(t)} r\left(-\widehat{c}_i^{(t)} \mid \widehat{\nu}^{(t)}\right), \tag{29}$$

where $\widehat{c}_i^{(t)} = \mathbf{x}_i^T \boldsymbol{\beta}^{(t)}/\sigma^{(t)}$ and $c_{2i}^{(t)} = \sqrt{(\widehat{\nu}^{(t)} + 2)/\widehat{\nu}^{(t)}}\,\widehat{c}_i^{(t)}$. Thus, from (24)–(29), (23) can be rewritten as

$$\begin{aligned} Q\left(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(t)}\right) = &-\frac{1}{2} n \log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\Big\{n_0\widehat{\sigma}^{2(t)} + \boldsymbol{\beta}^T \mathbf{X}^T \left(\boldsymbol{I}_n - \mathbf{D}\right)\widehat{\mathbf{D}}_{0T}^{(t)}\mathbf{X}\boldsymbol{\beta} \\ &+ \left(\widehat{\boldsymbol{\beta}}^{(t)} - 2\boldsymbol{\beta}\right)^T \mathbf{X}^T \left(\boldsymbol{I}_n - \mathbf{D}\right)\widehat{\mathbf{D}}_{0T}^{(t)}\left(\mathbf{X}\widehat{\boldsymbol{\beta}}^{(t)} - \widehat{\boldsymbol{\eta}}_\nu^{(t)}\right) \\ &+ \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)^T \mathbf{D}\widehat{\mathbf{D}}_{1T}^{(t)}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\Big\} + n\left\{\frac{\nu}{2}\log\left(\frac{\nu}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right)\right\} \\ &+ \frac{\nu}{2}\Bigg\{n\psi\left(\frac{\widehat{\nu}^{(t)} + 1}{2}\right) - n\log\left(\frac{\widehat{\nu}^{(t)}}{2}\right) \\ &- \mathbf{1}_n^T\left[\left(\boldsymbol{I}_n - \mathbf{D}\right)\widehat{\mathbf{b}}_0^{(t)} + \mathbf{D}\widehat{\mathbf{b}}_1^{(t)}\right] - \mathbf{1}_n^T\left(\widehat{\mathbf{D}}_{0T}^{(t)} + \widehat{\mathbf{D}}_{1T}^{(t)}\right)\mathbf{1}_n\Bigg\}, \end{aligned} \tag{30}$$

where as before $\mathbf{X} = (\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T)^T$, $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$, $\mathbf{1}_n$ is a n-dimensional vector of ones, and

$$\begin{aligned} \widehat{\boldsymbol{\eta}}_\nu^{(t)} &= (\widehat{\sigma}^{(t)} r(\widehat{c}_1^{(t)} \mid \nu^{(t)}), \ldots, \widehat{\sigma}^{(t)} r(\widehat{c}_n^{(t)} \mid \nu^{(t)}))^T, \\ \widehat{\mathbf{b}}_0^{(t)} &= \left(b_{01}\left(-\widehat{c}_1^{(t)} \mid \widehat{\nu}^{(t)}\right), \ldots, b_{01}\left(-\widehat{c}_n^{(t)} \mid \widehat{\nu}^{(t)}\right)\right)^T, \\ \widehat{\mathbf{b}}_1^{(t)} &= \left(b_{01}\left(\widehat{c}_1^{(t)} \mid \widehat{\nu}^{(t)}\right), \ldots, b_{01}\left(\widehat{c}_n^{(t)} \mid \widehat{\nu}^{(t)}\right)\right)^T, \end{aligned}$$

$$\widehat{\mathbf{D}}^{(t)}_{T0} = \mathrm{diag}\left(\frac{T\left(-\widehat{c}^{(t)}_{21} \mid \widehat{\nu}^{(t)}+2\right)}{T\left(-\widehat{c}^{(t)}_{1} \mid \widehat{\nu}^{(t)}\right)},\ldots,\frac{T\left(-\widehat{c}^{(t)}_{2n} \mid \widehat{\nu}^{(t)}+2\right)}{T\left(-\widehat{c}^{(t)}_{n} \mid \widehat{\nu}^{(t)}\right)}\right),$$

$$\widehat{\mathbf{D}}^{(t)}_{T1} = \mathrm{diag}\left(\frac{T\left(\widehat{c}^{(t)}_{21} \mid \widehat{\nu}^{(t)}+2\right)}{T\left(\widehat{c}^{(t)}_{1} \mid \widehat{\nu}^{(t)}\right)},\ldots,\frac{T\left(\widehat{c}^{(t)}_{2n} \mid \widehat{\nu}^{(t)}+2\right)}{T\left(\widehat{c}^{(t)}_{n} \mid \widehat{\nu}^{(t)}\right)}\right).$$

The above results show that for the Student-t censored regression model, computationally attractive expressions that can be easily implemented have been found. Thus, the EM algorithm can be implemented via a simple modification, called ECM algorithm; see Meng and Rubin (1993), Liu and Rubin (1995). A key feature of this algorithm is that it preserves the stability of the EM algorithm, namely its monotone convergence property. The steps of the ECM algorithm are presented as follows.

E-step: Given $\widehat{\boldsymbol{\theta}}^{(t)}$, compute the conditional expectations in (24)–(25), for each $i \in N_0$, and those in (26)–(29), for each $i \in N_1$.

CM-step 1: By maximization of (30) over $\boldsymbol{\beta}$ and σ^2, update $\widehat{\boldsymbol{\beta}}^{(t+1)}$ and $\widehat{\sigma}^{2(t+1)}$ using the following equations:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(t+1)} &= \left[\mathbf{X}^T\left((\boldsymbol{I}_n-\mathbf{D})\widehat{\mathbf{D}}^{(t)}_{T0}+\mathbf{D}\widehat{\mathbf{D}}^{(t)}_{T1}\right)\mathbf{X}\right]^{-1}\left[\mathbf{X}^T(\boldsymbol{I}_n-\mathbf{D})\widehat{\mathbf{D}}^{(t)}_{T0}\left(\mathbf{X}\widehat{\boldsymbol{\beta}}^{(t)}-\widehat{\boldsymbol{\eta}}^{(t)}_{\nu}\right)\right.\\
&\quad\left.+\mathbf{X}^T\mathbf{D}\widehat{\mathbf{D}}^{(t)}_{T1}\mathbf{y}\right],\\
\widehat{\sigma}^{2(t+1)} &= \frac{1}{n}\left\{n_0\widehat{\sigma}^{2(t)}+\widehat{\boldsymbol{\beta}}^{(t+1)T}\mathbf{X}^T\widehat{\mathbf{D}}^{(t)}_{T0}(\boldsymbol{I}_n-\mathbf{D})\mathbf{X}\widehat{\boldsymbol{\beta}}^{(t+1)}+\left(\widehat{\boldsymbol{\beta}}^{(t)}-2\widehat{\boldsymbol{\beta}}^{(t+1)}\right)^T\mathbf{X}^T\right.\\
&\quad\left.\times(\boldsymbol{I}_n-\mathbf{D})\widehat{\mathbf{D}}^{(t)}_{T0}\left(\mathbf{X}\widehat{\boldsymbol{\beta}}^{(t)}-\widehat{\boldsymbol{\eta}}^{(t)}_{\nu}\right)+\left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}^{(t+1)}\right)^T\widehat{\mathbf{D}}^{(t)}_{1T}\mathbf{D}\left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}^{(t+1)}\right)\right\}.
\end{aligned}$$

CM-step 2: Given $\widehat{\boldsymbol{\beta}}^{(t+1)}$ and $\widehat{\sigma}^{2(t+1)}$ obtained in the previous step, update $\widehat{\nu}^{(t+1)}$ by maximization of

$$\begin{aligned}
\psi\left(\frac{\widehat{\nu}^{(t+1)}}{2}\right)-\log\left(\frac{\widehat{\nu}^{(t+1)}}{2}\right) &= 1+\psi\left(\frac{\widehat{\nu}^{(t)}+1}{2}\right)-\log\left(\frac{\widehat{\nu}^{(t)}}{2}\right)\\
&\quad-\frac{1}{n}\mathbf{1}^T_n\left[(\boldsymbol{I}_n-\mathbf{D})\widehat{\mathbf{b}}^{(t+1)}_0+\mathbf{D}\widehat{\mathbf{b}}^{(t+1)}_1\right]\\
&\quad-\frac{1}{n}\mathbf{1}^T_n\left((\boldsymbol{I}_n-\mathbf{D})\widehat{\mathbf{D}}^{(t+1)}_{T0}+\mathbf{D}\widehat{\mathbf{D}}^{(t+1)}_1\right)\mathbf{1}_n,
\end{aligned}$$

over ν.

The E and CM steps are alternated repeatedly until a suitable convergence rule is satisfied, e.g., the difference in successive values of the estimates, namely $||\boldsymbol{\theta}^{(t+1)} -\boldsymbol{\theta}^{(t)}||$, is less than a tolerance value where $||\mathbf{z}|| = \sqrt{\mathbf{z}^T\mathbf{z}}$. The ECM algorithm is often criticized because it tends to get stuck at local modes. A convenient way to avoid this

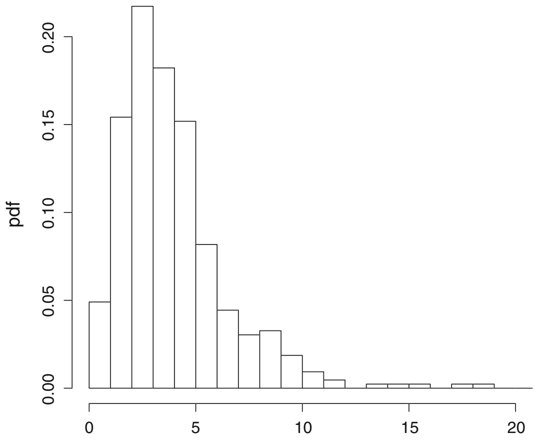

Fig. 1 Housewives wage rates

problem is to try several ECM iterations with a variety of starting values. If there exist several modes, one can find the global mode by comparing their relative masses and log-likelihood values.

Finally, note that the ECM algorithm for the Tobit model to estimate $\widehat{\boldsymbol{\beta}}^{(t+1)}$ and $\widehat{\sigma}^{2(t+1)}$ follows from letting $\widehat{\boldsymbol{\eta}}_\nu = \widehat{\boldsymbol{\eta}}_0$ and $\widehat{\mathbf{D}}_{T0} = \widehat{\mathbf{D}}_{T1} = \boldsymbol{I}_n$ in the above relations, which correspond justly to the limit case of $\nu \to \infty$.

4 Application

The developed method is illustrated with data from Mroz (1987), the University of Michigan Panel Study of Income Dynamics (PSID) for the year 1975 (interview year 1976). This year was particularly special since the PSID interviewed directly the wives in the households (during the other years, the head of the household's interview supplied information about the wife's labor market experiences during the previous year). The data consists of 753 married white women between the ages of 30 and 60 in 1975, with 428 women that worked at some point during that year. The dependent variable used in this application is a measure of the wage of the housewife known as the average hourly earnings (wage rates). It is important to stress that wage rates are set equal to zero (i.e. they are censored or simply not observed) for wives who did not work in 1975. This assumption is usually adopted in economy; see DaVanzo and Lee (1978). Since the 43.16 % of the housewives respondents had no work at the moment of the interviews, the data set presents a very high degree of censoring. For the rest of the housewives, the density function is sketched by the histogram in Fig. 1.

The data are analyzed using the Student-t censored regression model of the housewife wage rate, and a set of control variables such as the wife's age, her years of schooling, the number of children younger than six years old in the household and the number of children between the ages of six and nineteen.

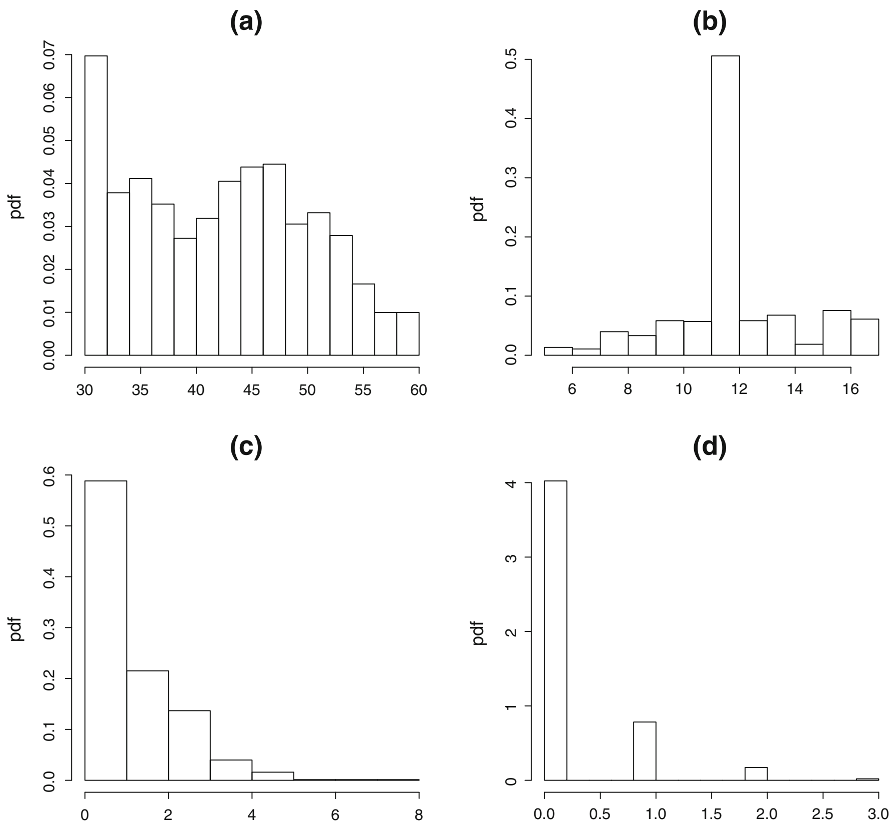

Fig. 2 Control variables: **a** age, **b** education, **c** children younger than 6 years old, **d** children between the ages of five and nineteen

In order to support the election of the Student-t censored model, we analyzed the martingale-type residuals (MT) proposed by Barros et al. (2010) for censored models. These residuals are defined by

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

where $r_{M_i} = d_i + \log(S(y_i; \widehat{\boldsymbol{\theta}}))$ is the martingale residual defined by Therneau et al. (1990), with $S(y_i; \widehat{\boldsymbol{\theta}})$ the survival function of $\mathbf{y}$ evaluated at the ML estimator of $\boldsymbol{\theta}$, $d_i = 0, 1$ indicating whether the observation is censored or not respectively and $sign(z)$ denoting the sign of z. The plots of the MT residuals with generated confidence envelopes,[1] presented in Fig. 3, confirm that the Student-t assumption is well supported by the data.

[1] The confidence envelopes are a graphical model-checking device for linear models based on point-wise confidence bounds at certain confidence level (for more details about confidence envelopes see Venables and Ripley 2000 and references therein).

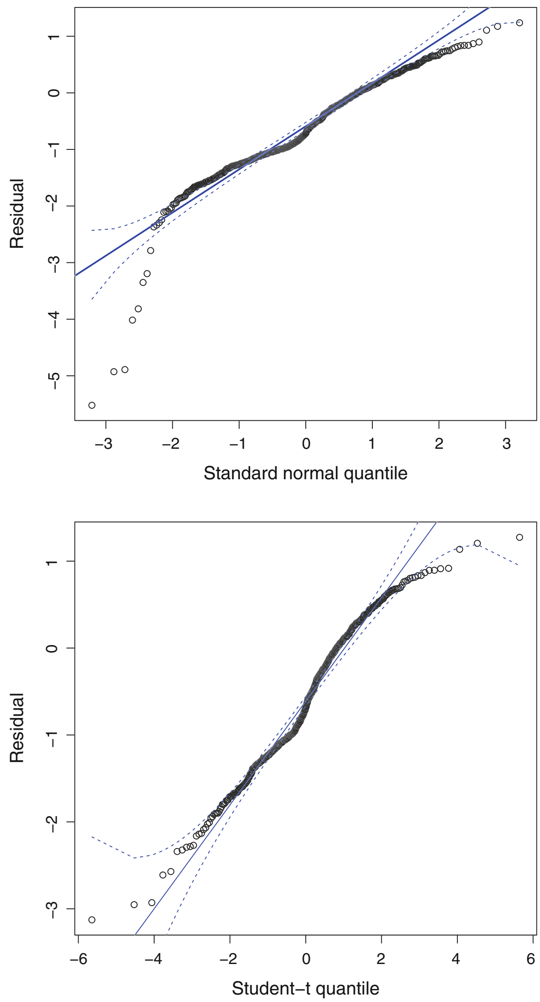

Fig. 3 QQ plot with envelopes of MT residuals for wage rates

The obtained results are compared to the estimations of Tobit model. Table 1 contains the ML estimates of the parameters from both models, together with their corresponding standard errors calculated using the observed information matrix given in Sect. 2.3. The AIC model selection criteria indicate that the Student-t model, with heavy tails, presents a better fit than the Tobit model.

5 Final conclusion

In this work, an extension of the normal censored regression model, known in the econometrical and statistical literature as the Tobit model, has been developed by considering the case where the error terms are independent and have a Student-t

Table 1 PSID 1975 data

Parameter	Student-t model		Tobit model	
	Estimate	SE	Estimate	SE
Constant	−3.27	1.47e-02	−2.75	1.89e-02
Age	−0.11	3.36e-03	−0.10	3.32e-03
Education	0.81	6.12e-04	0.73	6.22e-04
Kids $\leq$ 6 years old	−0.23	2.35e-02	−0.21	2.43e-02
Kids $\geq$ 6 years old	−3.28	1.04e-e03	−3.03	1.01e-03
σ	4.44	5.16e-08	4.57	2.86e-08
ν	6.61	3.23e-01	–	–
Log likelihood	−1,459.26		−1,481.66	
AIC	2,932.52		2,975.31	

distribution. A convenient EM-type algorithm is developed by exploring the statistical properties of the Student-t truncated distribution. In this context, attractive expressions that can be easily implemented are obtained. The observed and expected matrices are analytically derived, allowing for the direct implementation of the inference on this type of models. The Student-t censored regression model with heavy tails seems to be more appropriate to fit the Mroz (1987) data set. `R` programs are available from the authors upon request. Although the methodology proposed considers the case when ν is unknown, some computational difficulties arise (for example, slow convergence). For that reason, a more efficient ECM algorithm; see Meng and Rubin (1993), Liu and Rubin (1995), should be used. Finally, it is important to stress that the methodology proposed in this paper can be extended to other types of mixture distributions, leaving a future topic for research.

Acknowledgments The authors thank to one anonymous referee whose constructive comments led to an improved presentation of the manuscript. The research of Arellano-Valle was partially supported by grant FONDECYT 1085241 and 1120121 from the Chilean government. The research of Castro was partially supported by grant FONDECYT 11100076 from the Chilean government. The research of González-Farías was partially supported by CONACYT Ciencia Básica No. 105657 and ECO2010-19357. The research of Muñoz-Gajardo was partially supported by grant FONDECYT 1085241 from the Chilean government and by DIPUC and VRAID from Pontificia Universidad Católica de Chile.

Appendix 1

In this appendix, expressions for truncated expectations of some functions of a Student-t random variable are provided. They are used in particular to compute the information matrix.

Lemma 1 *Let $Z \sim t(0, 1, \nu)$. Then for any integrable function g:*

$$E\{g(Z) \mid Z < c\} = E\{g(-Z) \mid Z + c > 0\}.$$

Lemma 2 *Let $t(z \mid \nu)$ be the $t(0, 1, \nu)$-pdf. Then:*

$$\left(1 + \frac{z^2}{\nu}\right)^{-m/2} t(z \mid \nu) = \frac{t(0 \mid \nu)}{t(0 \mid \nu + m)} t\left(\sqrt{\frac{\nu + m}{\nu}}\, z \mid \nu + m\right).$$

Lemma 3 *Let $m_k(c \mid \nu) = E\{Z^k \mid Z + c > 0\}$, $\nu > k$, where $Z \sim t(0, 1, \nu)$. Then, for $k = 1, 2, 3, 4$ it follows that:*

$$\begin{aligned}
m_1(c \mid \nu) &= \frac{\nu}{\nu - 1}\left(1 + \frac{c^2}{\nu}\right) r(c \mid \nu), \quad \nu > 1,\\
m_2(c \mid \nu) &= \frac{\nu}{\nu - 2}\frac{T(c_{-2} \mid \nu - 2)}{T(c \mid \nu)} - c\, m_1(c \mid \nu), \quad \nu > 2,\\
m_3(c \mid \nu) &= \frac{2\nu^2}{(\nu - 1)(\nu - 3)}\left(1 + \frac{c^2}{\nu}\right)^2 r(c \mid \nu) + c^2 m_1(c \mid \nu), \quad \nu > 3,\\
m_4(c \mid \nu) &= \frac{3\nu^2}{(\nu - 2)(\nu - 4)}\frac{T(c_{-4} \mid \nu - 4)}{T(c \mid \nu)} - \frac{3}{2} c\, m_3(c \mid \nu)\\
&\quad + \frac{1}{2} c^3 m_1(c \mid \nu), \quad \nu > 4,
\end{aligned}$$

where $r(c \mid \nu) = \frac{t(c|\nu)}{T(c|\nu)}$ and $c_{-k} = \sqrt{\frac{\nu-k}{\nu}}\, c$, $\nu > k$.

Lemma 4 *Let $Z_m \sim t(0, 1, \nu + m)$, with $Z_0 = Z$. Then:*

$$\begin{aligned}
E\left\{\left(1 + \frac{Z^2}{\nu}\right)^{-1} \mid Z + c > 0\right\} &= \left(\frac{\nu}{\nu + 1}\right)\frac{T(c_2 \mid \nu + 2)}{T(c \mid \nu)},\\
E\left\{\left(1 + \frac{Z^2}{\nu}\right)^{-1} Z \mid Z + c > 0\right\} &= \left(\frac{\nu}{\nu + 1}\right) r(c \mid \nu),\\
E\left\{\left(1 + \frac{Z^2}{\nu}\right)^{-1} Z^2 \mid Z + c > 0\right\} &= \left(\frac{\nu}{\nu + 1}\right)\{1 - c\, r(c \mid \nu)\},\\
E\left\{\left(1 + \frac{Z^2}{\nu}\right)^{-2} Z^2 \mid Z + c > 0\right\} &= \left(\frac{\nu}{\nu + 1}\right)\left(\frac{\nu}{\nu + 3}\right)\frac{T(c_2 \mid \nu + 2)}{T(c \mid \nu)}\\
&\quad \times \{1 - c_2\, r(c_2 \mid \nu + 2)\},\\
E\left\{\left(1 + \frac{Z^2}{\nu}\right)^{-2} Z^3 \mid Z + c > 0\right\} &= \left(\frac{\nu}{\nu + 1}\right)\left(\frac{\nu}{\nu + 3}\right)\Bigg\{2\, r(c \mid \nu)\\
&\quad + \left(\frac{\nu}{\nu + 2}\right)^{1/2}\frac{T(c_2 \mid \nu + 2)}{T(c \mid \nu)} c_2^2\, r(c_2 \mid \nu + 2)\Bigg\},
\end{aligned}$$

$$E\left\{\left(1+\frac{Z^2}{\nu}\right)^{-2} Z^4 \mid Z+c>0\right\} = \left(\frac{\nu}{\nu+1}\right)\left(\frac{\nu}{\nu+3}\right)\{3(1-c\,r(c\mid\nu)) - \left(\frac{\nu}{\nu+2}\right)\frac{T(c_2\mid\nu+2)}{T(c\mid\nu)}c_2^3\, r(c_2\mid\nu+2)\},$$

where as before $c_k = \sqrt{\frac{\nu+k}{\nu}}\, c$.

The proofs of Lemma 1 and 2 are straightforward. For the proof of Lemma 3 see Arellano-Valle and Genton (2008). Finally, the proof of Lemma 4 is direct from Lemma 2 and 3.

Appendix 2

Observe here some auxiliary results to compute the conditional moments used in the EM algorithm.

Lemma 5 *Let* $Y \sim \mathcal{N}(\mu, \sigma^2)$. *Then,*

$$E(Y \mid Y \le a) = \mu - \sigma r(\bar{a}),$$
$$E(Y^2 \mid Y \le a) = \mu^2 - 2\sigma\mu r(\bar{a}) + \sigma^2\{1 - \bar{a} r(\bar{a})\},$$

where $\bar{a} = (a-\mu)/\sigma$ *and* $r(z) = \phi(z)/\Phi(z)$.

Lemma 6 *Let* $W \sim \mathcal{G}a(\alpha, \lambda)$. *Then,* $E(\log W) = \psi(\alpha) - \log\lambda$.

Lemma 7 *Let* $Y \mid V = v \sim \mathcal{N}(\mu, v^{-1}\sigma^2)$ *and* $V \sim \mathcal{G}a(\nu/2, \nu/2)$. *Then, the conditional pdf of* V *given* $[Y \le a]$ *is*

$$f_{V\mid Y\le a}(v) = \frac{(\nu/2)^{\nu/2}\, v^{(\nu+1)/2-1} e^{-\frac{\nu}{2}v}}{\Gamma(\nu/2)} \frac{\Phi\left(\sqrt{v}\,\bar{a}\right)}{T(\bar{a}\mid\nu)}, \quad v>0,$$

where $\bar{a}$ *is defined above. In particular, we have*

$$E(V \mid Y \le a) = \frac{T(\bar{a}_2 \mid \nu+2)}{T(\bar{a}\mid\nu)},$$
$$E\left\{\sqrt{V}\, r\left(\sqrt{V}\,\bar{a}\right) \mid Y \le a\right\} = r(\bar{a}\mid\nu),$$
$$E(\log V \mid Y \le a) = \psi\left(\frac{\nu+1}{2}\right) - \log\left(\frac{\nu}{2}\right) - b_{01}(\bar{a}),$$

where $\bar{a}_2 = \sqrt{(\nu+2)/\nu}\,\bar{a}$. *Moreover, for any integrable function* g, $E(g(V) \mid Y > a) = E(g(V) \mid Y \le -\bar{a})$.

The results in Lemma 5 and 6 are well-know; see e.g., Johnson et al. (1994) and Wilks (1932), respectively. For a proof of the selection *pdf* of $[V \mid Y \leq a]$ given in Lemma 7, see Arellano-Valle et al. (2002), Arellano-Valle et al. (2006); the truncated expectations presented in this lemma follow after some algebra and considering Lemma 5.

References

Amemiya T (1973) Regression analysis when the dependent variable is truncated normal. Econometrica 41:997–1016

Amemiya T (1984) Tobit models: a survey. Econometrica 24:2–61

Andrews D, Mallows C (1974) Scale mixtures of normal distributions. J R Stat Soc Ser B 36:99–102

Arellano-Valle R, Branco M, Genton M (2006) A unified view on skewed distributions arising from selections. Can J Stat 34:581–601

Arellano-Valle R, Genton M (2008) On the exact distribution of the maximum of absolutely continuous dependent random variables. Stat Probab Lett 78:27–35

Arellano-Valle R, del Pino G, San Martín E (2002) Definition and probabilistic properties of skew-disstributions. Stat Prob Lett 58:111–121

Barros M, Galea M, González M, Leiva V (2010) Influence diagnostics in the Tobit censored response model. Stat Methods Appl 19:379–397

Blundell R, Meghir C (1987) Bivariate alternatives to the Tobit model. J Econ 34:179–200

DaVanzo J, Lee D (1978) The compatibility of child care with labor force participation and non-market activities. Rand Corporation, Santa Monica

Dempster A, Laird N, Rubin D (1977) Maximum likelihood for incomplete data via the EM algorithm. J R Stat Soc Ser B 39:1–22

He X, Cui H, Simpson D (2004) Longitudinal data analysis using t-type regression. J Stat Plan Inf 122:253–269

He X, Simpson D, Wang G (2000) Breakdown points of t-type regression estimators. Biometrika 87:675–687

Hutton J, Stanghellini E (2011) Modelling bounded health scores with censored skew-normal distributions. Stat Med 30(4):368–376

Johnson N, Kotz S, Balakrishnan N (1994) Continuous univariate distributions, Vol 1. Wiley, Amsterdam

Lange K, Little R, Taylor J (1989) Robust statistical modeling using the t distribution. J Am Stat Assoc 84:881–896

Lin T, Lee J, Ni H (2004) Bayesian analysis of mixture modelling using the multivariate t distribution. Stat Comput 14:119–130

Liu C, Rubin D (1995) ML estimation of the t distribution using EM and its extensions, ECM and ECME. Stat Sin 5:19–39

Maddala G (1983) Limited-dependent and qualitative variables in econometrics. Econometric Society Monographs in Quantitative Econometrics. Econometric Society Publication No. 3. Cambridge University Press

Meng X, Rubin D (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika 80:267–278

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

Olsen R (1978) Note on the uniqueness of the maximum likelihood estimator for the Tobit model. Econometrica 46:1211–1215

Powell J (1994) Estimation of semiparametric models. In: Heckman J, Leamer E, (eds) Handboook of econometrics. Elsevier, Amsterdam 6B, pp 5307–5368

Therneau T, Grambsch P, Fleming T (1990) Martingale-based residuals for survival models. Biometrika 77:147–160

Tobin J (1958) Estimation of relationships for limited dependent variables. Econometrica 26:24–36

Venables W, Ripley D (2000) S Programming. Springer, New York

Wilks S (1932) Certain generalizations in the analysis of variance. Biometrika 24:471–494