CENSORED REGRESSION MODELS WITH DOUBLE EXPONENTIAL ERROR DISTRIBUTIONS: AN ITERATIVE ESTIMATION PROCEDURE BASED ON MEDIANS FOR CORRECTING BIAS

Carmen ANIDO and Teófilo VALDÉS

Abstract

In this paper, we consider a simple iterative estimation procedure for censored regression models with symmetrical exponential error distributions. Although each step requires to impute the censored data with conditional medians, its tractability is guaranteed as well as its convergence at geometrical rate. Finally, as the final estimate coincides with a Huber M-estimator, its consistency and asymptotic normality are easily proved.

1   Introduction

In censored regression models, it is well known that least squares estimation methods cannot be directly applied without previously correcting the bias derived from the unobserved dependent values. When a missing data mechanism is assumed, the maximum likelihood estimate is consistent and efficient although, occasionally, its practical implementation may be tedious when the likelihood possesses several critical points. When this is the case, the search for global maximum requires the checking in detail of all the above mentioned critical points and, consequently, it may be preferable to count with alternative estimation procedures which secure fast convergence and good asymptotic properties as happens with the one here proposed. The motor idea has been motivated by the missing information principle of Orchard and Woodbury (1972) and also by earlier algorithms (e.g. Healy and Westmacott (1956)).

They suggest at each step to fill all the missing individual data with micro-imputations based on the current estimates and, later on, to actualise these last ones by least squares applied on both observed and imputed dependent values. When the errors follow a normal distribution and the imputations equal the mean of the truncated distributions it is known that the iterative procedure leads to the maximum likelihood estimation as it coincides with the EM algorithm (Dempster, Laird and Rubin (1977)) whose general convergence properties, behaviour and practical applications may be found, for instance, in Wu (1983), Little and Rubin (1987) and Tanner (1993). When the errors are not normally distributed and we maintain the imputation processes based on the conditional means, the algorithm differs from the EM and, usually, the coincidence between the final estimate (assuming convergence) and the maximum likelihood estimator cannot be assured. Schmee and Hahn (1979), James and Smith (1984), Chatterjee and McLeish (1986) and Ritov (1990) have investigated several aspects of this last situation. Finally, with non normal errors and imputations not based on the conditional expectations, the behaviour of the algorithm is unknown. This is the case considered in this paper. The errors are assumed to be double exponentially distributed (e.g. symmetrised waiting time errors) and the censored data are filled with conditional medians instead of means. The reckoning of the successive steps is very easily implemented and the convergence to an unique point may be assured. This last convergence point will define the final estimate for the regression coefficients and, being a Huber type M-estimator, its consistency and asymptotic normality will be shown. Two general comments must be pointed out at this moment. The first one hinges on a naive fact, namely, that estimates derived from the proposed procedure are distribution-dependent on errors. The selection of Laplace distributions has been motivated on two grounds. On the one hand, this law has historically been considered as a firm candidate (though not unique) to substitute the usual normality assumption and also it is the most preferred under certain special circumstances (e.g., with environment radiological measures due to their exponential decay). On the other hand, the symmetric exponential distribution may be used to easily highlight the capability of the median-based imputation process to correct bias as well as to maintain desirable asymptotic properties of the final estimates before trying to

widen its scope of applicability. The second comment relates to one of the most usual fields of application of censored data models, to wit the analysis of survival data. It is well known with this sort of data that the most common distributions used to describe lifetimes tend to be positively skewed. Authors are conscious that the Laplace distribution does not suit to the mentioned asymmetric assumption, as neither does the normal distribution. At this respect, some additional comments will be postponed till section 6 where the skewness of the direct survival data is diminished by one of the usual monotone transformations resulting in a linear model equally censored. Notwithstanding this, the bias correcting process here proposed has an application ability that lies far beyond survival data scope, e.g. data recorded by means of apparatuses with different detection limits due, for instance, to their varying calibration capabilities: here the symmetric distribution assumption could undoubtedly make sense with no previous transformation of data.

The organisation of this paper is as follows: section 2 describes the model, the missing data mechanism and the estimation procedure. In section 3, we guarantee the convergence (at least, at a linear rate) to a fixed point which is independent on the starting values and, later on, section 4 develops some asymptotic distribution properties of the proposed estimate. In section 5, we include some natural extensions of the former results. Section 6 approaches a quite extensive simulation study that shows the general capabilities of the procedure and compares its results with those derived from other alternatives, all under several scenarios some of which set up against certain assumptions previously stated. Finally, an application with real life survival data has been tackled in section 7.

2   Model and estimation procedure

For $i = 1, \cdots, n$, assume a usual regression model

$$y_i = x_i^T \alpha + \varepsilon_i \tag{1}$$

where $y_i's$ are scalars and $x_i's$ are $d-$dimensional vectors. It will be assumed all along this work that the regression hyperplane may not contain the origin so that $x_{i,1} = 1$ for all i. Errors ϵ_i are $i.i.d$ following a double exponential distribution. Without loss in generality, it may be

assumed that the error density function corresponds to the symmetrical exponential distribution with parameter one, that is

$$f(x) = 2^{-1}\exp(-\mid x\mid)I_{(-\infty,\infty)}(x).$$

Finally, consider that the dependent variable is lower censored being only observed when $y_i > 0$. After reordering the data, we can write $y = (y^{obs}, y^{nobs})^T$ and accordingly the $n \times d$ matrix X, although completely observed, may be symbolically partitioned as $X = (X^{obs}, X^{nobs})^T$. Obviously we seek to estimate the vector parameter α. For this purpose, we suggest the following sequential procedure based on conditional medians. In each iteration, assuming a current value for α and being $X^T X$ of full rank, we first impute the censored y_i values by the medians using the available information; secondly, we actualise the estimator for α by the usual projection process. Formally, the iterative procedure may be written as

Initialization:

Fix a starting value $\alpha^{(0)}$.

Iteration:

Assuming the current estimate $\alpha^{(r)}$ to be known, we must implement the following two steps

M-STEP (Imputation on medians): For all $i \in I = \{1, \cdots, n\}$, do

$$y_i(\alpha^{(r)}) = \begin{cases} y_i, & i \in I^{obs} \\ x_i^T\alpha^{(r)} + \gamma(-x_i^T\alpha^{(r)}), & i \in I^{nobs} \end{cases} \qquad (2)$$

P-STEP (Projection step): Improve the current α−estimator by projecting the vector $y(\alpha^{(r)})$ on the space spanned by the columns of matrix X, that is

$$\alpha^{(r+1)} = (X^T X)^{-1}X^T y(\alpha^{(r)}). \qquad (3)$$

In the former steps $I^{obs} = \{i \mid y_i$ is observed$\}$, $I^{nobs} = I - I^{obs}$ and, for any real z, the function $\gamma(z)$ represents the conditional medians $\gamma(z) = Me(\epsilon \mid \epsilon < z)$ which it may be verified to coincide with

$$\gamma(z) = I_{z \leq 0)}(z)(z - ln2) + I_{(z>0)}(z)ln(1 - 2^{-1}e^{-z}). \qquad (4)$$

It is clear that an accumulation point α^* of any sequence $\{\alpha^{(r)}\}$ generated by the algorithm must be a fixed point of the transformation $T(\alpha) = (X^T X)^{-1} X^T y(\alpha)$. In particular, if $\alpha^{(r)} \to \alpha^*$ necessarily

$$\alpha^* = (X^T X)^{-1} X^T y(\alpha^*). \qquad (5)$$

On the one hand, we will prove in the next section that the former transformation $T(z)$ has an unique fix point and, consequently, any convergent algorithm sequence must tend to it, not depending on the initial value $\alpha^{(0)}$. Secondly, it will also be proved that all sequences generated by the iteration of steps $M - P$ necessarily converges. All as a whole indicates that the unique solution α^* of equation (5) consistently may define our proposed estimator for the vector regression parameter α.

For posterior convenience, observe that the $M-$step admits a peculiar matrix representation. Lets decompose the cross products matrix $X^T X$ as the sum

$$X^T X = (X^T X)^{obs} + (X^T X)^{nobs} \qquad (6)$$

being $(X^T X)^{obs} = (X^{obs})^T X^{obs}$ and lets also assign a similar interpretation to $(X^T X)^{nobs}$. The expression (2) may be written in the form

$$y(\alpha) = (y^{obs}, X^{nobs}\alpha)^T + (0, D(\alpha)1^{nobs})^T \qquad (7)$$

where $D(\alpha) = diag(\gamma(-x_i^T \alpha))$ and $1^{nobs} = (1, \cdots, 1)^T$ both of an order equal to the cardinal of I^{nobs}. Finally, observe that $\gamma(z) = Me(\epsilon \mid \epsilon < z)$ is continuously differentiable and its derivative

$$\frac{d}{dz}\gamma(z) = I_{(z \leq 0)}(z) + I_{(z>0)}(z)(2e^z - 1)^{-1} > 0 \qquad (8)$$

is bounded by one. As a consequence, the integral function

$$\Gamma(z) = \int_0^z \gamma(t)dt$$

is strictly convex on $\mathbb{R}$ and, for any fixed $x_i \in \mathbb{R}^d(x_i \neq 0), \Gamma(-x_i^T \alpha)$ is also strictly convex as a function on α.

3   Convergence properties of the procedure

In order to prove the above mentioned convergence properties of the iterative procedure, lets define the auxiliary function on α

$$H(\alpha) = 2^{-1} \sum_{i \in I^{obs}} (y_i - x_i^T \alpha)^2 + \sum_{i \in I^{nobs}} \Gamma(-x_i^T \alpha). \tag{9}$$

It is obvious that $H(\alpha)$ is strictly convex and, as $H(\alpha) \to \infty$ if $\| \alpha \| \to \infty$, it satisfies the cone condition

$$H(\alpha) \geq \delta \mid \alpha \mid + \rho$$

for some $\delta > 0$ and $\rho \in \mathbb{R}$. As a consequence, the next lemma 1 follows.

Lemma 1. *The function $H(\alpha)$ has an unique minimum α^* which necessarily is the unique fixed point of the transformation $T(z)$ defined by (5).*

Proof. The former cone condition guarantees that function $H(\alpha)$ has an unique critical point, say α^*, for which its gradient vanishes. This means that

$$0 = \frac{\partial}{\partial \alpha} H(\alpha^*) = - \sum_{i \in I^{obs}} x_i(y_i - x_i^T \alpha^*) - \sum_{i \in I^{nobs}} x_i \gamma(-x_i^T \alpha^*)$$

which, after $\pm (X^T X)^{nobs} \alpha^*$, is equivalent to expression (5), which concludes the proof.

  The previous lemma implies that any convergent algorithm sequence has α^* as its limit and also that any bounded sequence generated by the procedure converges to this unique fixed point. The following theorem secures the convergence of any sequence $\{\alpha^{(r)}\}$ whatever the starting value may be, only by assuming the slight hypothesis of being $(X^T X)^{obs}$ positive definite what obviously implies a similar condition for $X^T X$.

Theorem 2.1. *If $(X^T X)^{obs}$ is positive definite, then any sequence generated by the procedure $M - P$ necessarily converges at a linear rate to the unique fixed point α^*.*

Proof. Observe first that from (3)

$$\alpha^{(r+1)} - \alpha^{(r)} = S^{-1} S^{-1} X^T (y(\alpha^{(r)}) - y(\alpha^{(r-1)})) \tag{10}$$

where S represents the symmetric and definite positive square root of $X^T X (S^2 = X^T X)$ which must exist since $X^T X$ is a symmetric and definite positive matrix. Additionally, from (7)

$$y(\alpha^{(r)}) - y(\alpha^{(r-1)}) = (0, X^{nobs}(\alpha^{(r)} - \alpha^{(r-1)})^T + (0, (D(\alpha^{(r)}) - D(\alpha^{(r-1)}))1^{nobs})^T. \tag{11}$$

Define now the diagonal functional matrix $M(\alpha) = \mathrm{diag}(m_i(\alpha))$ of order n being

$$m_i(\alpha) = 1 - \delta_i \left(1 - \frac{d}{dz}\gamma(z) \right) \tag{12}$$

with $\delta_i = 0$ if y_i is observed and $= 1$ otherwise and where the derivative of the function $\gamma(z)$ is evaluated at the point $z = -x_i^T\alpha$. The former expression (11) may be also written as

$$y(\alpha^{(r)}) - y(\alpha^{(r-1)}) = (I - M^{(r)})X(\alpha^{(r)} - \alpha^{(r-1)})$$

where $M^{(r)} = \mathrm{diag}(m_i^{(r)})$ coincides with M evaluated in an intermediate point of the segment joining $\alpha^{(r-1)}$ and $\alpha^{(r)}$. Substituting in (10) we obtain

$$\alpha^{(r+1)} - \alpha^{(r)} = S^{-1}S^{-1}X^T(I - M^{(r)})X(\alpha^{(r)} - \alpha^{(r-1)})$$

from where

$$S(\alpha^{(r+1)} - \alpha^{(r)}) = S^{-1}X^T(I - M^{(r)})XS^{-1}S(\alpha^{(r)} - \alpha^{(r-1)}).$$

Since $S^{-1}X^T(I - M^{(r)})XS^{-1} = Q^{(r)}$ is symmetric it follows that

$$\left\| S(\alpha^{(r+1)} - \alpha^{(r)}) \right\| \leq \tau^{(r)} \left\| S(\alpha^{(r)} - \alpha^{(r-1)}) \right\| \tag{13}$$

being $\tau^{(r)}$ the spectral radii of matrix $Q^{(r)}$, that is, the maximum of the eigenvalues of $Q^{(r)}$ in absolute value. Clearly

$$\tau^{(r)} = \max_{u \neq 0} \left| \frac{u^T S^{-1}X^T(I - M^{(r)})XS^{-1}u}{u^T u} \right| = \max_{u \neq 0} \left| \frac{u^T X^T(I - M^{(r)})Xu}{u^T X^T Xu} \right|.$$

It can easily be checked that, for double exponentially distributed errors, $0 < d\gamma(z)/dz \leq 1$ from where $1 - m_i^{(r)} \leq 1$ for $i \in I^{nobs}$ and $1 - m_i^{(r)} = 0$ otherwise and consequently

$$u^T X^T(I - M^{(r)})Xu \leq u^T(X^T X)^{nobs}u$$

which, from (6), means that

$$
\begin{aligned}
\tau^{(r)} &\leq \max_u \frac{u^T(X^TX)^{nobs}u}{u^T(X^TX)^{obs}u + u^T(X^TX)^{nobs}u} \\
&\leq \left(1 + \frac{\min\limits_u u^T(X^TX)^{obs}u}{\max\limits_u u^T(X^TX)^{nobs}u}\right)^{-1} = \left(1 + \frac{\sigma}{\omega}\right)^{-1} = \tau < 1
\end{aligned}
$$

being $\sigma > 0$ the minimum of the eigenvalues of $(X^TX)^{obs}$ and $\omega \geq 0$ the spectral radii of the definite nonnegative matrix $(X^TX)^{nobs}$. It follows from (13) that

$$
\left\| S(\alpha^{(r+1)} - \alpha^{(r)}) \right\| < \tau \left\| S(\alpha^{(r)} - \alpha^{(r-1)}) \right\|.
$$

This last expression ensures that

$$
\left\| S(\alpha^{(r+1)} - \alpha^{(r)}) \right\| < \tau^r \left\| S(\alpha^{(1)} - \alpha^{(0)}) \right\| \underset{r \to \infty}{\to} 0
$$

which implies that $\{\alpha^{(r)}\}$ converges necessarily to α^* and also that $\| \alpha^* - \alpha^{(r)} \| \leq c\tau^r$, completing the proof.

From now on, it will be assumed that $(X^TX)^{obs}$ is definite positive and we will take the unique fixed point α^* of (5) as the regression coefficient estimator

$$
\widehat{\alpha} = \alpha^* = \arg\min_\alpha H(\alpha)
$$

which is the unique limit point of all the algorithm sequences and, since determined by the implicit equation (5), it may be recognised as a particular case of the Huber M-estimates whose general behaviour may be seen in Huber (1981).

4   Stochastic asymptotic properties of the estimator

Consider now the former estimate as a function of n (say α_n, omitting the $^\wedge$ from now on) and let us investigate its asymptotic properties as

$n \to \infty$. Suppose that α is the real value of the regression parameter and write $y(\alpha)$ in the form

$$y(\alpha) = X\alpha + v \tag{14}$$

where the components of $v = (v_1, \cdots, v_n)^T$ are statistically independent and defined as

$$v_i = \begin{cases} \varepsilon_i / \varepsilon_i > -x_i^T \alpha & \text{with probability} \quad 1 - F(-x_i^T \alpha) \\ \gamma(-x_i^T \alpha) & \text{with probability} \quad F(-x_i^T \alpha) \end{cases}$$

being F the cumulative probability function of the double exponential distribution. It may be easily checked that

$$E(v_i) = F(-x_i^T \alpha)\gamma(-x_i^T \alpha) - \int_{-\infty}^{-x_i^T \alpha} t f(t) dt$$

$$V(v_i) \le E(v_i^2) = F(-x_i^T \alpha)\gamma(-x_i^T \alpha)^2 + \int_{-x_i^T \alpha}^{\infty} t^2 f(t) dt < \infty.$$

Theorem 4.1. *Let λ_n^{min} be the minimum of the eigenvalues of $n^{-1}(X^T X)^{obs}$. If $(X^T X)^{obs}$ is positive definite and the three following technical conditions hold (for some $0 \le \beta < 1/2$ and $\delta > 0$)*

i) $\displaystyle\liminf_n \lambda_n^{min} = \lambda > 0$

ii) $\displaystyle\limsup_n \frac{1}{n^{\beta/2}} \parallel x_n \parallel^{1+\delta} = K < \infty$

iii) $\displaystyle\limsup_n \frac{1}{n^{1-\beta}} \left(\sum_{i=1}^{n} \parallel x_i \parallel^{-\delta} \right)^2 = \eta < \infty$

then α_n converges to α in L_2 which implies that α_n is a consistent estimate of α and, also, the sequence $n^{1/2}(\alpha_n - \alpha)$ is bounded in L_2.

Proof. Consider again the matrix $M(\alpha) = \text{diag}(m_i(\alpha))$ whose elements are defined as in (12) and observe that

$$v_i = \varepsilon_i - \delta_i(\varepsilon_i - \gamma(-x_i^T \alpha))$$

with $\delta_i = 1$ if y_i has not being observed ($\varepsilon_i < -x_i^T\alpha$) and $= 0$ otherwise. As in section 3, we can write for all fixed n

$$y(\alpha_n) - y(\alpha) = (I - M^*)X(\alpha_n - \alpha) \tag{15}$$

where M^* coincides with M evaluated at an intermediate point of the linear segment joining α and α_n. Since each α_n is a fixed point of the transformation $T(z)$, it must satisfy (5) which, together with (15), leads to

$$\alpha_n = (X^TX)^{-1}X^T(X\alpha + v + (I - M^*)X(\alpha_n - \alpha)).$$

It follows that

$$X^Tv = X^TM^*X(\alpha_n - \alpha) = [(X^TM^*X)^{obs} + (X^TM^*X)^{nobs}](\alpha_n - \alpha). \tag{16}$$

Obviously, $(X^TM^*X)^{obs} = (X^TX)^{obs}$ and, since $\gamma(z)$ is strictly increasing, $(X^TM^*X)^{nobs}$ is positive semidefinite. It follows that the minimum eigenvalue of X^TM^*X cannot be smaller than $n\lambda_n^{\min}$, consequently from (16)

$$\|\alpha_n - \alpha\| = \left\|n(X^TM^*X)^{-1}n^{-1}X^Tv\right\| \le \lambda^{-1}\left\|n^{-1}X^Tv\right\|.$$

Finally observe that

$$E\|\alpha_n - \alpha\|^2 \le \lambda^{-2}E\left\|n^{-1}X^Tv\right\|^2 \underset{n\to\infty}{\to} 0$$

since, on the one hand

$$E\|n^{-1}X^Tv\|^2 \le n^{-2}d\sum_{i,i'}^n |E(v_iv_{i'})|\|x_i\|\|x_{i'}\|$$

and, on the other, it can be checked that all $E(v_iv_i')$ are bounded. To confirm this, first observe after simple calculations that

$$F(z)\gamma(z) - \int_{-\infty}^z tf(t)dt \underset{|z|\to\infty}{\to} 0$$

and also

$$F(z)\gamma(z)^2 + \int_z^\infty t^2f(t)dt \underset{z\to\infty}{\to} 0 \text{ and } \underset{z\to-\infty}{\to} \mu_2 = \int_{-\infty}^\infty t^2f(t)dt < \infty.$$

These facts, along with the continuity of both terms to the left of the limit symbols as functions of z, respectively imply the mentioned boundness of the first and the second moments of the sequence $\{v_i\}$. Assuming that $\mid E(v_i) \mid \leq C < \infty$ and $E(v_i^2) \leq \sigma^2 < \infty$, then we can write

$$E\left\|n^{-1}X^T v\right\|^2 \leq n^{-1}dK^2 \max(C^2, \sigma^2)\eta \underset{n\to\infty}{\to} 0.$$

As a consequence, α_n converges to α in L_2 what implies that α_n is a consistent estimate for α. In a similar way, it can be proved that

$$\limsup_n \left\|n^{1/2}X^T E(v)\right\|^2 < \infty$$

and also

$$\limsup_{n\to\infty} E\left\|\sqrt{n}(\alpha_n - \alpha)\right\|^2 < \infty \tag{17}$$

which means that the sequence $n^{1/2}(\alpha_n - \alpha)$ is L_2-bounded, completing the proof.

Observe, in addition, that under the hypothesis of theorem 4.1

$$\underset{n\to\infty}{\text{plim}} \left\|n^{-1}X^T(M^* - M)X\left(\sqrt{n}(\alpha_n - \alpha)\right)\right\| = 0 \tag{18}$$

where $M = M(\alpha)$ and plim (or $\to_P$) denotes convergence in probability. To see (18), let represent by τ_n the spectral radii of $n^{-1}X^T(M^* - M)X$ and take into consideration that $(d/dz)\gamma(z)$ is a flat null function on $(-\infty, 0)$ and

$$0 < \left|\frac{d^2}{dz^2}\gamma(z)\right| \leq 2 \quad (\text{for } z > 0)$$

which ensures that $\gamma'(z)$ is an uniformly continuous function being 2 its module of uniform continuity. The following chain of inequalities holds

$$\left\|n^{-1}X^T(M^* - M)X(\sqrt{n}(\alpha_n - \alpha))\right\| \leq \tau_n \left\|\sqrt{n}(\alpha_n - \alpha)\right\|$$

$$= \left(\max_{|\xi|=1} \left|\xi^T n^{-1}X^T(M^* - M)X\xi\right|\right) \left\|\sqrt{n}(\alpha_n - \alpha)\right\|$$

$$\leq Trace(n^{-1}X^T X)\left(\max_{i\leq n}|m_i(\alpha_n^*) - m_i(\alpha)|\right)\left\|\sqrt{n}(\alpha_n - \alpha)\right\| \leq$$

$$\leq Trace(n^{-1}X^T X)2\left(\max_{i\leq n}\| x_i \|\right)\|\alpha_n^* - \alpha\|\left\|\sqrt{n}(\alpha_n - \alpha)\right\|.$$

Since α_n^* belongs to the segment joining α_n and α, it follows from (17) that

$$
\left\| n^{-1} X^T (M^* - M) X \left(\sqrt{n}(\alpha_n - \alpha) \right) \right\|
$$
$$
\leq 2n^{-(1-\beta)/2} K^3 \eta \left\| \sqrt{n}(\alpha_n - \alpha) \right\|^2 \underset{n \to \infty P}{\to} 0
$$

which implies (18). Equally, it can be proved that

$$
\underset{n \to \infty}{\text{plim}} \left\| n^{-1} X^T (M - E(M)) X \left(\sqrt{n}(\alpha_n - \alpha) \right) \right\| = 0 \tag{19}
$$

where $E(M) = \text{diag}(E(m_i(\alpha)))$ being

$$
E(m_i(\alpha)) = 1 - E(\delta_i) \left(1 - \frac{d}{dz} \gamma(z) \right) = 1 - F(-x_i^T \alpha) \left(1 - \frac{d}{dz} \gamma(z) \right) .
$$

To see (19), let denote $n^{-1} X^T (M - E(M)) X = P = (p_{j,j'})$, then observe that

$$
\left\| n^{-1} X^T (M - E(M)) X (\sqrt{n}(\alpha_n - \alpha)) \right\|^2 \leq \left(\sum_{j,j'} p_{j,j'}^2 \right) \left\| \sqrt{n}(\alpha_n - \alpha) \right\|
$$

being

$$
p_{j,j'} = -n^{-1} \sum_{i=1}^{n} x_{i,j} x_{i,j'} (1 - \gamma'(-x_i^T \alpha))(\delta_i - F(-x_i^T \alpha)).
$$

Since $0 < \gamma'(z) \leq 1$, we conclude that

$$
\sum_{j,j'} p_{j,j'}^2 \leq n^{-1} d^2 K^2 \eta \underset{n \to \infty}{\to} 0.
$$

This, after considering (17), leads us to write

$$
E \left\| n^{-1} X^T (M - E(M)) X (\sqrt{n}(\alpha_n - \alpha)) \right\|^2
$$
$$
\leq o(1) E \left\| \sqrt{n}(\alpha_n - \alpha) \right\|^2 \underset{n \to \infty}{\to} 0
$$

which guarantees (19).

We are now in condition to formulate the following theorem which establishes the asymptotic distribution of the regression parameter estimate.

Theorem 4.2. *Under the hypothesis of the theorem 4.1, $n^{1/2}(\alpha_n - E(\alpha_n))$ is asymptotically distributed as a normal*

$$\sqrt{n}(\alpha_n - E(\alpha_n)) \underset{n\to\infty D}{\to} N(0,\Lambda)$$

where the asymptotic variance Λ will be given during the course of the proof.

Observe that this theorem indicates that, for n large enough, $\alpha_n - E(\alpha_n)$ is approximately distributed as

$$\alpha_n - E(\alpha_n) \approx N(0, n^{-1}\Lambda)$$

and considering that

$$\|E(\alpha_n) - \alpha\| \leq E\|\alpha_n - \alpha\| \leq \left(E\|\alpha_n - \alpha\|^2\right)^{1/2} \underset{n\to\infty}{\to} 0$$

it can be concluded that $\alpha_n - E(\alpha_n)$ and $\alpha_n - \alpha$ have identical asymptotic distributions; therefore the following approximation corollary directly derives.

Corollary 4.1. *Under the assumptions of the theorem 4.2, the distribution of $\alpha_n - \alpha$ can be approximated by*

$$\alpha_n - \alpha \approx N(0, n^{-1}\Lambda).$$

Proof (of theorem 4.2): First (after $\pm(M + M^*)$) write

$$n^{-1}X^T E(M)X\sqrt{n}(\alpha_n - \alpha) = n^{-1}X^T M^* X\sqrt{n}(\alpha_n - \alpha) + \zeta_n = n^{-1/2}X^T v + \zeta_n$$

with $\text{plim}\,\zeta_n = 0$. Then, lets show that $n^{-1}X^T E(M)X$ is nonsingular, for large enough n. It is sufficient to prove that the minimum eigenvalue $\eta_n^{\min}$ of the symmetric matrix $n^{-1}X^T E(M)X$ is bigger than zero. To see this, take into consideration that

$$\eta_n^{\min} \geq \lambda - \sqrt{\sum_{j,j'} p_{j,j'}^2} \underset{n\to\infty}{\to} \lambda > 0$$

which means that, for large enough n, $\eta_n^{\min} > \lambda/2 > 0$. As a consequence, if $W = \lim_{n \to \infty} n^{-1} X^T E(M) X$ then W is nonsingular and also, if n is large enough, it can be written

$$\sqrt{n}(\alpha_n - \alpha) = (n^{-1} X^T E(M) X)^{-1} n^{-1/2} X^T v + \varphi_n$$

where again plim $\varphi_n = 0$ because

$$0 \leq \left\| (n^{-1} X^T E(M) X)^{-1} \zeta_n \right\| \leq 2\lambda^{-1} \left\| \zeta_n \right\| \underset{n \to \infty P}{\to} 0.$$

Therefore, the asymptotic distributions of $n^{1/2}(\alpha_n - \alpha)$ and $(n^{-1} X^T E(M) X)^{-1} n^{-1/2} X^T v = n^{-1/2} Q v$ must coincide. Finally, since it was proved that $\sup \| n^{-1} X^T E(v) \|^2 < \infty$, it follows from

$$\left\| n^{-1/2} Q E(v) \right\| \leq 2\lambda^{-1} \left\| n^{-1/2} X^T E(v) \right\|$$

that $n^{-1/2} Q E(v)$ is a bounded sequence of $d-$dimensional vectors. After centering on the mean, the stochastic vector $n^{-1/2} Q(v - E(v))$ is

$$n^{-1/2} Q(v - E(v)) = n^{-1/2} \sum_{i=1}^{n} (n^{-1} X^T E(M) X)^{-1} x_i (v_i - E(v_i)).$$

If f_j is the row $j (j = 1, ., d)$ of $(n^{-1} X^T E(M) X)^{-1}$, then the jth component of $n^{-1/2} Q(v - E(v))$ equals

$$(n^{-1/2} Q(v - E(v)))_j = n^{-1/2} \sum_{i=1}^{n} f_j x_i (v_i - E(v_i)).$$

It will be enough to show that each one of these components is asymptotically normally distributed. As it is well known (Laha and Roatgi (1979), p. 289), the following two relations suffice it since they together imply the Lindeberg-Feller central limit theorem conditions

$$\max_{i \leq n} (n^{-1/2} f_j x_i (v_i - E(v_i))) \underset{n \to \infty}{\to} 0$$

$$\sum_{i=1}^{n} V(n^{-1/2} f_j x_i (v_i - E(v_i))) \leq C < \infty \quad \text{(for all } n \geq 1\text{)}.$$

To see both, observe that each candidate to be included in either the maximum or the sum is upper bounded by

$$n^{-1}(f_j x_i)^2 \sigma^2 \leq n^{-1} \left\| (n^{-1}X^T E(M)X)^{-1} x_i \right\|^2 \leq 4\lambda^{-2} n^{-1} \sigma^2 K^2 \underset{n\to\infty}{\to} 0.$$

It has been proved that

$$n^{-1/2} Q(v - E(v)) \underset{n\to\infty D}{\to} N(0, \Lambda) \tag{20}$$

for some square matrix Λ of order d, concluding the proof.

Finally, let us represent by Ψ the diagonal covariance matrix of the random vector v and assume that $\Phi = lim_n n^{-1} X^T \Psi X$ and $W = lim_{n\to\infty} n^{-1} X^T E(M)X$. Then it is clear that $n^{1/2}(\alpha_n - E(\alpha_n))$ converges in distribution to the $d-$variate normal distribution $N(0, W^{-1}\Phi W^{-1})$. Therefore, for each fixed n, let define $W_n = n^{-1} X^T M(\alpha_n)X$. We claim that $W_n \to_n W$ a.e.. To see this, observe that if $P(\alpha_n) = (p_{j,j'}(\alpha_n)) = n^{-1} X^T (M(\alpha_n) - E(M))X$, then reasoning as when proving (19)

$$\left| p_{j,j'}(\alpha_n) \right| \leq n^{-1} \sum_{i=1}^{n} |x_{i,j}| |x_{i,j'}| (\delta_i + F(-x_i^T \alpha))$$

from where

$$\| n^{-1} X^T (M(\alpha_n) - E(M))X \|^2 \leq 2n^{-1} d^2 K^2 \eta \underset{n\to\infty}{\to} 0.$$

This means that

$$\| W_n - W \|^2 = \left\| n^{-1} X^T (M(\alpha_n) - E(M))X \right\|^2 \underset{n\to\infty}{\to} 0 \quad \text{a.e.}$$

In the same way, by defining, for all real π, the diagonal matrix $\Psi(\pi) = \text{diag}(V(v_i(\pi))$ where $v_i(\pi)$ represents the random variable

$$v_i(\pi) = \begin{cases} \varepsilon/\varepsilon_i > -x_i^T \pi & \text{with probability} \quad 1 - F(-x_i^T \pi) \\ \gamma(-x_i^T \pi) & \text{with probability} \quad F(-x_i^T \pi) \end{cases}$$

it follows that

$$\Phi_n = n^{-1} X^T \Psi(\alpha_n)X \underset{n\to\infty}{\to} \Phi \quad \text{a.e.} .$$

As a consequence, also

$$W_n^{-1}\Phi_n W_n^{-1} \underset{n\to\infty}{\to} W^{-1}\Phi W^{-1} \quad \text{a.e.}$$

and the following corollary directly derives under the assumptions above stated.

Corollary 4.2. *Under the hypothesis of the theorem 4.1, $W_n^{-1}\Phi_n W_n^{-1}$ consistently estimates the asymptotic variance of $n^{1/2}(\alpha_n - \alpha)$.*

This last estimate of the asymptotic covariance matrix may be used to carry out hypothesis testing as well as confidence regions for α and also for linear transformations of regression parameters, in the usual way.

5   Some natural extensions

Lets assume that each y_i is only observed when $y_i > c_i$ being all $c_i's$ known. Very usually in this kind of models it is assumed the natural condition

$$\limsup_{n\to\infty} | c_i | < \infty$$

to prevent an excessive number of censored data. Redefine (2) as

$$y_i(\alpha^{(r)}) = \begin{cases} y_i, & i \in I^{obs} \\ x_i^T\alpha^{(r)} + \gamma(-x_i^T\alpha^{(r)} + c_i), & i \in I^{nobs} \end{cases}$$

The following results generalise the above mentioned ones and may easily be shown by simply readjusting the proofs.

(a) For n fixed, the lemma 2.1 and the theorem 2.1 remain true if the function $H(\alpha)$ of expression (9) is substituted by

$$H(\alpha) = 2^{-1}\sum_{i\in I^{obs}} (y_i - x_i^T\alpha)^2 + \sum_{i\in I^{nobs}} \Gamma(-x_i^T\alpha + c_i).$$

(b) After redefining each v_i of (18) as

$$v_i = \begin{cases} \varepsilon_i/\varepsilon_i > -x_i^T\alpha + c_i, & \text{with probability} \quad 1 - F(-x_i^T\alpha + c_i) \\ \gamma(-x_i^T\alpha + c_i), & \text{with probability} \quad F(-x_i^T\alpha + c_i) \end{cases}$$

the asymptotic results contained in theorems 4.1 and 4.2 remain valid with no restriction.

For both, it is necessary to slightly mend the matrix function $M(\alpha) = \text{diag}(m_i(\alpha))$ by redefining (12) as

$$m_i(\alpha) = 1 - \delta_i \left(1 - \frac{d}{dz}\gamma\left(-x_i^T\alpha + c_i\right)\right).$$

Finally, it was assumed from the beginning of this paper that the regression hyperplane we have been seeking for might not contain the origin so that $x_{i1} = 1 (\Rightarrow \| x_i \| \geq 1)$ for all i. At this point, it may be checked that this initial assumption is completely irrelevant with respect to the convergence properties included in the epigraph 3. The reader may revise the proofs concerning the asymptotic results maintained in section 4 to realise that all remain true if this initial restriction is merely substituted by the slight one $\| x_i \| \geq k > 0$. Therefore, after incorporating this last as a new technical condition in theorem 4.1, the former may be released with no additional consequences.

6   A simulation study

In this section we intend, on the one hand, to show the general performance of the median-based recursive estimation process under several scenarios and, on the other hand, to compare the results got by this method with those derived from some of its usual alternatives. We have started from a simple regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

after having fixed the arbitrary values $\beta_0 = -3$ and $\beta_1 = 5$. Later, we have generated a list of n random numbers x_i from a $N(1,2)$ distribution which were treated as fixed in all the following scenarios commented below. The first one concerns the error distributions, so that we have generated, independently on the $x_i's$, specific values for ε_i following respectively a Laplace distribution and also a standard normal distribution, both having variance one. The censored data have been implemented by a two-stage controlled process which easily allows us to foresee the expected percent of censored data under normality. Firstly, we have drawn (randomly and independently on the above simulated $x_i/\varepsilon_i's$) the values c_i^* from a $N(\mu_0, 1)$ distribution and, afterwards, we

have fixed the real censure values $c_i = \beta_0 + \beta_1 x_i + c_i^*$. Thus, the probability of y_i being censored is $P(\varepsilon_i < c_i^*) = P(N(-\mu_0, 2^{1/2}) < 0)$ and, as a consequence, the expected percentage of censored data may be controlled by simply varying the position parameter μ_0. For instance, direct calculations indicate that we must respectively assign the position values $\mu_0 = -0.742, -0.358, 0$ in order to get expected %'s of censored data equal to $30\%, 40\%, 50\%$ as displayed in the results. To end up, we have made $N = 1000$ replications of the above mentioned generation of errors and have run for each one of their corresponding data files the new median-based algorithm as well as the mean-based one. At this respect, note that the last mean-based algorithm has good asymptotic properties only when errors are normally distributed since it coincides with the well known EM algorithm. In a similar way, it has been only proved in this article that the median-based algorithm here proposed has good asymptotic abilities when errors follow a Laplace distribution. Table 1 depicts the results that we have achieved under all the above mentioned circumstances for different values of $n(= 150, 100, 50)$ seeking by them to also have an empirical feeling about the sensibility of the estimates on the sample size. In each case, the performances of the different regression parameter estimates has been reflected both individually and globally. The first by means of the empirical estimates of bias (B) and variance (V) corresponding to each one of the parameters β_0 and β_1, namely

$$\widehat{B}(\widehat{\beta}_j) = N^{-1} \sum_{i=1}^{N} \widehat{\beta}_j - \beta_j$$

$$\widehat{V}(\widehat{\beta}_j) = N^{-1} \sum_{i=1}^{N} (\widehat{\beta}_j^{(i)} - \overline{\widehat{\beta}}_j)^2.$$

The second, on its part, from the empirical estimate of the multidimensional mean square error given by

$$M\widehat{S}E = \sum_j \left\{ \widehat{B}\left(\widehat{\beta}_j\right)^2 + \widehat{V}\left(\widehat{\beta}_j\right) \right\}$$

where $\widehat{\beta}_j^{(i)}$ is the estimate of β_j at the ith replication of data and $\overline{\widehat{\beta}}_j = N^{-1} \sum_i \widehat{\beta}_j^{(i)}$ is the empirical mean of the former estimates ($i = 1, \cdots, N; j = 0, 1$).

Sample size n	Expected % of censored data	Median-based estimates		Mean-based estimates	
		Laplace errors	N(0,1) errors	Laplace errors	N(0,1) errors*
Empirical bias for $\hat{\beta}_0$					
150	30	-0.153	-0.172	-0.226	0.162
	40	0.177	-0.226	-0.251	0.181
	50	0.231	0.219	-0.230	-0.179
100	30	0.154	-0.166	0.184	-0.180
	40	-0.188	-0.195	-0.226	0.174
	50	0.194	0.238	-0.253	-0.175
50	30	-0.160	0.218	0.174	-0.181
	40	-0.179	0.202	0.194	-0.185
	50	0.192	-0.197	0.225	0.217
Empirical bias for $\hat{\beta}_1$					
150	30	-0.314	0.402	0.423	-0.293
	40	0.300	-0.397	0.477	0.331
	50	-0.352	-0.463	-0.465	-0.340
100	30	0.334	-0.436	-0.453	-0.318
	40	-0.326	-0.511	-0.446	-0.369
	50	-0.401	0.484	0.487	-0.355
50	30	-0.378	-0.392	0.443	-0.403
	40	0.346	-0.426	0.460	0.373
	50	-0.415	0.438	-0.455	-0.399
Empirical variance for $\hat{\beta}_0$					
150	30	1.323	1.526	1.613	1.183
	40	1.479	1.574	1.779	1.633
	50	1.673	2.025	2.106	1.784
100	30	1.438	1.609	1.838	1.352
	40	1.720	1.871	2.180	1.910
	50	2.083	2.293	2.369	2.273
50	30	2.103	2.365	2.557	1.798
	40	2.384	2.691	3.009	2.575
	50	2.899	3.354	3.477	3.106
Empirical variance for $\hat{\beta}_1$					
150	30	2.105	2.339	2.647	1.948
	40	2.518	2.579	3.201	2.647
	50	3.050	3.369	3.609	3.226
100	30	2.516	2.807	3.205	2.293
	40	3.009	3.108	3.851	3.164
	50	3.645	3.927	4.391	3.856
50	30	3.685	3.995	4.475	3.299
	40	4.228	4.409	5.331	4.335
	50	5.116	5.699	6.201	5.353
Empirical bivariate mean square errors for $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$					
150	30	3.550	4.056	4.490	3.243
	40	4.118	4.362	5.271	4.422
	50	4.900	5.656	5.984	5.158
100	30	4.089	4.634	5.282	3.779
	40	4.871	5.278	6.281	5.240
	50	5.926	6.511	7.061	6.286
50	30	5.956	6.561	7.259	5.292
	40	6.764	7.322	8.589	7.083
	50	8.224	9.284	9.936	8.665

results, it can be noted that the empirical performance of the median-based estimate with Laplace errors is quite similar to the EM-estimate's (this last with normally distributed errors). Their values are of the same order of magnitude either in bias or in variance terms (whence also in MSE) and, additionally, the smaller ciphers alternate from one to another depending on the scenarios what seemingly points to non systematic low-scale improvements between them. Both are unarguable better (save for isolated exceptions) to the estimates obtained after interchanging their error distributions. In all cases, their effectiveness decreases, on the one hand, as the mean fraction of censored data increases and, on the other hand, as the sample size diminishes, although this feature coincides with what might be expected in advance.

7   An application to real data

Schmee and Hahn (1979) presented certain motorette censored data which later were also considered by Tanner (1993; p. 41-43) who assumed normally distributed errors and applied the EM algorithm to a simple regression model with right censored data. The motorettes were tested at temperatures of $150°, 170°, 190°$ and $220°$ in degrees $°C$ and their recorded times to failure are given below:

150°	170°	190°	220°
8064*	1764	408	408
8064*	2772	408	408
8064*	3444	1344	504
8064*	3542	1344	504
8064*	3780	1440	504
8064*	4860	1680*	528*
8064*	5196	1680*	528*
8064*	5448*	1680*	528*
8064*	5448*	1680*	528*
8064*	5448*	1680*	528*

A star indicates that a motorette was taken off study without failing at the indicated event time. For this data Tanner fitted the model

$$t_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where $\varepsilon_i \sim N(0, \sigma)$, $x_i = 1000(\text{temperature} + 273.2)^{-1}$ and $t_i = \log_{10}(\text{ith}$ failure time). The standard deviation σ was estimated to be equal to 0.2592 and also

$$\widehat{\beta}_0 = -6.019, \quad \widehat{\beta}_1 = 4.311.$$

In order to make the situation comparable with the assumptions of this paper, we will fit the model

$$y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i$$

where the errors ε_i are independently distributed as the symmetrised exponential with parameter one, $y_i = -2^{1/2} 0.2592^{-1} t_i$ and each α_i should be equal to $-2^{1/2} 0.2592^{-1} \beta_i$ if the influence of the different assumed error distributions were negligible. The discrepancies between $\widehat{\alpha}_i$ and this last expression evaluated at $\widehat{\beta}_i$ will measure the combined effect due to their different (median versus mean-based) imputation processes as well as to their distint (Laplace versus normal) error distributions. None of these last are skewed but, at this respect, the logarithm transformation on the survival data undoubtedly seeks to diminish the positive skewness of the original survival data. For each censored data (with star), we have assumed that its correspondent y_i has not been observed since

$$y_i < -\frac{\sqrt{2}}{0.2592} \log_{10}(\text{ith failure time}^*).$$

After running the algorithm on the Schmee and Hahn data, we have found the convergence point to be

$$\widehat{\alpha}_0 = 28.5545, \quad \widehat{\alpha}_1 = -21.2926.$$

These limit values, as proved before, are independent on the starting points. The adjoining Table 2 shows this fact and also reflects the algorithm sequences $\{\alpha^{(r)}\}$ corresponding to several initial values $\alpha^{(0)}$ which have been chosen trying to cover a great spectrum both in the magnitude as well as in the signature of its components. For each sequence, we have shaded and printed in bold characters the first iteration r that fulfils the condition $\mid \alpha_i^{(r)} - \alpha_i^{(r-1)} \mid < 10^{-4}$ for all i. As it can be observed, this usual stopping rule would has been exerted after a number of iterations that slightly varies amidst 17 and 21 depending on the cases.

| Iterations | Algorithm starting values $\alpha^{(0)}$ | | | |
r	(0,0)	(-40,10)	(32,25)	(60,-40)
1	(27.4741,-20.7798)	(44.6638,-29.7694)	(77.3736,-46.0576)	(12.0202,-14.0553)
2	(28.1005,-21.0771)	(42.5746,-28.2017)	(67.2825,-40.2408)	(22.1794,-18.4771)
3	(28.3605,-21.2006)	(37.9248,-25.7926)	(54.6656,-33.8830)	(26.5752,-20.3961)
4	(28.4711,-21.2531)	(33.5260,-23.6603)	(44.0218,-28.7134)	(28.0744,-21.0648)
5	(28.5185,-21.2756)	(30.9506,-22.4300)	(37.0087,-25.3311)	(28.3495,-21.1953)
6	(28.5390,-21.2853)	(29.6363,-21.8061)	(32.8330,-23.3274)	(28.4664,-21.2508)
7	(28.5478,-21.2894)	(29.0321,-21.5193)	(30.5732,-22.2508)	(28.5165,-21.2746)
8	(28.5516,-21.2912)	(28.7628,-21.3915)	(29.4604,-21.7227)	(28.5381,-21.2849)
9	(28.5532,-21.2920)	(28.6448,-21.3355)	(28.9531,-21.4818)	(28.5474,-21.2893)
10	(28.5539,-21.2924)	(28.5936,-21.3112)	(28.7281,-21.3750)	(28.5514,-21.2912)
11	(28.5542,-21.2925)	(28.5714,-21.3006)	(286297,-21.3283)	(28.5531,-21.2920)
12	(28.5544,-21.2926)	(28.5618,-21.2961)	(28.5870,-21.3081)	(28.5539,-21.2924)
13	(28.5544,-21.2926)	(28.5576,-21.2941)	(28.5685,-21.2993)	(28.5542,-21.2925)
14	(28.5544,-21.2926)	(28.5558,-21.2933)	(28.5605,-21.2955)	(28.5544,-21.2926)
15	(28.5544,-21.2926)	(28.5550,-21.2929)	(28.5571,-21.2939)	(28.5544,-21.2926)
16	(28.5545,-21.2926)	(28.5547,-21.2927)	(28.5556,-21.2932)	(28.5544,-21.2926)
▉	▉	(28.5546,-21.2927)	(28.5549,-21.2929)	(28.5544,-21.2926)
18	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5547,-21.2927)	(28.5545,-21.2926)
▉	(28.5545,-21.2926)	▉	(28.5545,-21.2927)	▉
20	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5545,-21.2926)
▉	(28.5545,-21.2926)	(28.5545,-21.2926)	▉	(28.5545,-21.2926)
22	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5545,-21.2926)
23	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5545,-21.2926)
24	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5545,-21.2926)
25	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5545,-21.2926)	(28.5545,-21.2926)

References

[1] S. Chatterjee, and D.L. McLeish, (1986). *Fitting linear regression models to censored data by least squares and maximum likelihood methods*, Comm. Statist. Theory Methods **15**, 3227-3243.

[2] A.P. Dempster, N.M. Laird, and D.B Rubin, (1977). *Maximum likelihood from incomplete data via the EM algorithm*, J. Roy. Statist. Soc. **B 39**, 1-22.

[3]   M.J.R. Healy and M. Westmacott, (1956). *Missing values in experiments analized on automatic computers*, Appl. Statist. **5**, 203-206.

[4]   P.J. Huber, (1981). *Robust statistics*, Wiley.

[5]   I.R. James and P.J. Smith, (1984). *Consistency results for linear regression with censored data*, Ann. Statist. **12**, 590-600.

[6]   R.G. Laha and V.K. Rohatgi, (1979). *Probability Theory*, Wiley.

[7]   R.J.A. Little and D.B. Rubin, (1987). *Statistical analysis with missing data*, Wiley.

[8]   T. Orchard and M.A. Woodbury, (1972). *A missing information principle: Theory and applications*, Proc. Sixth Berkeley Symp. Math. Statist. Probab. 697-715. Univ. California Press, Berkeley.

[9]   Y. Ritov, (1990). *Estimation in linear regression model with censored data*, Ann. Statist. **18**, 303-328.

[10] J. Schmee and G.J. Hahn, (1979). *A simple method for regression analysis with censored data*, Technometrics **21**, 417-432.

[11] M.A. Tanner, (1993), *Tools for statistical inference. Methods for the exploration of posterior distributions and likelihood functions*, Springer-Verlag.

[12] C.F.J. Wu, (1983). *On the convergence of the EM algorithm*, Ann. Statist. 11, 95-103.

Universidad Autónoma de Madrid
Universidad Complutense de Madrid