Computational Statistics and Data Analysis 55 (2011) 874–887

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust accelerated failure time regression

Isabella Locatelli [a], Alfio Marazzi [a,*], Victor J. Yohai [b]

[a] Institute for Social and Preventive Medicine, Centre Hospitalier Universitaire Vaudois and University of Lausanne, route de la Corniche 2, CH 1006 Epalinges, Switzerland

[b] Departamento de Matematicas, Facultad de Ciencias Exactas y Naturales, University of Buenos Aires and CONICET, Argentina

ARTICLE INFO

Article history:
Received 30 January 2010
Received in revised form 19 July 2010
Accepted 19 July 2010
Available online 29 July 2010

Keywords:
Accelerated failure time models
Robust regression
Censoring

ABSTRACT

Robust estimators for accelerated failure time models with asymmetric (or symmetric) error distribution and censored observations are proposed. It is assumed that the error model belongs to a log-location-scale family of distributions and that the mean response is the parameter of interest. Since scale is a main component of mean, scale is not treated as a nuisance parameter. A three steps procedure is proposed. In the first step, an initial high breakdown point S estimate is computed. In the second step, observations that are unlikely under the estimated model are rejected or down weighted. Finally, a weighted maximum likelihood estimate is computed. To define the estimates, functions of censored residuals are replaced by their estimated conditional expectation given that the response is larger than the observed censored value. The rejection rule in the second step is based on an adaptive cut-off that, asymptotically, does not reject any observation when the data are generated according to the model. Therefore, the final estimate attains full efficiency at the model, with respect to the maximum likelihood estimate, while maintaining the breakdown point of the initial estimator. Asymptotic results are provided. The new procedure is evaluated with the help of Monte Carlo simulations. Two examples with real data are discussed.

1. Introduction

Positive random variables with asymmetric distributions arise in many applications (e.g., analysis of survival times, income and expenditures, output of biological systems). Often the population mean is the parameter of interest and depends upon a number of covariates. The data may contain censored observations as well as outliers; these features make the mean a difficult parameter to estimate.

As an example, we can mention the problem of estimating the expected length of stay (LOS) in a hospital as a function of available patient characteristics, such as diagnosis, treatment, and type of admission (regular or emergency). LOS is often used as a substitute of cost of stay and the expected LOS is necessary for comparing hospital activities, planning, and budgeting purposes. Stays may be censored because a patient may die or be transferred to a different hospital before the ordinary home discharge. In addition, LOS distributions are skewed and often contain an important number of "outliers". Outliers are observations markedly different from most others, often extremely long or surprisingly short stays, but also patients with unusual characteristics. When a small number of outliers are observed, the classical estimates of the conditional mean can be much different than when none is observed. Since the values and the frequency of outliers fluctuate from sample to sample, the mean and the associated inferences are unreliable. For this reason, various empirical rules are currently used by the practitioners in the domain of hospital management to distinguish typical cases, or "inliers", from outliers

* Corresponding author. Tel.: +41 213147260; fax: +41 213147373.
E-mail addresses: isabella.locatelli@chuv.ch (I. Locatelli), alfio.marazzi@chuv.ch (A. Marazzi), vyohai@dm.uba.ar (V.J. Yohai).

0167-9473/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2010.07.017

(e.g., Cots et al., 2003). Inliers are paid on the ground of standard rates based on estimates of the mean cost; outliers are reimbursed on the ground of special negotiations.

The proportional hazards model (Cox, 1972) is most commonly used in practice to describe censored data. However, as noted by Cox (Reid, 1994, p. 450), accelerated failure time (AFT) regression (e.g., Cox and Oakes, 1984, Chap. 5; Kalbfleisch and Prentice, 2002, Chap. 2; Lawless, 2003, Chap. 6) is "in many ways more appealing than the proportional hazards model because of its quite direct physical interpretation". AFT models assume a log-linear relationship between the response and the explanatory variables, where the error distribution belongs to a given parametric family. However, "answers are quite insensitive to the parametric formulation" (Cox, in Reid, 1994, p. 450). Thus, AFT models are very convenient in many prediction problems, especially when one has to predict outside the range of the sample, or when the sample size is small. Usually, the maximum likelihood (ML) method is used to estimate the model parameters and the conditional mean. Unfortunately, the ML estimate is extremely sensitive to outliers.

Several distribution free, rank based, and semiparametric methods for AFT models have been developed (e.g., Jin et al., 2003; Zeng and Lin, 2007). These methods are fairly stable with respect to outliers in the response variable but are very sensitive to leverage points, i.e., to outliers in the covariate components.

In this paper, we will consider robust procedures for AFT regression with censored data that are robust with respect to both outliers in the response and leverage points, that provide stable inferences for the inliers, and identify outliers. To the best of our knowledge, the sole published methodology meeting these requirements are the high breakdown point (bdp) S and MM estimates proposed in Salibian-Barrera and Yohai (2008). We will pay a special attention to the nonparametric S estimate defined by these authors. This procedure, based on a symmetric loss, can consistently estimate the conditional mean of the (log-) response when the error distribution is symmetric. However, it is not designed for asymmetric errors. In addition, the procedure uses the conditional expectation approach introduced by Buckley and James (1979), where functions of censored residuals are replaced by their estimated conditional expectation given that the response is larger than the recorded censored value. The conditional expectation is based on the Kaplan–Meier (KM) distribution $K(r)$ of the residuals. From the point of view of robustness, the KM estimate has an important drawback, because it distributes the mass of each censored residual r_i^* among all the noncensored residuals r_j such that $r_j > r_i^*$. The worst case occurs when all the censored residuals are between the good noncensored residuals and the outliers. Suppose that we have a fraction ϵ of extremely large residuals and a proportion λ of censored residuals. Then, the mass given to the largest residuals by the KM estimate is $\lambda + \epsilon$. Therefore, the regression estimate is more affected by outliers than in the uncensored case. In addition, when the largest residual r_k^* is censored, the KM estimate $K(r)$ is not defined for $r > r_k^*$. To define $K(r)$ it is common to treat the largest censored residual as a noncensored one. Clearly, this procedure exacerbates the outlier's effect.

To overcome these problems, we propose to modify the S estimate of Salibian-Barrera and Yohai replacing the KM estimate with a parametric estimate of the error distribution. We assume that the error model belongs to a location-scale family of asymmetric or symmetric distributions. Examples are the Log–Weibull and the Gaussian distributions. Monte Carlo results indicate that the new parametric S estimate has a higher degree of robustness than the original proposal because it does not suffer the drawback of the KM estimate.

Finally, we use the new S estimate as the first step of an adaptive weighted maximum likelihood (WML) procedure which extends the adaptive truncated maximal likelihood estimate of Marazzi and Yohai (2004) to censored regression models. The adaptive WML attains full relative efficiency with respect to ML at the model, while maintaining the bdp of the initial S estimate.

In Section 2 we introduce the model and some notations. Section 3 describes the proposed estimates. In Section 4 we discuss their bdp and in Section 5 their asymptotic behavior. A resampling algorithm to compute the parametric S estimate is described in Section 6. Section 7 summarizes some empirical results described in more details in a technical report (Locatelli et al., 2010, referenced as LMY in the following). Section 8 illustrates the new procedure with two real data sets and compares it with the estimates of Salibian-Barrera and Yohai (2008), Jin et al. (2003), and Zeng and Lin (2007). One of the examples concerns modeling of hospital length of stay. A discussion section concludes the paper. Proofs are given in LMY.

2. Model, notations, and expectations

We consider an accelerated failure time model for n pairs of variates $(\mathbf{x}_i, y_i)$

$$y_i = \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_i + \sigma_0 u_i, \quad i = 1, \ldots, n, \tag{1}$$

where y_i represents the duration on the logarithmic scale. The errors u_i are i.i.d. with cdf F and independent of $\mathbf{x}_i$; $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ is an unknown vector of coefficients, the first component being an intercept term and σ_0 an unknown scale parameter. The distribution of the carriers $\mathbf{x}_i$ is unknown. We consider single censoring on the right, i.e., the true value of y_i is not observed. Instead, the censored variable $y_i^* = \min(y_i, v_i)$ is observed, where $v_1, \ldots, v_n$ are i.i.d. censoring log-times, which are independent of the y_i's. We define the indicator $\delta_i = 1$ if $y_i^* = y_i$ and $\delta_i = 0$ if $y_i^* = v_i$. Thus, δ_i indicates whether observation i is complete ($\delta_i = 1$) or censored ($\delta_i = 0$).

In practice, we use a hypothetical model cdf F_0 as an approximation of the real error distribution F. We assume that $F_0(z) = F_{0,1}(z)$, where $F_{0,1}$ is the standard member of a parametric location-scale family of asymmetric (or symmetric) distributions with cdf $F_{\mu,\sigma}(z) = F_{0,1}((z-\mu)/\sigma)$. We denote by $f_{\mu,\sigma}$ and f_0 the densities of $F_{\mu,\sigma}$ and F_0, respectively, and by

$H_{\beta,\sigma}$ the corresponding cdf of $(y, \mathbf{x})$ when $(\boldsymbol{\beta}, \sigma)$ are the true parameters. Examples of location-scale error models are the Gaussian model with density

$$f_{\mu,\sigma}(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left((z-\mu)^2/(2\sigma^2)\right), \quad -\infty < z < \infty,$$

and the Log–Weibull model with density

$$f_{\mu,\sigma}(z) = \frac{1}{\sigma} \exp\left[\left((z-\mu)/\sigma\right) - \exp\left((z-\mu)/\sigma\right)\right], \quad -\infty < z < \infty.$$

Other models are mentioned in Lawless (2003, Chaps. 1.3.6 and 6). The negative log-likelihood function is denoted by $\rho_{\mu,\sigma}(u) = -\ln f_{\mu,\sigma}(u)$. We put $\rho_0(u) = \rho_{0,1}(u)$ and assume that this function is convex and that $E_0[\rho_0(u)] < \infty$, where E_0 denotes expectation under F_0. Finally let

$$\psi_0(u) = \rho_0'(u) = -f_0'(u)/f_0(u) \quad \text{and} \quad \psi_1(u) = u\psi_0(u). \tag{2}$$

The triplets $(y_i^*, \mathbf{x}_i, \delta_i)$ are observed and we want to estimate $(\boldsymbol{\beta}_0, \sigma_0)$. Since σ_0 is necessary for the computation of the conditional expectation of the response $\exp(y)$, we are not going to treat σ_0, as often is the case, as a nuisance parameter.

We note that, for any measurable function $h(y, \mathbf{x})$, $E[h(y, \mathbf{x})|y^*, \mathbf{x}, \delta = 1] = h(y^*, \mathbf{x})$, $E[h(y, \mathbf{x})|y^*, \mathbf{x}, \delta = 0] = E[h(y, \mathbf{x})|y > y^*, \mathbf{x}]$, and thus

$$E\left[h(y, \mathbf{x})|y^*, \mathbf{x}, \delta\right] = \delta h(y^*, \mathbf{x}) + (1-\delta)E\left[h(y, \mathbf{x})|y > y^*, \mathbf{x}\right]. \tag{3}$$

In particular, under the model, we have

$$E_{\beta,\sigma}\left[h(y, \mathbf{x})\,|y > y_i^*, \mathbf{x}_i\right] = \frac{\int_{(y_i^* - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})/\sigma}^{\infty} h(\sigma u + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \mathbf{x}_i) f_0(u)\mathrm{d}u}{1 - F_0((y_i^* - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})/\sigma)}. \tag{4}$$

Using the model, we define an *empirical cdf for censored observations* $(y_i^*, \mathbf{x}_i)$ as

$$H_{n,\beta,\sigma}(z, \mathbf{z}) = \frac{1}{n}\sum_{i=1}^{n} E_{\beta,\sigma}\left[I(y \le z)\,|y_i^*, \mathbf{x}_i, \delta_i\right] I(\mathbf{x}_i \le \mathbf{z}). \tag{5}$$

When there is no censoring, $H_{n,\beta,\sigma}(z, \mathbf{z})$ coincides with the usual empirical cdf

$$H_n(z, \mathbf{z}) = \frac{1}{n}\sum_{i=1}^{n} I(y_i \le z)\, I(\mathbf{x}_i \le \mathbf{z}). \tag{6}$$

We note that $E_{\beta,\sigma}\left[I(y \le z)\,|y_i^*, \mathbf{x}_i, \delta_i\right] I(\mathbf{x}_i \le \mathbf{z})$ are i.i.d. random variables and that

$$E_{\beta,\sigma}\left[E_{\beta,\sigma}\left\{I(y \le z)\,|y_i^*, \mathbf{x}_i, \delta_i\right\} I(\mathbf{x}_i \le \mathbf{z})\right] = H_{\beta,\sigma}(z, \mathbf{z}).$$

Therefore, by the Law of the Large Numbers, $H_{n,\beta_0,\sigma_0}(z, \mathbf{z})$ is a consistent estimate of $H_{\beta_0,\sigma_0}(z, \mathbf{z})$. In general, for any measurable function $h(y, \mathbf{x})$, we have $\lim_{n\to\infty} E_{n,\beta,\sigma}[h(y, \mathbf{x})] = E_{\beta,\sigma}[h(y, \mathbf{x})]$ a.s., where $E_{n,\beta,\sigma}$ denotes expectation under $H_{n,\beta,\sigma}$.

We finally note that the ML equations of the estimates of $\boldsymbol{\beta}_0$ and σ_0 (Lawless, 2003, p. 293) can be written as follows,

$$E_{n,\beta,\sigma}\left[\psi_0\left(\left(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\right)/\sigma\right)\mathbf{x}\right] = \mathbf{0}, \tag{7}$$

$$E_{n,\beta,\sigma}\left[\psi_1\left(\left(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\right)/\sigma\right)\right] = 1. \tag{8}$$

Remark 1. The conditional expectation in (4) can also be estimated using the Kaplan–Meier distribution of the residuals $y_i^* - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$, an idea used by Buckley and James (1979) to extend the least squares estimate to censored observations and by Salibian-Barrera and Yohai (2008) to define nonparametric high bdp regression estimates for censored data.

3. The proposed estimates

3.1. The initial estimate

The initial step of the proposed procedure is the computation of a high bdp S estimate. This class of regression estimates was introduced by Rousseeuw and Yohai (1987) for noncensored data. S estimates can be calibrated so that they have a bdp of 50%. However, when this occurs, the S estimates are inefficient. For details about S estimates see Maronna et al. (2007). In this section, we extend the S estimates for censored observations proposed by Salibian-Barrera and Yohai (2008) to the case where data follow the parametric model of Section 2.

Suppose that ρ is a given function $\mathbb{R} \to \mathbb{R}^+$ satisfying the following conditions:

A: (i) $\rho(0) = 0$; (ii) ρ is even; (iii) if $|z_1| < |z_2|$, then $\rho(z_1) \leq \rho(z_2)$; (iv) ρ is bounded; (v) ρ is continuous at 0.

For example, ρ is a member of the Tukey's biweight family

$$\rho^T(z, k) = \begin{cases} 3(z/k)^2 - 3(z/k)^4 + (z/k)^6 & \text{if } |z| \leq k, \\ 1 & \text{if } |z| > k, \end{cases} \tag{9}$$

where k is a user chosen tuning parameter. For any μ, let the function $S(\mu)$ be the M scale (Huber, 1981) $E_0[\rho((u-\mu)/S(\mu))] = b$, where expectation is based on F_0 and $b = \max \rho(z)/2$. Thus, $b = 0.5$ for Tukey's biweight. In the following, we will assume that there exists a unique μ_0 such that $\mu_0 = \arg\min_\mu S(\mu)$. This holds, for example, under the following conditions (Mizera, 1993):

(a) f_0 is strictly unimodal,
(b) for any $c \geq \inf \rho(u)$, the function $\log(\max(c - \rho(x), 0))$ is strictly concave.

Then, we define the S scale of F_0 as $s_0 = S(\mu_0)$. Without loss of generality, we assume that $s_0 = 1$ and $\mu_0 = 0$ (if not, we replace $\rho(u)$ with $\rho((u-\mu_0)/s_0)$). For any $\boldsymbol{\beta} \in \mathbb{R}^p$, let the residual scale $s(\boldsymbol{\beta})$ be defined by

$$E_0\left[\rho\left(\left(y - \mathbf{x}^T\boldsymbol{\beta}\right)/s(\boldsymbol{\beta})\right)\right] = b.$$

Marazzi et al. (2009, Lemma 2) prove that $\sigma_0 = \min_{\boldsymbol{\beta}} s(\boldsymbol{\beta})$ and $\boldsymbol{\beta}_0 = \arg\min_{\boldsymbol{\beta}} s(\boldsymbol{\beta})$. Then, for the noncensored case, the S estimate $(\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n)$ is defined by

$$\tilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} s_n(\boldsymbol{\beta}), \qquad \tilde{\sigma}_n = s_n(\tilde{\boldsymbol{\beta}}_n),$$

where $s_n(\boldsymbol{\beta})$ solves

$$E_{H_n}\left[\rho\left(\left(y - \mathbf{x}^T\boldsymbol{\beta}\right)/s_n(\boldsymbol{\beta})\right)\right] = b. \tag{10}$$

Since H_n is a consistent estimate of $H_{\boldsymbol{\beta}_0,\sigma_0}$, $(\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n)$ is consistent for $\left(\boldsymbol{\beta}_0, \sigma_0\right)$.

Consider now the censored case; here, H_n is not available. Then, we define $s_n(\boldsymbol{\gamma})$ by

$$E_{n,\boldsymbol{\gamma},s_n(\boldsymbol{\gamma})}\left[\rho\left(\left(y - \mathbf{x}^T\boldsymbol{\gamma}\right)/s_n(\boldsymbol{\gamma})\right)\right] = b \tag{11}$$

and $\bar{\boldsymbol{\beta}}_n$ by $\bar{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\gamma}} s_n(\boldsymbol{\gamma})$. Since $s_n(\boldsymbol{\gamma})$ is a consistent estimate of σ_0 only when $\boldsymbol{\gamma} = \boldsymbol{\beta}_0$, the estimate $\bar{\boldsymbol{\beta}}_n$ is not consistent. Therefore, we have to proceed in a different way. Assume for one moment that we know $\boldsymbol{\beta}_0$. Then, we can find $s_n^*(\boldsymbol{\gamma})$ by solving

$$E_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)}\left[\rho\left(\left(y - \mathbf{x}^T\boldsymbol{\gamma}\right)/s_n^*(\boldsymbol{\gamma})\right)\right] = b \tag{12}$$

and define the "pseudo estimate" $\bar{\boldsymbol{\beta}}_n^*$ by

$$\bar{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{\gamma}} s_n^*(\boldsymbol{\gamma}). \tag{13}$$

Then, since $H_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)}$ is a consistent estimate of $H_{\boldsymbol{\beta}_0,\sigma_0}$ (Section 2), we have that $\bar{\boldsymbol{\beta}}_n^* \to \boldsymbol{\beta}_0$. Unfortunately, $\bar{\boldsymbol{\beta}}_n^*$ is clearly not a feasible estimate but it suggests the following procedure to find a feasible and consistent estimate.

For any $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\boldsymbol{\gamma} \in \mathbb{R}^p$, let $S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ be defined by

$$E_{n,\boldsymbol{\beta},s_n(\boldsymbol{\beta})}\left[\rho\left(\left(y - \mathbf{x}^T\boldsymbol{\gamma}\right)/S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})\right)\right] = b, \tag{14}$$

where $s_n(\boldsymbol{\beta})$ is defined in (11). Let

$$\tilde{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\gamma}} S(\boldsymbol{\beta}, \boldsymbol{\gamma}). \tag{15}$$

We note that $\tilde{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_0) = \bar{\boldsymbol{\beta}}_n^*$, where $\bar{\boldsymbol{\beta}}_n^*$ is defined in (13) and therefore $\tilde{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_0) \to \boldsymbol{\beta}_0$, i.e., $\boldsymbol{\beta}_0$ is an "almost fixed point" of $\tilde{\boldsymbol{\gamma}}_n$. Then, it is natural to define the S estimate $\tilde{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}_0$ by the fixed point equation

$$\tilde{\boldsymbol{\gamma}}_n(\tilde{\boldsymbol{\beta}}_n) = \tilde{\boldsymbol{\beta}}_n \tag{16}$$

and the S estimate $\tilde{\sigma}_n$ of σ_0 by $\tilde{\sigma}_n = s_n(\tilde{\boldsymbol{\beta}}_n)$.

The estimate $(\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n)$ is a parametric version of the S estimate for censored data introduced by Salibian-Barrera and Yohai (2008). Differentiating (14) with respect to $\boldsymbol{\gamma}$ and using (16), we obtain the following system of estimating equations for $(\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n)$:

$$E_{n,\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}[\psi((y - \mathbf{x}^T\tilde{\boldsymbol{\beta}}_n)/\tilde{\sigma}_n)\mathbf{x}] = 0, \tag{17}$$

$$E_{n,\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}[\rho((y - \mathbf{x}^T\tilde{\boldsymbol{\beta}}_n)/\tilde{\sigma}_n)] = b. \tag{18}$$

Since the function ψ is redescending, Eqs. (17) and (18) may have more than one solution, especially when the sample contains outliers. However, not all these solutions satisfy (16). Therefore it is important, when solving these equations using a numerical iterative algorithm, to start with a robust estimate unaffected by outliers. Such a procedure is described in Section 6.

Remark 2. To avoid existence problems with the solution of (16), we can alternatively define $\tilde{\boldsymbol{\beta}}_n$ by

$$\tilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \left| \left(\tilde{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) - \boldsymbol{\beta}\right)^{\mathrm{T}} A_n \left(\tilde{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) - \boldsymbol{\beta}\right) \right|,$$

where A_n is any robust equivariant estimator of the covariance matrix of the explanatory variables. The matrix A_n is needed to maintain the affine equivariance of the estimator.

3.2. The outlier rejection rule

We now suppose that $(\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n)$ is an "initial" high bdp and consistent but maybe inefficient estimate, such as the parametric S estimate defined above. To obtain a "final" estimate that keeps the bdp of the initial estimate but which is highly efficient, we have to reject the outliers. In the following, we define a rejection rule based on a proposal by Marazzi and Yohai (2004).

We want to reject observations whose likelihoods under the initial model are smaller than a given cut-off value. For this purpose, we consider the cdf M_0 of the negative log-likelihood $l = \rho_0(u)$ under the model and the estimate of M_0 given by

$$M_{n,\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}(z) = \frac{1}{n}\sum_{i=1}^{n}[\delta_i I\left(l_i^* \le z\right) + (1-\delta_i)P_{\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}(l \le z | y > y_i^*)],$$

where $l_i^* = \rho_0(\tilde{r}_i^*)$ and $\tilde{r}_i^* = (y_i^* - \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_n)/\tilde{\sigma}_n$. For simplicity, we write M_n in place of $M_{n,\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}$. Using the argument of Section 2, one can show that M_n is a consistent estimate of M_0. A fixed cut-off ζ on the likelihood scale can be defined as a large quantile of M_0, e.g., $\zeta = M_0^{-1}(0.99)$. To define an adaptive cut-off ϑ_n, that depends on the observed degree of contamination, we compare the tails of M_0 and M_n. Let $M_{n,\vartheta}$ denote M_n truncated at ϑ, i.e.,

$$M_{n,\vartheta}(z) = \begin{cases} M_n(z)/M_n(\vartheta) & \text{if } z \le \vartheta, \\ 1 & \text{otherwise.} \end{cases} \tag{19}$$

We look for the largest ϑ such that $M_{n,\vartheta}(z) \geqq M_0(z)$ for all $z \ge \zeta$, i.e.,

$$\vartheta_n = \sup\{\vartheta \mid M_{n,\vartheta}(z) \geqq M_0(z) \text{ for all } z \ge \zeta\}. \tag{20}$$

Note that $\vartheta_n \ge \zeta$. As in Gervini and Yohai (2002), one can prove that if the sample does not contain outliers, $\vartheta_n \to \infty$ a.s.

3.3. The final estimate

Let $\omega(z)$ be a function satisfying conditions B:

B: (i) $\omega(z)$ is nonincreasing; (ii) $\lim_{z\to-\infty}\omega(z) = 1$; (iii) $\omega(z) = 0$ for $z > 0$.
For example, let $c > 0$ and consider the function

$$\omega(z) = \rho^T(z, c) \cdot I(z \le 0), \tag{21}$$

where $\rho^T(z, c)$ is in the biweight family (9). Then, define the weight function

$$w_{\vartheta_n}(z) = \omega\left(\rho_0(z) - \vartheta_n\right), \tag{22}$$

where ϑ_n is a fixed or adaptive cut-off for outlier rejection. When $\vartheta_n \to \infty$, $w_{\vartheta_n}(u) \to 1$ for all u. The "final" estimate $(\hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n)$ is defined by the equations

$$E_{n,\hat{\boldsymbol{\beta}}_n,\hat{\sigma}_n}[w_{\vartheta_n}((y - \mathbf{x}^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_n)/\tilde{\sigma}_n)\psi_0((y - \mathbf{x}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_n)/\hat{\sigma}_n)\mathbf{x}] = \mathbf{0}, \tag{23}$$

$$E_{n,\hat{\boldsymbol{\beta}}_n,\hat{\sigma}_n}[w_{\vartheta_n}((y - \mathbf{x}^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_n)/\tilde{\sigma}_n)\psi_1((y - \mathbf{x}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_n)/\hat{\sigma}_n)] = b_{\vartheta_n}, \tag{24}$$

where $b_{\vartheta_n} = E_0\left[w_{\vartheta_n}(u)\psi_1(u)\right]$, and ψ_0, ψ_1 are given in (2).

The estimate $(\hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n)$ is a natural extension of the truncated maximum likelihood estimate for noncensored observations proposed in Marazzi and Yohai (2004), which uses $\omega(z) = I(z \le 0)$ ("hard rejection"). In the adaptive case, where $\vartheta_n \to \infty$ and $b_{\vartheta_n} \to 1$, the Eqs. (23)–(24) approach to the ML equations (7)–(8). In the nonadaptive case, where ϑ_n is fixed, the estimator $(\hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n)$ will be called *weighted maximum likelihood estimate* or *WML estimate*; in the adaptive case, it will be referred to as the *adaptive WML estimate*.

4. Breakdown point

Intuitively, the bdp of an estimator is the proportion of incorrect observations (i.e. arbitrarily large observations) the estimator can handle before giving an arbitrarily large result. Formally, given a sample $\mathbf{Z}_n$ of size n, the finite-sample bdp of an estimator $T_n = T_n(\mathbf{Z}_n)$ is defined (Maronna et al., 2007) as:

$$\epsilon_n^*(T_n, \mathbf{Z}_n) = \min_{1\le k\le n}\{k/n : \sup \|T_n(\mathbf{Z}_{k,n}^*) - T_n(\mathbf{Z}_n)\| = \infty\},$$

where the supremum is taken over all possible samples $\mathbf{Z}_{k,n}^*$ which are obtained by replacing k observations from $\mathbf{Z}_n$ with arbitrary values and $\|\cdot\|$ is the L_2 norm. Let $\mathbf{Z}_n = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be a sample from a censored linear regression model, where $\mathbf{z}_i = (y_i^*, \mathbf{x}_i, \delta_i)$, $\mathbf{x}_i \in \mathbb{R}^p$. Assume that the rank of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is p and let $q = \max_{\|\boldsymbol{\beta}\|=1} \#\{i : \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i = 0\}$. The following theorem gives a lower bound of the bdp of the S estimate and the adaptive (or the non adaptive) WML estimate.

Theorem 1. *We assume that ρ satisfies conditions* A *and $b = \max \rho(z)/2$.* (a) *A lower bound for the finite sample bdp of $\tilde{\boldsymbol{\beta}}_n$ and $\tilde{\sigma}_n$ is $\epsilon_1^* = 0.5 - q/n - m/n$, where m is the number of censored observations in the sample.* (b) *The finite sample bdp of $\hat{\sigma}_n$ and $\hat{\boldsymbol{\beta}}_n$, starting with $\tilde{\sigma}_n$ and $\tilde{\boldsymbol{\beta}}_n$, are larger or equal than ϵ_1^*.*

Remark 3. Salibian-Barrera and Yohai (2008) show that ϵ_1^* is also a lower bound of the bdp of their nonparametric S estimate. In LMY, we show that, for a given $\boldsymbol{\beta}$, a lower bound of the bdp of the initial scale estimate $s_n(\boldsymbol{\beta})$ defined by (11) is $\epsilon_2^* = 0.5 - 0.5m/n$. This bound is larger than ϵ_1^*. Unfortunately we cannot prove that the bdp of $\tilde{\sigma}_n$ and $\tilde{\boldsymbol{\beta}}_n$ is larger than ϵ_2^*. However, the numerical results in Section 7.2 support this conjecture.

5. Asymptotic behavior

We first discuss the Fisher consistency of the adaptive WML estimate. We consider the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \theta_2, \boldsymbol{\theta}_3, \theta_4) = (\boldsymbol{\beta}, \sigma, \boldsymbol{\beta}, \sigma)$ where $\boldsymbol{\beta}$ and σ are duplicated. The true value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \sigma_0, \boldsymbol{\beta}_0, \sigma_0)$. According to (17), (18), (23) and (24), the estimate $\hat{\boldsymbol{\theta}}_n = (\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n, \hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n)$ is defined by

$$\sum_{i=1}^{n} \boldsymbol{\eta}(y_i^*, \mathbf{x}_i, \delta_i, \hat{\boldsymbol{\theta}}_n, \phi_n) = \mathbf{0},$$

where $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \eta_2, \boldsymbol{\eta}_3, \eta_4)^{\mathrm{T}}$, $\phi_n = 1/\vartheta_n$, and

$$\boldsymbol{\eta}_1(y^*, \mathbf{x}, \delta, \boldsymbol{\theta}, \phi) = E_{\boldsymbol{\theta}_1,\theta_2}\left[\psi\left(\left(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\theta}_1\right)/\theta_2\right)|y^*, \mathbf{x}, \delta\right]\mathbf{x},$$

$$\eta_2(y_i^*, \mathbf{x}_i, \delta_i, \boldsymbol{\theta}, \phi) = E_{\boldsymbol{\theta}_1,\theta_2}\left[\rho\left(\left(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\theta}_1\right)/\theta_2\right)|y^*, \mathbf{x}, \delta\right] - b,$$

$$\boldsymbol{\eta}_3(y^*, \mathbf{x}, \delta, \boldsymbol{\theta}, \phi) = E_{\boldsymbol{\theta}_3,\theta_4}\left[w_{1/\phi}\left(\left(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\theta}_1\right)/\theta_2\right)\psi_0\left(\left(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\theta}_3\right)/\theta_4\right)|y^*, \mathbf{x}, \delta\right]\mathbf{x},$$

$$\eta_4(y_i^*, \mathbf{x}_i, \delta_i, \boldsymbol{\theta}, \phi) = E_{\boldsymbol{\theta}_3,\theta_4}\left[w_{1/\phi}\left(\left(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\theta}_1\right)/\theta_2\right)\psi_1\left(\left(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\theta}_3\right)/\theta_4\right)|y^*, \mathbf{x}, \delta\right] - b_{1/\phi}.$$

Theorem 2. *Assume that the function ρ satisfies conditions* A *and that $\mu_0 = 0$ and $s_0 = 1$. Then, $\hat{\boldsymbol{\theta}}_n$ is Fisher consistent for $\boldsymbol{\theta}_0$, i.e.*

$$E\left[\boldsymbol{\eta}(y^*, \mathbf{x}, \delta, \boldsymbol{\theta}_0, \phi)\right] = \mathbf{0} \quad \textit{for all } \phi. \tag{25}$$

A complete proof of consistency would require to show that the S estimate of $\boldsymbol{\beta}_0$ defined in (16) is unique, i.e., that if $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ then $\tilde{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})$ remains asymptotically away from $\boldsymbol{\beta}_0$. This seems a difficult problem and is still open. However, in all our numerical experiments this property holds. A simple illustration is provided in Section 7.3.

We now put $u = (y - \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x})/\sigma_0$,

$$\boldsymbol{\eta}_{\boldsymbol{\theta}}(y^*, \mathbf{x}, \delta, \boldsymbol{\theta}, \phi) = \partial\boldsymbol{\eta}(y^*, \mathbf{x}, \delta, \boldsymbol{\theta}, \phi)/\partial\boldsymbol{\theta},$$

$$R_0 = E\left[\boldsymbol{\eta}(u, \mathbf{x}, \delta, (\mathbf{0}, 1, \mathbf{0}, 1), \phi_0)\boldsymbol{\eta}(u, \mathbf{x}, \delta, (\mathbf{0}, 1, \mathbf{0}, 1), \phi_0)^{\mathrm{T}}\right],$$

$$Q_0 = E\left[\boldsymbol{\eta}_{\boldsymbol{\theta}}(u, \mathbf{x}, \delta, (\mathbf{0}, 1, \mathbf{0}, 1), \phi_0)\right].$$

The following theorem states the asymptotic normality of the WLM estimate. The notation "$\to_P$" means convergence in probability.

Theorem 3. *Suppose that:* (i) *the function ρ satisfies conditions* A*;* (ii) *the functions ρ and ρ_0 are two times continuously differentiable;* (iii) *the function ω satisfies conditions* B*;* (iv) *the function ω is continuously differentiable;* (v) *the vector $\mathbf{x}$ has second order moments;* (vi) $\hat{\boldsymbol{\theta}}_n \to_P \boldsymbol{\theta}_0$*;* (vii) $\phi_n \to_P \phi_0$ (viii) *the matrix Q_0 is nonsingular. Then,* (a) $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ *is asymptotically normal, with mean $\mathbf{0}$ and covariance matrix $\sigma_0^2 Q_0^{-1} R_0 Q_0^{-1\mathrm{T}}$;* (b) *If $\phi_0 = 0$, the asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{\sigma}_n - \sigma_0)$ is the same as the one of the maximum likelihood estimate defined by Eqs. (7) and (8).*

Remark 4. If ϑ_n is defined by (20) then, under the model, $\vartheta_n \to \infty$ a.s. and $\phi_0 = 0$. Thus, (b) holds.

6. Computation

For a given initial estimate and a given cut off-value, the final estimate can be computed with the help of standard numerical tools for solving nonlinear equations. To compute the initial S estimate, we propose the following resampling algorithm based on Salibian-Barrera and Yohai (2008) and Salibian-Barrera and Yohai (2006). Without loss of generality, we assume that $s_0 = 1$.

Phase 1 (*Subsampling and concentration*). Draw N random subsamples of size $q \geq p$ of noncensored observations. Let $\boldsymbol{\beta}_0^{(1)}, \ldots, \boldsymbol{\beta}_0^{(N)}$ be the coefficients of the corresponding ML fits. Compute the scales $s_0^{(j)} = s_n(\boldsymbol{\beta}_0^{(j)})$ $(j = 1, \ldots, N)$ according to

$$E_{n,\boldsymbol{\beta}_0^{(j)}, s_n(\boldsymbol{\beta}_0^{(j)})}[\rho((y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0^{(j)})/s_n(\boldsymbol{\beta}_0^{(j)}) - \mu_0)] = b. \tag{26}$$

For each pair $(\boldsymbol{\beta}_0^{(j)}, s_0^{(j)})$ $(j = 1, \ldots, N)$ calculate the "residuals"

$$r_i^{(j)} = \delta_i(y_i^* - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0^{(j)})/s_0^{(j)} + (1 - \delta_i)\int_{(y_i^* - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0^{(j)})/s_0^{(j)}}^{\infty} uf_0(u)\mathrm{d}u, \quad i = 1, \ldots, n,$$

set $y_i^{(j)} = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0^{(j)} + s_0^{(j)}r_i^{(j)}$, $i = 1, \ldots, n$, and compute the (noncensored) truncated maximum likelihood fit $\boldsymbol{\beta}^{(j)}$ to $(\mathbf{x}_i, y_i^{(j)})$, according to Marazzi and Yohai (2004), where the rejection fraction is set to 50%. The coefficient vectors $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(N)}$ are the "candidates".

Phase 2 (*Selection*). Take $\boldsymbol{\beta}^{(k_j)}$ as an approximate value of $\tilde{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}^{(j)})$, where

$$k_j = \arg\min_{1 \leq k \leq N} S_n(\boldsymbol{\beta}^{(j)}, \boldsymbol{\beta}^{(k)}), \tag{27}$$

and $S_n(\boldsymbol{\beta}^{(j)}, \boldsymbol{\beta}^{(k)})$ is the solution of

$$E_{n,\boldsymbol{\beta}^{(j)}, s_n(\boldsymbol{\beta}^{(j)})}\left[\rho\left(\left(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}^{(k)}\right)/S_n(\boldsymbol{\beta}^{(j)}, \boldsymbol{\beta}^{(k)}) - \mu_0\right)\right] = b. \tag{28}$$

Select a tentative S-estimate as $\tilde{\boldsymbol{\beta}}_n^* = \boldsymbol{\beta}^{(j^*)}$ and $\tilde{s}_n^* = s_n(\tilde{\boldsymbol{\beta}}_n^*)$, where $j^* = \arg\min_{1 \leq j \leq N} \|\boldsymbol{\beta}^{(k_j)} - \boldsymbol{\beta}^{(j)}\|$.

The selection can be accelerated, avoiding computing the N^2 values $S_n(\boldsymbol{\beta}^{(j)}, \boldsymbol{\beta}^{(k)})$, as follows.

Step 1. Compute k_1 using (27) with $j = 1$. Put $j_1^* = 1$ and $\lambda_1 = \|\boldsymbol{\beta}^{(k_1)} - \boldsymbol{\beta}^{(1)}\|$.

Step 2. Divide the N candidates $\boldsymbol{\beta}^{(k)}$ into two sets:

$$A = \{\boldsymbol{\beta}^{(k)} \mid \|\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}^{(2)}\| \leq \lambda_1\} \quad \text{and} \quad B = \{\boldsymbol{\beta}^{(k)} \mid \|\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}^{(2)}\| > \lambda_1\}.$$

Let $F_2 = \min_{\boldsymbol{\beta}^{(k)} \in A} S_n(\boldsymbol{\beta}^{(2)}, \boldsymbol{\beta}^{(k)})$. Then compute, once at the time, $S_n(\boldsymbol{\beta}^{(2)}, \boldsymbol{\beta}^{(k)})$ for $\boldsymbol{\beta}^{(k)} \in B$. Suppose that for some $\boldsymbol{\beta}^{(k)} \in B$, $S(\boldsymbol{\beta}^{(2)}, \boldsymbol{\beta}^{(k)}) < F_2$. Then, $\|\boldsymbol{\beta}^{(k_2)} - \boldsymbol{\beta}^{(2)}\| \geq \lambda_1$ and the remaining values of B can therefore be ignored. In this case, put $j_2^* = j_1^*$, and $\lambda_2 = \lambda_1$. If $S(\boldsymbol{\beta}^{(2)}, \boldsymbol{\beta}^{(k)}) \geq F_2$ for all $\boldsymbol{\beta}^{(k)} \in B$, put $\lambda_2 = \|\boldsymbol{\beta}^{(k_2)} - \boldsymbol{\beta}^{(2)}\|$ and $j_2^* = 2$.

Step 3 to Step N. Proceed as in step 2, replacing $\boldsymbol{\beta}^{(2)}$ by $\boldsymbol{\beta}^{(3)}$, λ_1 by λ_2, etc. At the end of the procedure put $j^* = j_N^*$, $\tilde{\boldsymbol{\beta}}_n^* = \boldsymbol{\beta}^{(j^*)}$ and $\tilde{s}_n^* = s_n(\tilde{\boldsymbol{\beta}}_n^*)$.

Phase 3 (*Refinement*). Solve

$$\sum_{i=1}^{n} E_{\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n}[\psi((y - \mathbf{x}^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_n)/\tilde{\sigma}_n - \mu_0)\mathbf{x}|y_i^*, \mathbf{x}_i, \delta_i] = 0,$$

$$\sum_{i=1}^{n} E_{\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n}[\psi((y - \mathbf{x}^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_n)/\tilde{\sigma}_n - \mu_0)\mathbf{x}|y_i^*, \mathbf{x}_i, \delta_i] = 0,$$

using an iterative algorithm starting at $(\tilde{\boldsymbol{\beta}}_n^*, \hat{s}_n^*)$. The vector $(\tilde{\boldsymbol{\beta}}_n, \tilde{\sigma}_n)$ defines the refined tentative S estimate.

7. Empirical results

7.1. Monte Carlo simulation

We performed a Monte Carlo study for the simple regression model

$$\begin{aligned} &y_i = \alpha_0 + \beta_0 x_i + \sigma_0 u_i, \quad i = 1, \ldots, n, \\ &v_i \sim N(\mu, 1), \qquad x_i \sim N(0, 1), \qquad u_i \sim F_0, \end{aligned} \tag{29}$$

with $\alpha_0 = 0$, $\beta_0 = 1$ and $\sigma_0 = 1$. We considered both the standard Gaussian and the standard Log–Weibull distributions of

Table 1
Simulated root mean square errors at the nominal Gaussian model.

Parameter	Estimate	Sample size			
		100	200	500	1000
Intercept	S_{NP}	0.231	0.155	0.100	0.071
	MM	0.136	0.092	0.059	0.040
	S_P	0.187	0.129	0.082	0.059
	WML	0.118	0.083	0.055	0.037
	ML	0.116	0.082	0.054	0.036
Slope	S_{NP}	0.222	0.158	0.094	0.066
	MM	0.151	0.110	0.064	0.045
	S_P	0.211	0.156	0.097	0.066
	WML	0.123	0.092	0.055	0.038
	ML	0.124	0.090	0.054	0.037
Scale	S_{NP}	0.122	0.087	0.054	0.038
	MM	0.122	0.087	0.054	0.038
	S_P	0.116	0.083	0.052	0.037
	WML	0.097	0.070	0.044	0.031
	ML	0.090	0.063	0.040	0.029

Table 2
Simulated maximum root mean square errors under point contamination at (x_0, mx_0), $\epsilon = 10\%$, $n = 100$, and Gaussian errors. The maximum has been computed over $m \in (1.0, 1.5, \ldots, 5.5, 6.0)$.

Parameter	$x_0 = 1$					$x_0 = 10$				
	S_{NP}	MM	S_P	WML	ML	S_{NP}	MM	S_P	WML	ML
Intercept	0.600	0.449	0.456	0.417	1.007	0.466	0.386	0.351	0.310	2.399
Slope	0.646	0.510	0.557	0.434	0.998	0.760	0.756	0.664	0.652	4.661
Scale	0.342	0.342	0.268	0.304	1.018	0.343	0.343	0.268	0.122	3.357

the error term. The mean μ of the censoring variables v_i was chosen in order to have a probability of censoring of around 0.35. For normal errors, we have $\mu = 0.668$; in the Log–Weibull case, $\mu = 0.213$. The initial S estimates were based on the Tukey's biweight ρ-function with $k = 1.548$ in the Gaussian case and $k = 1.718$ in the Log–Weibull case. With these values of k, we have $s_0 = 1$. In addition, $\mu_0 = 0$ in the Gaussian case and $\mu_0 = -0.135$ in the Log–Weibull case. The S estimates were computed using the algorithm described in Section 6 with $N = 100$ and $q = 4$. The fixed cut-off on the negative log-likelihood scale was set to $\zeta = M_0^{-1}(0.99)$. The weight function (21) was used in the calculation of the adaptive WML estimator with $c = 0$, so that the adaptive WML estimate behaves like a truncated maximum likelihood estimate. All simulations were based on 1000 samples.

In the tables, ML denotes the ML estimate, S_{NP} and MM denote the nonparametric S estimate and the final MM estimate of Salibian-Barrera and Yohai (2008) respectively, S_P the parametric S estimate defined in this paper, and WML the adaptive weighted maximum likelihood estimate.

Table 1 shows the simulated root mean square errors (rMSE) of the intercept, slope, and scale estimates at the nominal Gaussian model. As expected, WML attains a much higher performance than the initial S_P, approaching the ML values when n increases. The efficiency of S_P is higher than the one of S_{NP} and WML performs better than MM.

In order to investigate the behavior of the estimates in the presence of outliers, the simulated samples were contaminated with a fixed fraction $\epsilon = 10\%$ of outlying observations at (x_0, mx_0) for $x_0 = 1$ (low leverage point) and $x_0 = 10$ (high leverage point) and m varying over the regular grid 1.0, 1.5, ..., 6.0. Detailed results of this simulation can be found in a LMY. Table 2 reports the maximum rMSEs (maxrMSE) over the grid of the estimated parameters for $n = 100$. In general, the maxrMSE of the adaptive WML estimates are smaller than the maxrMSEs of the other estimates. Moreover, S_P and WML perform better than S_{NP} and MM. For Log–Weibull errors, we obtained similar results reported in LMY. Since S_{NP} and MM require a symmetric error distribution, they were not included this case.

7.2. Breakdown point

In order to compare the bdp of the parametric and the non-parametric S estimates, we took a sample of size $n = 1000$ from the model $y_i = \alpha_0 + \beta_0 x_i + \sigma_0 u_i$, $i = 1, \ldots, n$, $v_i \sim N(\mu, 1)$, $x_i \sim N(0, 1)$ and $u_i \sim N(0, 1)$. We took $\alpha_0 = 0$, $\beta_0 = 1$, $\sigma_0 = 1$, and $\mu = 0.44$, so that the censoring fraction was around 40%. In a first experiment, an increasing number t of noncensored observations was replaced by t noncensored outliers at (x_0, y_0) with $x_0 = 1$ and y_0 varying from 1 to 500. In a second experiment, x_0 was set to 10 and y_0 varied from 10 to 500. For each t, we computed the maximum bias of the estimates of β_0, which is represented in Fig. 1. We observe that the parametric estimate has a lower maximum bias and a higher bdp than the nonparametric one. For the parametric estimate the bdp occurs around $\epsilon = 30\%$ for both values of x_0, which is close to $0.5 - 0.5m/n$ (see Remark 3, Section 4). For the nonparametric estimate the bdp occurs around $\epsilon = 27.5\%$

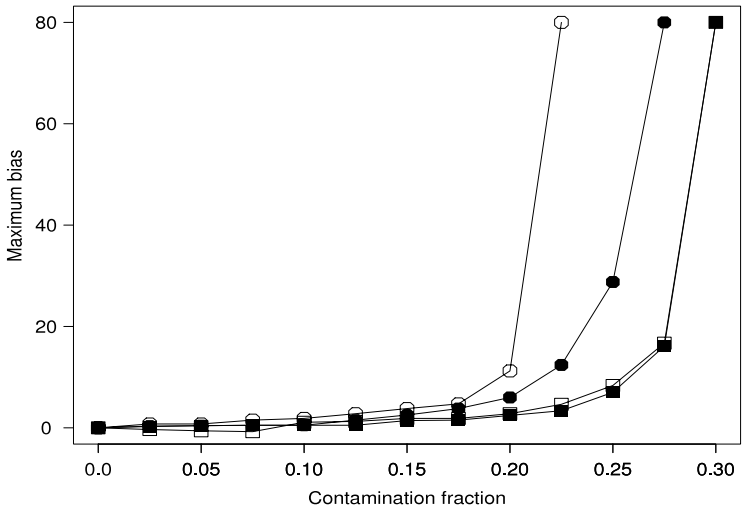

Fig. 1. Maximum bias of the parametric (squares) and non-parametric (circles) S estimates of slope as a function of the contamination fraction (filled marks: $x_0 = 1$; empty marks: $x_0 = 10$).

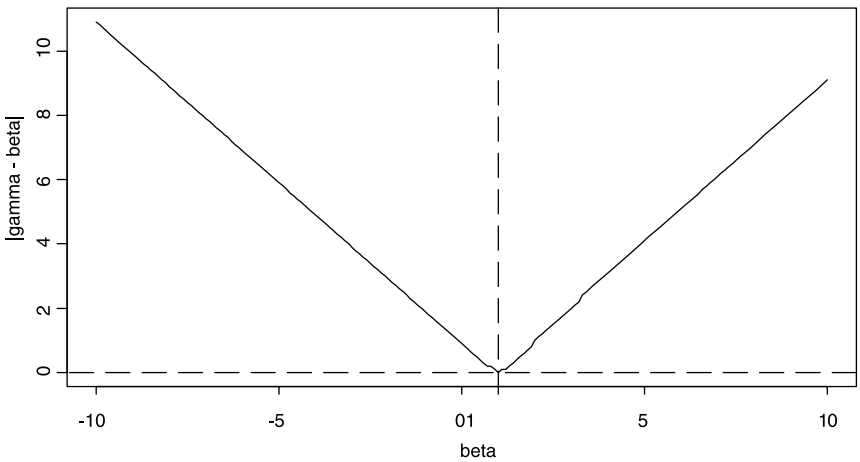

Fig. 2. Plot of $|\tilde{\gamma}_n(\beta) - \beta|$ as a function of β based on simulated data.

when $x_0 = 1$ and $\epsilon = 22.5\%$ when $x_0 = 10$. All these values are larger that the theoretical lower bound $0.5 - q/n - m/n$ given in Salibian-Barrera and Yohai (2008).

7.3. Unicity of the S estimate

In order to explore the unicity of the parametric S estimate, we considered $n = 1000$ observations from model (29) with $\alpha_0 = 0$, $\beta_0 = 1$, $\sigma_0 = 1$ and standard Gaussian errors. The censoring fraction was 0.35. Fixing the intercept at its true value ($\alpha_0 = 0$), we computed $\tilde{\gamma}_n(\beta)$ according to (15) for β varying in $(-10, 10)$. The plot of $|\tilde{\gamma}_n(\beta) - \beta|$ as a function of β in Fig. 2 shows that the only value of β satisfying (16) is the true value $\beta_0 = 1$. Using Log–Weibull errors we obtained similar results.

7.4. Sensitivity analysis

To assess the sensitivity of the estimates to an incorrect choice of the model error distribution, a Monte Carlo simulation has been performed. We considered the regression model (29) with a standard Log–Weibull error and $n = 100$. (For large n, model inadequacies can usually be detected with the help of diagnostic tools.) We computed 1000 simulated values of the estimates based on the Log–Weibull (correct model) and the Gaussian error models (incorrect model). However, since α_0 was a model dependent intercept, we redefined the intercept as the median of the response distribution for $x = 0$ (i.e., we required that the median of the error distribution was 0). Thus, with $\alpha_0 = 0$, the new intercept was $\alpha_0^* = m_0$, where $m_0 = -0.577$ was the median of a standard Log–Weibull distribution. If $(\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ was an estimate of $(\alpha_0, \beta_0, \sigma_0)$ based on

Table 3
Simulated means and root mean square errors of ML and WML estimates based on the correct (Log–Weibull) and a wrong (Gaussian) model when the errors are Log–Weibull. MM estimates are not based on a parametric model.

Parameter	Estimate	Correct model		Wrong model	
		rMSE	Mean	rMSE	Mean
α_0^*	ML	0.146	−0.578	0.176	−0.490
	WML	0.154	−0.588	0.226	−0.409
	MM	0.267	−0.367	0.267	−0.367
β_0	ML	0.144	1.014	0.182	1.050
	WML	0.158	1.034	0.169	1.026
	MM	0.178	1.018	0.178	1.018

Table 4
Estimates of the regression model for the Length of Stay dataset.

	Complete data					Outliers removed				
	ML	JIN	ZNG	MM	WML	ML	JIN	ZNG	MM	WML
α	2.93	2.42	2.49	2.21	2.46	2.43	2.22	2.41	2.27	2.39
β_1	2.34	2.11	0.37	−0.19	0.60	1.25	0.77	−0.31	0.54	0.60
$10\beta_2$	0.11	0.11	0.07	0.08	0.11	0.13	0.11	0.08	0.11	0.10
10γ	−0.30	−0.24	0.02	0.09	−0.04	−0.14	−0.05	0.10	−0.02	−0.02
σ	1.00	–	–	0.77	0.60	0.65	–	–	0.56	0.52

Log–Weibull errors, the estimate of α_0^* was $\hat{\alpha} + \hat{\sigma} m_0$. On the other hand, if $(\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ was based on the Gaussian model, the estimate of α_0^* was $\hat{\alpha}$ (since the median of the Gaussian error was 0). Table 3 reports the mean values and the root mean square errors (rMSE) of the intercept and slope estimates (scale estimates are not comparable). We note that the bias of the Gaussian slope estimates – especially the robust ones – were very small and comparable with the bias of the correct estimates. In addition, the Gaussian robust slope estimates had smaller rMSEs than the Gaussian ML. Finally, the rMSE of the Gaussian WML estimate of α_0^* was lower than the rMSE of the nonparametric MM estimate of Salibian-Barrera and Yohai (2008).

7.5. Computing times

Computing times of the resampling algorithm for the parametric S estimate are reported in LMY.

8. Illustrations with real data

In a first example, we consider a sample of 75 hospital stays for "Major cardiovascular interventions". The data (made available in LMY) are shown in Fig. 3. 45 stays were censored because the patients were transferred to a different hospital before dismissal. The LOS of two young patients were exceptionally high. We study the relationship between length of stay (LOS) and two covariates usually available on administrative files: age of the patient (x_1) and admission type ($x_2 = 0$ for planned admissions, $x_2 = 1$ for emergency admissions.). This kind of relationship is used as a basis to determine reimbursement rates. We consider the model $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \gamma x_1 x_2 + \sigma u$, where $y = \log(\text{LOS})$. We computed the ML estimate, the adaptive WML estimate (with normal errors), the MM estimate of Salibian-Barrera and Yohai (2008), the rank based estimate of Jin et al. (2003), JIN, and the nonparametric estimate of Zeng and Lin (2007), ZNG. Note that JIN and ZNG do not directly provide the intercept and we used the median of the KM distribution of the residuals $y_i - \beta_1 x_1 - \beta_2 x_2 - \gamma x_1 x_2$ to estimate α. The results are given in Table 4 (complete data); the prediction lines for the regular cases are shown in Fig. 3, panel (a) and for the emergency cases in panel (b). Standard errors based on Theorem 3 are reported in LMY.

We remark the two strong negative interactions γ given by ML and JIN, which are meaningless and disagree with the very small values of γ provided by WML, MM, and ZNG. Clearly, the ML and JIN prediction lines for emergency cases suffer the leverage effect of the two outlying observations with very long stay. Surprisingly, ZNG does not seem to be affected. The next example shows however, that it might also be impaired by leverage outliers.

The WML estimate points out five observations with zero weights. These observations are indicated by prominent marks in panels (a) and (b) of Fig. 3. If we remove these outliers we obtain the second set of estimates given in Table 4 and the prediction lines in Fig. 3, panels (c) and (d). These estimates look much more alike than those based on the full data set and all values of γ are small. Note however, the large changes in β_1 for ML, JIN, and ZNG and that WML remains virtually unchanged after removal of the five outliers.

The plots in Fig. 4 show the parametric (normal) estimate, the Kaplan–Meier (KM) estimate, and the semiparametric estimate of the standardized residual cdf for ML, WML, and MM based on (5). Censored and noncensored residuals are marked by "0", respectively "1" on the horizontal axes. We note the very large steps of the KM distribution corresponding to the two extreme noncensored outliers. The reason is that KM puts the mass of 12 uncensored residuals in the interval (0.78, 2.64) on these two points. Panel (a) shows that the ML fit and the associated residual scale estimate are strongly affected

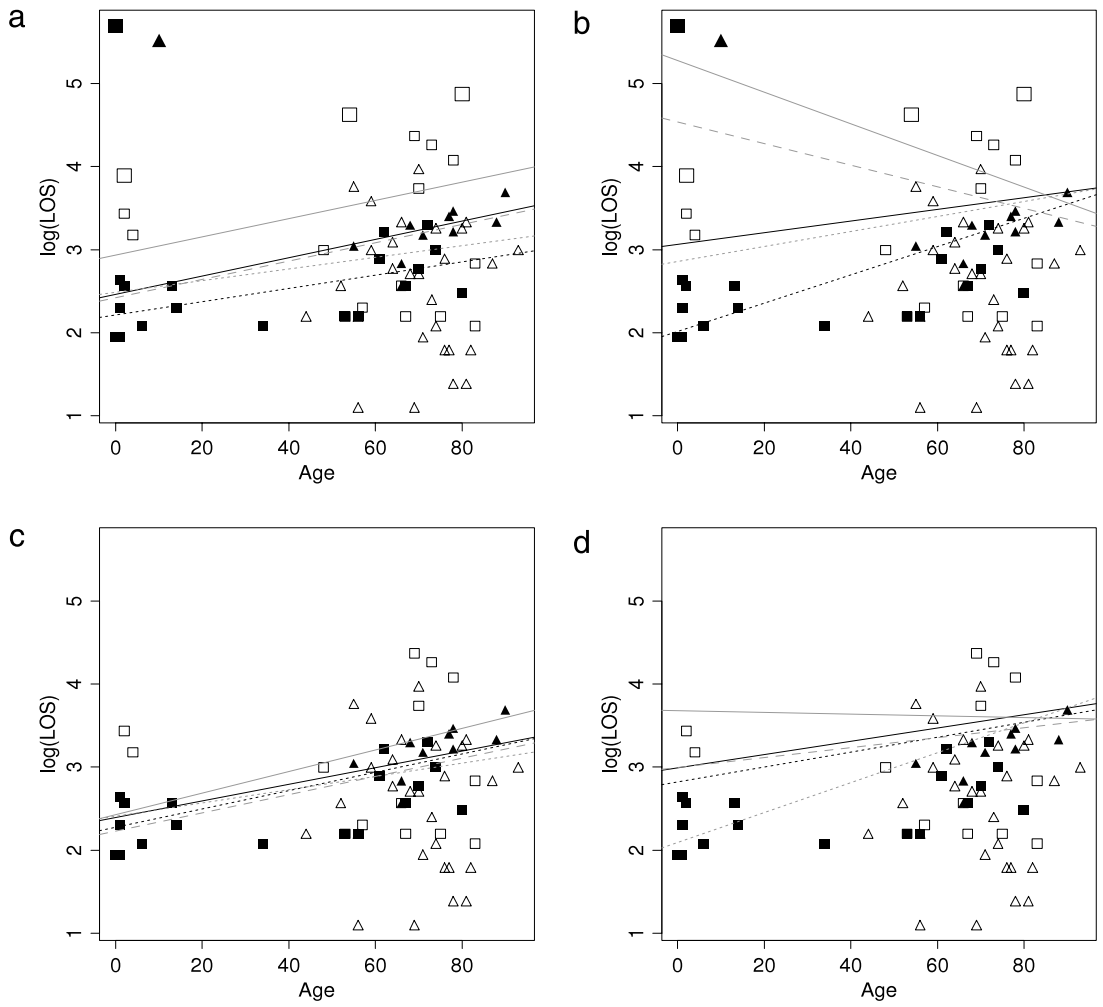

Fig. 3. Data: log(LOS) and Age of 75 hospital patients. Squares are regular admissions, triangles emergency admissions. Filled and empty marks indicate complete and censored cases. Panels (a) and (b) show the complete data. Large marks are outliers. In panels (c) and (d), outliers have been removed. The fits WML (black solid line), MM (black dotted line), ML (grey solid line), ZNG (grey dotted line) JIN (grey broken line) in panels (a) and (c) refers to regular cases. The fits in panels (b) and (d) refers to emergency cases.

by these two points. From panel (c), we learn that the MM estimate based on KM is also badly damaged. This explains the large negative value of β_1. As a matter of fact, if we remove the 12 observations with the largest censored residuals, the two extreme steps contract and we obtain the plot in Fig. 4, panel (d), i.e. a very nice fit. Finally, the WML estimate (panel (b)) behaves as desired, its residuals being – with two exceptions – almost perfectly normal.

In a second example, we consider the Heart dataset analyzed in Kalbfleisch and Prentice (2002) and available in the "survival" library of R. These data contain information on 69 heart transplant recipients, including their age and their time to death or censoring (survival). The model $y = \alpha + \beta x + \sigma u$, where $y = \log(\text{time})$ and $x = \text{age}$, has been considered in Salibian-Barrera and Yohai (2008). The results are given in Table 5 and shown in Fig. 5. Four cases – the large marks in Fig. 5 – receive a zero weight in the WML procedure. Table 4 also provides the results obtained after removal of these outliers.

We observe that, with the full data set, the ML, JIN, and ZNG estimates of β are similar and indicate that the effect of age on the mean survival time is very small. This is somewhat counterintuitive. The robust estimates yield a slope with a much smaller value, indicating that there is in fact a negative linear relationship, as it is expected from the data (older transplant patients tend to die sooner). Clearly, the outliers have an important leverage effect on ML, JIN, and ZNG. Removing the outliers, these estimates become similar to WML and MM which remain almost unchanged.

9. Discussion

We have introduced a robust procedure to estimate a linear regression model with censored observations, which may be considered a parametric counterpart of the estimates presented in Salibian-Barrera and Yohai (2008).

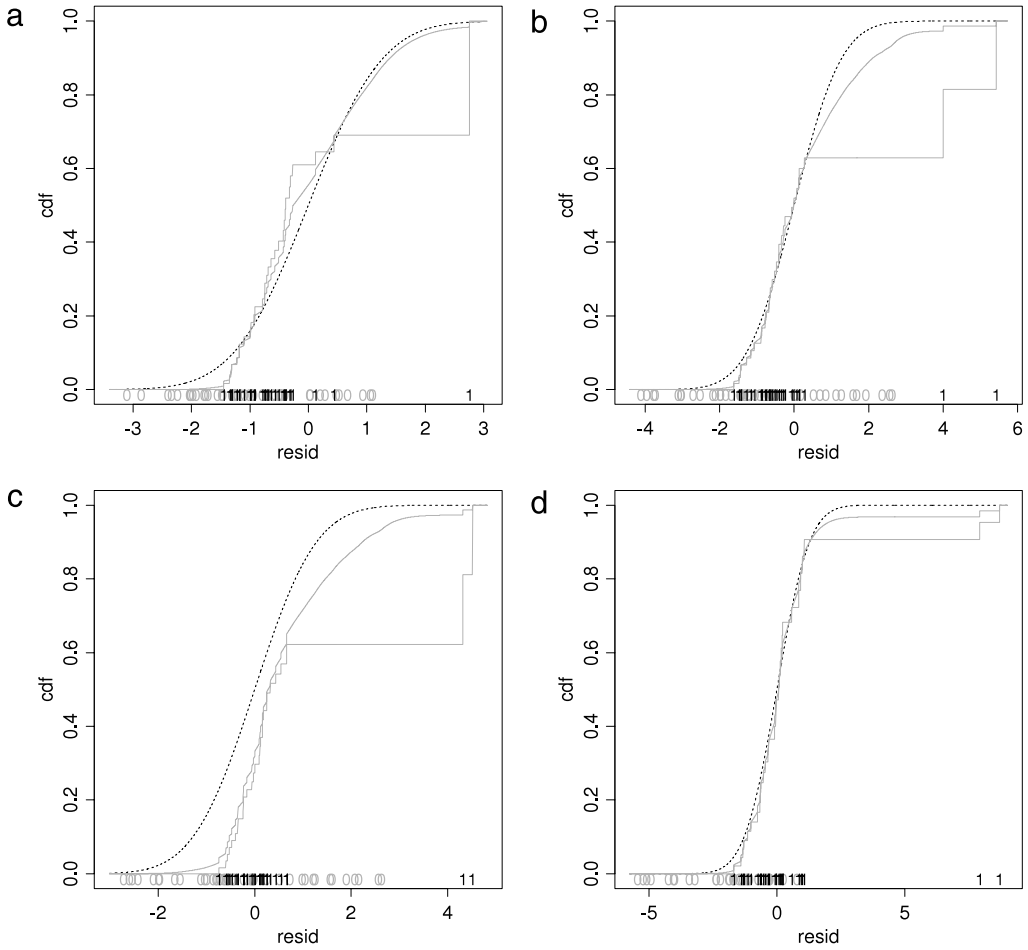

Fig. 4. Parametric (normal, dotted line), Kaplan–Meier (grey step function), and semiparametric (grey smooth function) estimates of the standardized error cdf for ML (panel (a)), WML (panel (b)), and MM (panel (c)). Censored and noncensored residuals are marked by "0", respectively "1" on the horizontal axes. Panel (d) is obtained with MM, after removal of the 10 observations with the largest censored residuals.

Table 5
Estimates of the regression model for the Heart dataset.

	Complete data					Outliers removed				
	ML	JIN	ZNG	MM	WML	ML	JIN	ZNG	MM	WML
α	8.90	8.79	7.10	13.17	13.63	12.91	13.66	12.16	14.48	13.67
β	−0.07	−0.07	−0.04	−0.16	−0.17	−0.15	−0.17	−0.14	−0.18	−0.17

Our Monte Carlo simulations show that, when the model is correct, the new estimates are more efficient and more robust than the previous ones. The higher degree of robustness was explained in the introduction by the fact that the Kaplan–Meier estimate creates artificial outliers. Another advantage of the parametric approach occurs when the error distribution is asymmetric. In fact, in this case, if the family of parametric models includes asymmetric distributions, a better and differential treatment of left and right outliers is made possible.

As usual, the plausibility of an initially selected model has to be checked with the help of diagnostic tools or goodness of fit measures. For example, in Fig. 4, we compared the Kaplan–Meier distribution of the standardized residuals with the assumed normal distribution. When inadequacy of the selected model is detected, the procedure has to be repeated using different distribution families until a satisfactory fit is found. Methods to choose among several competing distribution models in the presence of censoring and in the absence of covariates have been described in the literature; see for example Kim and Yum (2008) and the references mentioned by these authors. Some of these methods (e.g., the comparison of the likelihoods of competing models with the same number of parameters) could be naturally extended to regression. One could also robustify these procedures by deleting the detected outliers before their application. However, we consider that this is a matter of further research.

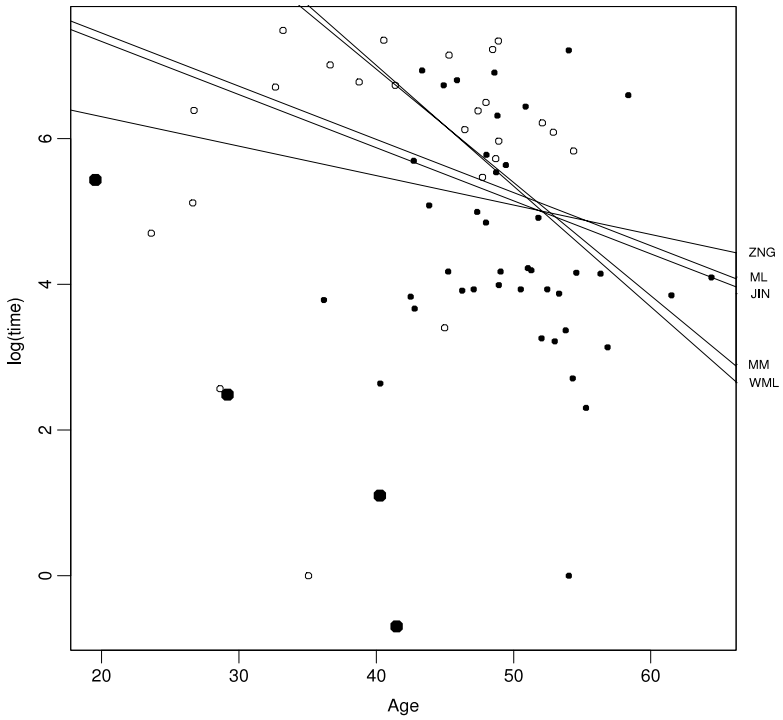

Fig. 5. Data: log(time) and Age of 69 heart transplant recepients. Filled and empty marks correspond to complete and censored observations. Large marks correspond to outliers.

One may also wonder how good is the performance of the parametric estimate if the selected family of model distributions is not correct. To answer this question we studied the performance of our procedure based on a Gaussian error model when the true distribution was Log–Weibull. The results can be considered as satisfactory.

Acknowledgements

This work was supported by Grant 205320-108424 and 205320-116357 from the Swiss National Science Foundation, Grant X-018 from the Universidad de Buenos Aires, Grant PIP 5505 from Conicet, Argentina, and Grant PICT 21407 from ANPCYT, Argentina. The authors thank A. Randriamiharisoa for the programming help.

Appendix. Supplementary data

Supplementary material related to this article can be found online at doi:10.1016/j.csda.2010.07.017.

References

Buckley, J., James, I., 1979. Linear regression with censored data. Biometrika 66, 429–436.
Cots, F., Elvira, D., Castells, X., Saez, M., 2003. Relevance of outlier cases in case mix systems and evaluation of trimming methods. Health Care Management Science 6, 27–35.
Cox, D.R., 1972. Regression models and life tables (with discussion). Journal of the Royal Statistical Society. Series B 34, 187–220.
Cox, D.R., Oakes, D., 1984. Analysis of Survival Sata. Chapman and Hall, London.
Gervini, D., Yohai, V.J., 2002. A class of robust and fully efficient estimates. The Annals of Statistics 30, 1–34.
Huber, P.J., 1981. Robust Statistics. Wiley, New York.
Jin, Z., Lin, D.Y., Wei, L.J., Jing, Z., 2003. Rank-based inference for the accelerated failure time model. Biometrika 90, 341–353.
Kalbfleisch, J.D., Prentice, R.L., 2002. The Statistical Analysis of Failure Time Data, 2nd ed. Wiley, New York.
Kim, J.S., Yum, B.-J., 2008. Selection between Weibull and lognormal distributions: a comparative simulation study. Computational Statistics & Data Analysis 53 (2), 477–485.
Lawless, J.F., 2003. Statistical Models and Methods for Lifetime Data, 3rd ed. Wiley, New York.
Locatelli, I., Marazzi, A., Yohai, V.J., 2010. Supplemental material for "Robust accelerated failure time regression". Available as a supplemental material together with the electronic version of the paper.
Marazzi, A., Villar, A.J., Yohai, V.J., 2009. Supplemental material for Robust response transformations based on optimal prediction. Available at the Jasa Site for supplemental material.
Marazzi, A., Yohai, V.J., 2004. Adaptively truncated maximum likelihood regression with asymmetric errors. Journal of Statistical Planning and Inference 122, 271–291.
Maronna, R.A., Martin, R.D., Yohai, V.J., 2007. Robust Statistics: Theory and Methods. Wiley, New York.
Mizera, I., 1993. On consistent *M*-estimators: tuning constants, unimodality and breakdown. Kybernetika 30, 289–300.

Reid, N., 1994. A conversation with Sir David Cox. Statistical Science 9, 439–455.
Rousseeuw, P.J., Yohai, V.J., 1987. Robust regression by means of *S*-estimates. In: Franke, J., Härdle, W., Martin, R. (Eds.), Robust and Nonlinear Time Series. In: Lecture Notes in Statistics, vol. 26. Springer, New York, pp. 256–272.
Salibian-Barrera, M., Yohai, V.J., 2006. A fast algorithm for *S*-regression estimates. Journal of Computational and Graphical Statistics 15, 1–14.
Salibian-Barrera, M., Yohai, V.J., 2008. High breakdown point robust regression with censored data. The Annals of Statistics 36, 118–146.
Zeng, D., Lin, D., 2007. Efficient estimation for the accelerated failure time model. Journal of the American Statistical Association 102, 1387–1396.