THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF STATISTICS

TECHNICAL REPORT # 225

HIGH BREAKDOWN POINT ROBUST REGRESSION WITH CENSORED DATA

MATIAS SALIBIAN-BARRERA AND VICTOR J. YOHAI

NOVEMBER 2006

High breakdown point robust regression with censored data

Matías Salibian-Barrera, University of British Columbia
and Víctor J.Yohai, Universidad de Buenos Aires

November 17, 2006

Abstract

In this paper we propose a class of high breakdown point estimators for the linear regression model when the response variable contains censored observations. These estimators are robust against high-leverage outliers and they generalize the LMS, S, MM and τ-estimators for linear regression. An important contribution of this paper is that we can define consistent estimators using a bounded loss function (or equivalently, a re-descending score function). Since the calculation of these estimators can be computationally costly we propose an efficient algorithm to compute them. We illustrate their use on an example and present simulation studies that show that these estimators also have good finite sample properties.

1 Introduction

Consider the linear regression model

$$y_i = \boldsymbol{\beta}_0' \mathbf{x}_i + u_i \,, \quad i = 1, \ldots, n \,, \tag{1.1}$$

where u_i are i.i.d. errors, and the covariates $\mathbf{x}_i \in \mathbb{R}^p$ are independent from the errors. When there is an intercept the first component of $\mathbf{x}_i$ is set to 1. In this paper we study the problem of robust estimation of $\boldsymbol{\beta}_0$ when the response variable is censored. Miller (1976) studied least squares estimators (LS) for censored responses. He proposed to modify the classical LS estimator

$$\hat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - \boldsymbol{\beta}' \mathbf{x}_i)^2 = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} E_{F_{n\boldsymbol{\beta}}}[u^2] \,, \tag{1.2}$$

replacing the empirical distribution of the residuals $F_{n\boldsymbol{\beta}}$ with the corresponding Kaplan-Meier (KM) estimator $F^*_{n\boldsymbol{\beta}}$ [Kaplan and Meier, (1958)]. Unfortunately, the resulting estimator is not consistent in general and the iterative algorithm to compute it may have several or no solutions.

Buckley and James (1979) studied a different extension of LS to censored response variables by modifying the LS scores equations

$$\sum_{i=1}^{n}(y_i - \hat{\boldsymbol{\beta}}_n'\mathbf{x}_i)\,\mathbf{x}_i = \mathbf{0}\,, \tag{1.3}$$

using a conditional distribution approach. This proposal replaces censored residuals by their estimated conditional expectation given that the response is larger than the recorded (censored) value. The conditional expectation is estimated using $F^*_{n\hat{\boldsymbol{\beta}}_n}$. James and Smith (1984) and Lai and Ying (1991) showed that this estimator is consistent. However, it has the same problems in terms of existence and unicity of the solution for finite samples as Miller's proposal.

Stute (1993, 1996) and Sellero *et al.* (2004) also considered linear regression estimators for censored responses. Their main assumption is that $P(Y \le C \,|\, X, Y) = P(Y \le C \,|\, Y)$ where C is the censoring variable. This holds, for example, when the censoring variables are independent from both the responses and the covariates. In our view, this is a rather strong condition. Another shortcoming of this approach is that the resulting estimators are not regression equivariant.

It is well known that LS estimators are very sensitive to the presence of outliers. The estimators proposed by Miller (1976) and Buckley and James (1979) inherit the lack of robustness of the LS estimators for uncensored observations. Robust regression estimators for uncensored observations have been actively studied for the last 30 years. Huber's original proposal [Huber (1973)] was to replace the square loss function in (1.2) with an even, convex and monotone on $[0, \infty)$ function $\rho(t)$ with bounded derivative. However, such a function $\rho(t)$ will be unbounded and the resulting estimator has breakdown point equal to zero because it cannot handle high-leverage outliers (see Lemma 5.2 in Yohai and Zamar, 1993). Generalized M-estimators (GM) [Krasker, (1980); Krasker and Welsch, (1982)] tried to overcome this difficulty by downweighting observations with high leverage. Unfortunately, the breakdown point of these estimators is at most $1/(p+1)$ where p is the number of covariates [Maronna *et al.* , (1979)]. The first equivariant regression estimators with high-breakdown point independent of the dimension of the problem were the Least Median of Squares (LMS) estimators [Hampel (1975), Rousseeuw (1984)] and the S-estimators [Rousseeuw and Yohai, (1984)]. Note, however, that the LMS and 50% breakdown point S estimators have very low efficiency for uncontaminated observations. Yohai introduced the class of MM-estimators

[Yohai, (1987)] which have simultaneous high breakdown point and high efficiency for normal errors. Another class of estimators with good robustness and efficiency properties are the τ-estimators [Yohai and Zamar, (1988)].

In recent years there has been some interest in extending robust regression estimators to the case of censored response variables. Ritov (1990) studied a generalization of Bukley and James's proposal for robust estimators. He considered monotone non-decreasing score functions ψ (that correspond to unbounded loss functions ρ) and showed that under certain regularity conditions there exists a sequence of $\sqrt{n}$-consistent solutions to the estimating equations. This sequence is also asymptotically normal. Unfortunately, since these estimators are based on an unbounded loss function ρ they are not robust against high-leverage outliers. More recently, Lai and Ying (1994) extended the conditional expectation approach of Bukley and James to M-regression estimators for censored and truncated data. Their proposal also requires a monotone score function.

If we allow for a re-descending score function ψ (equivalently, a bounded loss function ρ) then the estimating equations may have several solutions with different robustness properties. Moreover, if we define a robust estimator as the solution to a minimization problem similar to (1.2) but replacing the squared residuals with $\rho(u)$ for a bounded loss function ρ, then this estimator may not be consistent [Lai and Ying, (1991, 1994)]. Hence, unlike in the uncensored regression model, we do not have a way to identify which solutions of the re-descending score equations are not affected by the outliers.

In this paper we extend the approach of Bukley and James and Ritov to M-estimators with bounded loss functions ρ. We achieve this by proposing an estimator that is the solution to a minimization problem that has a consistent and robust solution. In particular we obtain extensions of the LMS, S, MM and τ estimators. We show that these estimates are Fisher and $\sqrt{n}$-consistent, asymptotically normal, and that they have high breakdown point.

It is important to realize that when there are censored observations the breakdown point of an estimator maybe much lower than in the uncensored case. For example, in the location model the worst contamination occurs when all the censored observations are between the outliers and the "good" non-censored points. Suppose that we have a fraction ϵ of outliers going to $+\infty$ and a proportion λ of censored observations. Since the KM estimator distributes the mass of the censored observations among the non-censored points to their right (Efron, 1967), in this case the mass given to the outliers by the KM estimators will be $\gamma = \lambda + \epsilon$. Consequently, the sample median will not break if $\gamma < 1/2$, or equivalently, if $\epsilon < 1/2 - \lambda = \eta$. It follows that the breakdown point of the median is equal to η, which is less than 1/2 when there are censored observations.

The rest of this paper is organized as follows. Section 2 contains our main definitions. The robustness properties of our proposal are discussed in Section 3 and their asymptotic properties in Section 4. In Section 5 we present an algorithm to compute these estimators. An example with real-life data is given in Section 6 and the results of a Monte Carlo experiment are discussed in Section 7. Finally, all the proofs are given in the Appendix.

2 Robust estimators

Consider the linear regression model (1.1). We assume that the sample may be right-censored, i.e., there are unobservable random variables $c_1, \ldots, c_n$ independent from the errors u_i's such that we observe $y_i^* = \min(y_i, c_i)$ for $i = 1, \ldots, n$. In other words, the observed data is $\mathbf{z}_i = (y_i^*, \mathbf{x}_i', \delta_i)'$, $i = 1, \ldots, n$, where $\delta_i = I\{y_i \leq c_i\}$, and $I\{A\}$ is the indicator function of the event A.

When the scale of the residuals is known, regression M-estimators for uncensored observations are defined by

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n} \rho(r_i(\boldsymbol{\beta})) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} E_{F_{n\boldsymbol{\beta}}}[\rho(u)]\,, \tag{2.1}$$

where $F_{n\boldsymbol{\beta}}$ is the empirical distribution of the residuals $r_i(\boldsymbol{\beta}) = y_i - \boldsymbol{\beta}'\mathbf{x}_i$, and $\rho : \mathbb{R} \to \mathbb{R}^+$ is a function satisfying

P1: $\rho(0) = 0$ and ρ is continuous at 0;

P2: $\rho(-u) = \rho(u)$ for $u > 0$;

P3: ρ is monotone non-decreasing on $u > 0$; and

P4: $\sup_u \rho(u) = a < +\infty$.

(see Huber, 1973). If $\psi(u) = \partial\rho(u)/\partial u$ then the estimator $\hat{\boldsymbol{\beta}}_n$ also satisfies the following vector equation:

$$\frac{1}{n}\sum_{i=1}^{n} \psi(r_i(\boldsymbol{\beta}))\mathbf{x}_i = E_{H_{n\boldsymbol{\beta}}}[\psi(u)\mathbf{x}] \;= \mathbf{0}\,, \tag{2.2}$$

where $H_{n\beta}$ is the empirical distribution of the vectors $(r_i(\boldsymbol{\beta}), \mathbf{x}_i')' \in \mathbb{R}^{p+1}$, $i = 1, \ldots, n$.

Since not all the residuals $r_i(\boldsymbol{\beta})$ are observed in the presence of censoring, we can define the censored residuals by $r_i^*(\boldsymbol{\beta}) = y_i^* - \boldsymbol{\beta}'\mathbf{x}_i$. Note that $r_i^*(\boldsymbol{\beta}) = \min(r_i(\boldsymbol{\beta}), c_i - \boldsymbol{\beta}'\mathbf{x}_i)$, and therefore we can think of the $r_i^*(\boldsymbol{\beta})$ as censored observations of $r_i(\boldsymbol{\beta})$ with censoring variables $c_i - \boldsymbol{\beta}'\mathbf{x}_i$, $i = 1, \ldots, n$. Then, in the case of a censored response variable, one way to generalize (2.1) is to replace it by

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n} E[\rho(r_i(\boldsymbol{\beta}))|\mathbf{z}_i] = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n} E_{F_{\boldsymbol{\beta}}}[\rho(u)|\mathbf{w}_i(\boldsymbol{\beta})]\,, \tag{2.3}$$

where $F_{\boldsymbol{\beta}}$ is the distribution of the residuals $r(\boldsymbol{\beta})$, $\mathbf{w}_i(\boldsymbol{\beta}) = (r_i^*(\boldsymbol{\beta}), \delta_i)$ and

$$E_{F_{\boldsymbol{\beta}}}(\rho(u)|\mathbf{w}_i(\boldsymbol{\beta})) = \begin{cases} \rho(r_i^*(\beta)) & \text{if } \delta_i = 1\,, \\ \\ \int_{r_i^*(\beta)}^{\infty} \rho(u)\ dF_{\boldsymbol{\beta}}(u) \Big/ [1 - F_{\boldsymbol{\beta}}(r_i^*(\boldsymbol{\beta}))] & \text{if } \delta_i = 0\,. \end{cases}$$

Intuitively, to obtain (2.3) from (2.1), for each censored observation we replace the term $\rho(r_i(\boldsymbol{\beta}))$ in (2.1) by the conditional expectation of $\rho(u)$ given that the (actual but unobserved) residual is larger than or equal to the observed censored residual $r_i^*(\boldsymbol{\beta})$.

The score equations in (2.2) can also be similarly modified to obtain

$$\frac{1}{n}\sum_{i=1}^{n} E_{F_{\boldsymbol{\beta}}}[\psi(u)|\mathbf{w}_i(\boldsymbol{\beta})]\,\mathbf{x}_i = \mathbf{0}\,. \tag{2.4}$$

Since the distribution of the residuals $F_{\boldsymbol{\beta}}$ in (2.3) and (2.4) is unknown, we can estimate it with the Kaplan-Meier estimator $F_{n\boldsymbol{\beta}}^*$ based on $r_i^*(\boldsymbol{\beta})$

To guarantee consistency of the estimator defined by

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n} E_{F_{n\boldsymbol{\beta}}^*}[\rho(u)|\mathbf{w}_i(\boldsymbol{\beta})] \tag{2.5}$$

we need that $F_{n\boldsymbol{\beta}}^*$ be consistent to $F_{\boldsymbol{\beta}}$ for all $\boldsymbol{\beta}$. Let F and D be the distribution functions of the errors u_i and censoring variables c_i, $i = 1, \ldots, n$, respectively. Let $\tau_F = \inf\{u : F(u) = 1\}$ and let τ_D be defined similarly. In what follows we will assume that

R1: $\tau_F < \tau_D$, or $\tau_F = \tau_D = \infty$, or $\tau_F = \tau_D$ and τ_F is a continuity point of F;

R2: F and D do not have jumps in common.

Under these conditions, a sufficient condition for the KM estimator to be consistent is the independence between the uncensored variables and the censoring times [see, for example, Breslow and Crowley (1974)]. When $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ we have $r_i(\boldsymbol{\beta}_0) = u_i$ which are independent from the corresponding censoring times $c_i - \boldsymbol{\beta}_0'\mathbf{x}_i$ because we have assumed that the errors are independent from the c_i's and the $\mathbf{x}_i$'s. However, for $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ it is not generally true that $r_i(\boldsymbol{\beta})$ is independent from $c_i - \boldsymbol{\beta}'\mathbf{x}_i$, $i = 1, \ldots, n$. Hence, we can only guarantee the consistency of $F^*_{n\boldsymbol{\beta}}$ to $F_{\boldsymbol{\beta}}$ when $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Therefore, the estimator defined in (2.5) may not be consistent (Lai and Ying, 1991, 1994).

On the other hand, note that the estimator $\hat{\boldsymbol{\beta}}_n$ defined as the solution to

$$\frac{1}{n}\sum_{i=1}^{n} E_{F^*_{n\boldsymbol{\beta}}}[\psi(u)|\mathbf{w}_i(\boldsymbol{\beta})]\,\mathbf{x}_i = \mathbf{0}\,, \tag{2.6}$$

is Fisher consistent. In fact, $F^*_{n\boldsymbol{\beta}_0} \to F_{\boldsymbol{\beta}_0}$ and therefore

$$\frac{1}{n}\sum_{i=1}^{n} E_{F^*_{n\boldsymbol{\beta}_0}}[\psi(u)|\mathbf{w}_i(\boldsymbol{\beta}_0)]\,\mathbf{x}_i \to E_{H_0}(\psi(u)\mathbf{x}) = \mathbf{0}\,,$$

where H_0 is the joint distribution of $(u, \mathbf{x}')'$.

It is important to note that, unlike in the uncensored regression case, equations (2.5) and (2.6) are not equivalent: we cannot obtain (2.6) by differentiating (2.5) because $F^*_{n\boldsymbol{\beta}}$ depends on $\boldsymbol{\beta}$.

M-estimators defined by (2.6) were first proposed by Ritov (1990) and further studied by Lai and Ying (1994) when $\psi(u)$ is monotone (which corresponds to a convex ρ). However, it is well know that M-estimators with monotone ψ functions are only robust against low leverage outliers. As mentioned in the Introduction, the main difficulty in using a re-descending ψ in (2.6) is that in general this equation may have several solutions with different robustness properties. Althought in the uncensored regression model this difficulty can be avoided by defining the estimator as the solution to the minimization problem (2.1), the corresponding minimization in the censored case (2.5) does not in general yield a consistent estimator. In other words, (2.5) cannot be used to select a consistent solution of (2.6).

For this reason, in the next sub-section we will define robust M-estimators as the solution of a minimization problem using a bounded loss function ρ that has a consistent sequence of solutions.

2.1 Consistent M-estimators

First note that to obtain scale equivariant regression estimators we need to standardize the residuals in the estimating equations using a robust error scale estimator s_n.

Let $\rho : \mathbb{R} \to \mathbb{R}^+$ satisfy regularity conditions P1-P4 above. For each $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in $\mathbb{R}^p$ define

$$C_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=1}^{n} E_{F^*_{n\boldsymbol{\beta}}}\left[\rho\left(\frac{u - \boldsymbol{\gamma}'\mathbf{x}_i}{s_n}\right)\middle|\, \mathbf{w}_i(\boldsymbol{\beta})\right]\,, \tag{2.7}$$

where s_n is a robust scale estimator of the residuals. For each $\boldsymbol{\beta} \in \mathbb{R}^p$ let

$$\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^p} C_n(\boldsymbol{\beta}, \boldsymbol{\gamma})\,. \tag{2.8}$$

Note that $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})$ can be considered an M-estimator of regression of the residuals $r_i(\boldsymbol{\beta})$ on the covariates $\mathbf{x}_i$. Since $F^*_{n,\boldsymbol{\beta}_0}$ is a consistent estimator of $F_{\boldsymbol{\beta}_0}$, the distribution of the u_i's, and since the errors are independent of the $\mathbf{x}_i$'s, it is reasonable to expect that

$$\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_0) \to \mathbf{0}\,.$$

This can be formally proved with similar arguments to those used in the proof of Theorem 5 below. Therefore, we define an estimator of $\boldsymbol{\beta}_0$ by the equation

$$\hat{\boldsymbol{\gamma}}_n(\hat{\boldsymbol{\beta}}_n) = \mathbf{0}\,, \tag{2.9}$$

To avoid existence problems, we can alternatively define $\hat{\boldsymbol{\beta}}_n$ as

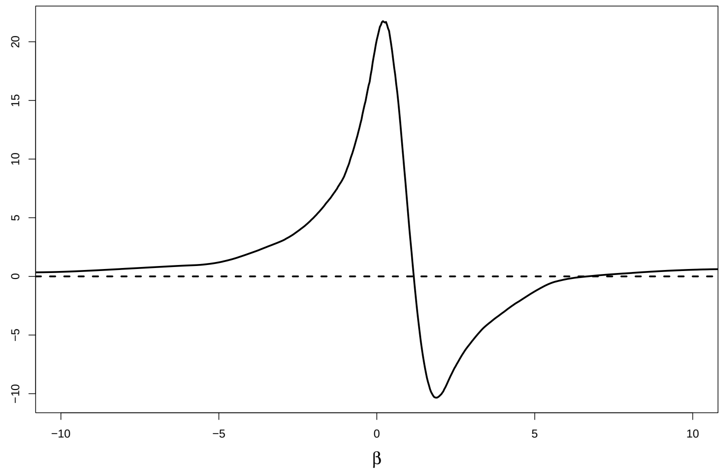
$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left[\, \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})'\, \mathbf{A}_n\, \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) \,\right]\,, \tag{2.10}$$

where $\mathbf{A}_n = \mathbf{A}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is any robust equivariant estimator of the covariance matrix of the explanatory variables $\mathbf{x}_i$, $1 \le i \le n$. The covariance matrix $\mathbf{A}_n$ is needed to maintain the affine equivariance of the estimator.

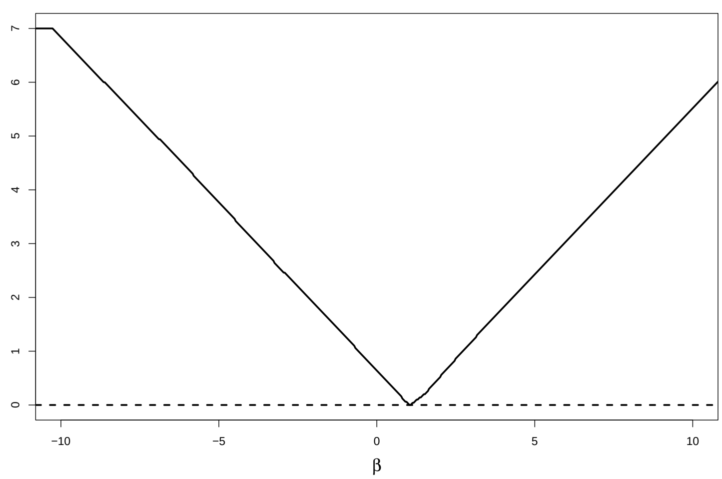
As an illustration of the difference between using (2.5) and (2.9) to define a robust estimator, in Figure 1 we plot $\|\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})\|$ and the score equations (2.5) as a function of $\boldsymbol{\beta}$ for a data set of $n = 200$ observations with $\boldsymbol{\beta}_0 = 1.5$ and a probability of censoring of approximately 32%. These data were generated following the same model we used in our simulation study described in Section 7. Note that although the score equation has two distinct solutions and only one is close to the true value of $\boldsymbol{\beta}_0 = 1.5$, our proposed optimization problem has a unique minimum and this minimum is close to $\boldsymbol{\beta}_0$. This definition may be considered an extension of Ritov's M-estimators for censored data to the case of bounded ρ functions. In particular, note that $\hat{\boldsymbol{\beta}}_n$ satisfies equation (2.6) with $\psi(u) = \rho'(u)$. It follows that this estimator will have the same asymptotic properties as the estimators considered in Ritov (1990).

2.2 S-estimators

The scale estimator s_n in (2.7) may be chosen to be the scale of the residuals of an initial (and scale-equivariant) estimator that does not require a scale estimator itself. One class of estimates



(a) Right hand side of score equation (2.5)



(b) $\|\hat{\gamma}_n(\boldsymbol{\beta})\|$

Figure 1: Panel (a) shows an example where the score equations (2.5) have two roots with only one of them close to $\boldsymbol{\beta}_0 = 1.5$ whereas Panel (b) shows that, for the same data set, the objective function of (2.9) has a unique minimum close to $\boldsymbol{\beta}_0$.

that satisfies this is the class of S-estimators [Rousseeuw and Yohai, (1984)]. We can extend this class of estimators to the case of censored observations following the same principle as above, i.e., for each $\boldsymbol{\beta}$ we fit an S-estimate to the residuals $r_i^*(\boldsymbol{\beta})$, and find the $\boldsymbol{\beta}$ whose residuals have the "smallest" S-estimator (i.e. the one with the smallest norm).

Let ρ_1 satisfy regularity conditions P1-P4 and let $b = E_F[\rho_1(u)]$ where F is the distribution of the errors u_i in (1.1). Define the M-scale $S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ by

$$\frac{1}{n} \sum_{i=1}^{n} E_{F^*_{n\boldsymbol{\beta}}} \left[\rho_1 \left(\frac{u - \boldsymbol{\gamma}' \mathbf{x}_i}{S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})} \right) \middle| \, \mathbf{w}_i(\boldsymbol{\beta}) \right] = b \,, \tag{2.11}$$

and let

$$\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) = \arg \min_{\boldsymbol{\gamma} \in \mathbb{R}^p} S_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) \,. \tag{2.12}$$

Note that $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})$ is the S-estimator of regression of the residuals $(r_i^*(\boldsymbol{\beta}), \mathbf{x}_i')'$, $i = 1, \ldots, n$. We define the S-regression estimator for censored responses as the vector $\tilde{\boldsymbol{\beta}}_n$ such that

$$\hat{\boldsymbol{\gamma}}_n(\tilde{\boldsymbol{\beta}}_n) = \mathbf{0} \,. \tag{2.13}$$

As before, to avoid existence problems, the following definition is also natural:

$$\tilde{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left[\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})' \, \mathbf{A}_n \, \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) \right] \,,$$

where $\mathbf{A}_n = \mathbf{A}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is any robust equivariant estimator of the covariance matrix of the covariates $\mathbf{x}_i$.

A robust residual scale estimate s_n can be defined by

$$s_n = S_n(\tilde{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\gamma}}_n(\tilde{\boldsymbol{\beta}}_n)) \,. \tag{2.14}$$

In particular, we can obtain a consistent version of the LMS using as ρ_1 a jump function

$$\rho_1(u) = \begin{cases} 0 & \text{if } |u| < 1 \\ 1 & \text{if } |u| \geq 1 \,, \end{cases} \tag{2.15}$$

and $b = 1/2$ in equation (2.11) above.

In Section 3 we will show that the choice $b = \sup_u \rho_1(u)/2$ yields regression estimators with high breakdown point. However, we know from the uncensored case that S-estimators cannot combine high breakdown point with high efficiency for normal errors (see Hössjer, 1992). To overcome this problem, in the next sub-section we will extend to the censored case a class of estimators that can achieve simultaneous high efficiency and high breakdown point.

2.3 MM-estimators

Yohai (1987) proposed a class of estimators, called MM-estimators, that simultaneously have breakdown point 50% and high efficiency for normal errors. In this sub-section, we extend this class of estimators to the case of censored responses.

Consider two functions ρ_1 and ρ_2 that satisfy the regularity conditions P1-P4. Moreover, assume that $\rho_2(u) \le \rho_1(u)$ for all u and that $\sup_u \rho_2(u) = \sup_u \rho_1(u)$. Let $\tilde{\boldsymbol{\beta}}_n$ and s_n be the S-regression and S-scale estimators calculated as in (2.13) and (2.14) respectively. For each $\boldsymbol{\gamma} \in \mathbb{R}^p$ define $R(\boldsymbol{\gamma})$ as

$$R(\boldsymbol{\gamma}) = \sum_{i=1}^{n} E_{F^*_{n\tilde{\beta}_n}} \left[\rho_2 \left(\frac{u - \boldsymbol{\gamma}'\mathbf{x}_i}{s_n} \right) \middle| \mathbf{w}_i(\tilde{\boldsymbol{\beta}}_n) \right] , \tag{2.16}$$

and let $\tilde{\boldsymbol{\gamma}}_n$ be a local minimum of $R(\cdot)$ such that $R(\tilde{\boldsymbol{\gamma}}_n) \le R(\mathbf{0})$. The MM-estimator $\hat{\boldsymbol{\beta}}_n$ for censored regression is defined by

$$\hat{\boldsymbol{\beta}}_n = \tilde{\boldsymbol{\beta}}_n + \tilde{\boldsymbol{\gamma}}_n \, . \tag{2.17}$$

The motivation for the definition in (2.17) is as follows. We improve the initial S-estimator $\tilde{\boldsymbol{\beta}}_n$ by fitting an efficient M-estimator to the residuals of $\tilde{\boldsymbol{\beta}}_n$. The resulting M-estimate $\tilde{\boldsymbol{\gamma}}_n$ is the required correction.

Expanding the conditional expectations in (2.16) we see that $R(\boldsymbol{\gamma})$ can also be written as

$$R(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \left[\delta_i \, \rho_2 \left(\frac{r_i(\tilde{\boldsymbol{\beta}}_n) - \boldsymbol{\gamma}'\mathbf{x}_i}{s_n} \right) + \frac{(1-\delta_i)}{1 - F^*_{\tilde{\boldsymbol{\beta}}_n}(r_i(\tilde{\boldsymbol{\beta}}_n))} \int_{r_i(\tilde{\boldsymbol{\beta}}_n)}^{\infty} \rho_2 \left(\frac{u - \boldsymbol{\gamma}'\mathbf{x}_i}{s_n} \right) dF^*_{\tilde{\boldsymbol{\beta}}_n}(u) \right] . \tag{2.18}$$

For each i such that $\delta_i = 0$ let $M_i = \left\{ j : r_j(\tilde{\boldsymbol{\beta}}_n) > r_i(\tilde{\boldsymbol{\beta}}_n), \delta_j = 1 \right\}$. Then, we have

$$\int_{r_i(\tilde{\boldsymbol{\beta}}_n)}^{\infty} \rho_2 \left(\frac{u - \boldsymbol{\gamma}'\mathbf{x}_i}{s_n} \right) dF^*_{\tilde{\boldsymbol{\beta}}_n}(u) = \sum_{j \in M_i} \rho_2 \left(\frac{r_j(\tilde{\boldsymbol{\beta}}_n) - \boldsymbol{\gamma}'\mathbf{x}_i}{s_n} \right) \pi_j \, , \tag{2.19}$$

and

$$1 - F^*_{\tilde{\boldsymbol{\beta}}_n}(r_i(\tilde{\boldsymbol{\beta}}_n)) = \sum_{j \in M_i} \pi_j \, ,$$

where π_j, $j \in M = \{j : \delta_j = 1\}$, are the probabilities given to the uncensored $r^*_j(\tilde{\boldsymbol{\beta}}_n)$ by the KM estimator $F^*_{\tilde{\boldsymbol{\beta}}_n}$. For $i, j = 1, \ldots, n$ let

$$\pi_{ij} = \begin{cases} \pi_j / \sum_{k \in M_i} \pi_k & \text{if } \delta_i = 0 \text{ and } j \in M_i \\ 1 & \text{if } \delta_i = 1 \text{ and } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Then, from (2.18) and (2.19) we have

$$R(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_2 \left(\frac{r_j(\tilde{\boldsymbol{\beta}}_n) - \boldsymbol{\gamma}' \mathbf{x}_i}{s_n} \right) \pi_{ij} \,. \tag{2.20}$$

Since the π_{ij}'s do not depend on $\boldsymbol{\gamma}$, a local minimum of $R(\boldsymbol{\gamma})$ will satisfy

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \rho_2' \left(\frac{r_j(\tilde{\boldsymbol{\beta}}_n) - \boldsymbol{\gamma}' \mathbf{x}_i}{s_n} \right) \mathbf{x}_i \pi_{ij} = \mathbf{0} \,.$$

Similarly to the uncensored case, this equation can be written as

$$\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (r_j(\tilde{\boldsymbol{\beta}}_n) - \mathbf{x}_i' \boldsymbol{\gamma}) \mathbf{x}_i = \mathbf{0} \,,$$

where

$$w_{ij} = \frac{\rho_2' \left(\frac{r_j(\tilde{\boldsymbol{\beta}}_n) - \boldsymbol{\gamma}' \mathbf{x}_i}{s_n} \right)}{\left(\frac{r_j(\tilde{\boldsymbol{\beta}}_n) - \boldsymbol{\gamma}' \mathbf{x}_i}{s_n} \right)} \pi_{ij} \,.$$

Hence, a local minimum of $R(\boldsymbol{\gamma})$ is the weighted least squares estimator for the points $(r_i(\tilde{\boldsymbol{\beta}}_n), \mathbf{x}_j)$ with weights w_{ij}, $i, j = 1, \ldots, n$. Note that, in principle, this formulation involves n^2 points, however in practice many of them receive weight zero.

On the other hand, since we need to find a local minimum such that $R(\boldsymbol{\gamma}) < R(\mathbf{0})$, and re-weighted least squares iterations reduce the objective function [see Remark 1 to Lemma 8.3 in Huber (1981, page 186), which was originally proved by Dutter (1975)], we can start the iterative weighted least squares algorithm at zero.

2.4 τ-estimators

Another way to obtain estimators with high breakdown and high efficiency for normal errors with censored responses, is to extend the class of τ-estimators [Yohai and Zamar (1988)]. These estimators are based on an efficient scale estimator, called τ-scale.

Let $\rho_1 : \mathbb{R} \to \mathbb{R}^+$ and $\rho_2 : \mathbb{R} \to \mathbb{R}^+$ satisfy conditions P1-P4, and let $b = E_F(\rho_1)$. Moreover, to obtain consistent estimators, we will assume that ρ_1 and ρ_2 satisfy

P5: ρ_i, $i = 1, 2$, are continuous, and if $0 \le v < w$ with $\rho_2(w) < \sup_u \rho_2(u)$ then $\rho_2(v) < \rho_2(w)$;

P6: $2\,\rho_2(u) - \rho_2'(u)\,u \ge 0$.

Given a sample $u_1, ..., u_n$ let s_n be the solution of

$$\frac{1}{n}\sum_{i=1}^{n} \rho_1\left(u_i/s_n\right) = b\,,$$

and define the τ-scale as

$$\tau_n^2 = s_n^2 \frac{1}{n}\sum_{i=1}^{n} \rho_2\left(u_i/s_n\right)\ .$$

The extension of the τ-estimators for censored data follows the same lines as the one for S-estimators but using a τ-scale instead of an S-scale.

More specifically, let $S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ be as in (2.11) and define $\tau_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ by

$$\tau_n(\boldsymbol{\beta}, \boldsymbol{\gamma})^2 = S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})^2\,\frac{1}{n}\sum_{i=1}^{n} E_{F^*_{n\boldsymbol{\beta}}}\left[\rho_2\left(\frac{u - \boldsymbol{\gamma}'\mathbf{x}_i}{S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})}\right)\middle|\,\mathbf{w}_i(\boldsymbol{\beta})\right]\ . \tag{2.21}$$

Let

$$\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^p} \tau_n(\boldsymbol{\beta}, \boldsymbol{\gamma})\,, \tag{2.22}$$

and define the τ-estimator $\hat{\boldsymbol{\beta}}_n$ as in (2.9) or (2.10). Note that $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})$ in (2.22) may be considered the τ-estimator of the residuals $(r_i^*(\boldsymbol{\beta}), \mathbf{x}_i')'$, $i = 1, \ldots, n$.

2.5 Alternative representation

In this section we show an alternative way of writing the estimating equations that define our estimators for censored data. This alternative representation is most useful when computing these estimators for non-smooth functions $\rho(u)$ (e.g. the LMS). We will also use this representation in our proofs in the Appendix. This approach also lets us understand better the connection between the estimators defined in the previous sections and their uncensored counterparts.

Let $r_1, \ldots, r_n$ be a random sample from a distribution F, and let $c_1, \ldots, c_n$ be unobservable censoring variables independent from the r_i's. Suppose that we observe $r_i^* = \min(r_i, c_i)$ and let $\delta_i = I\{r_i \le c_i\}$ where $I\{A\}$ is the indicator function of the event A. The Kaplan-Meier estimator of F assigns positive weights π_i to the non-censored observations and zero otherwise. The self-consistency property of the Kaplan-Meier estimator (Efron, 1967) implies that, if π_i is the probability assigned to r_i^* for $\delta_i = 1$, then

$$\pi_i = \frac{1}{n} + \sum_{r_i^* > r_{j,}^* \delta_j = 0} \pi_{ij} \, ,$$

where

$$\pi_{ij} = \frac{1}{n} \frac{\pi_i}{\sum_{r_h^* > r_{j,}^* \delta_h = 1} \pi_h} \qquad \text{if} \;\; r_i^* > r_j^* \, , \;\; \delta_i = 1 \, , \;\; \delta_j = 0 \; .$$

Observe that π_{ij} can be interpreted as the proportion of the mass from the censored j-th observation that is assigned to the i-th point. Note that the mass $1/n$ of each censored observation r_j^* is distributed among all the uncensored $r_i^* > r_j^*$ with $\delta_i = 1$ proportionally to π_i.

Suppose that $r_i^* = r_i^*(\boldsymbol{\beta})$ for $1 \leq i \leq n$ are residuals for some vector of regression parameters $\boldsymbol{\beta}$, and let $\mathbf{x}_i$, $1 \leq i \leq n$ be the corresponding vectors of covariates. The censored residual sample can be written as $\mathbf{z}_1 = (r_1^*, \delta_1, \mathbf{x}_1')', \ldots, \mathbf{z}_n = (r_n^*, \delta_n, \mathbf{x}_n')'$. Consider the following probability mass function

$$h_n^*(r_i^*, \mathbf{x}_j, \delta_i, \delta_j) = \begin{cases} 1/n & \text{if } \delta_i = 1, i = j \\ \pi_{ij} & \text{if } \delta_i = 1, \delta_j = 0, r_i^* > r_j^* \\ 0 & \text{otherwise,} \end{cases} \tag{2.23}$$

and let $H_{n\beta}^*$ be the corresponding distribution function.

Following the same arguments leading to (2.20) it is easy to show that for any function $g : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$ we have

$$\frac{1}{n} \sum_{i=1}^{n} E_{F_{n\boldsymbol{\beta}}^*} \left[g(u, \mathbf{x}_i) | \mathbf{z}_i \right] = E_{H_{n\boldsymbol{\beta}}^*} \left[g(u, \mathbf{x}) \right] \, . \tag{2.24}$$

Then, according to (2.24) the formula for $C_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ in (2.7) can be written as

$$C_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = E_{H_{n,\boldsymbol{\beta}}^*} \left[\rho \left(\frac{r - \boldsymbol{\gamma}'\mathbf{x}}{s_n} \right) \right]$$

which makes some computations faster. For example, consider the jump function ρ defined in (2.15) and the solution s_n to

$$E_{H_{n\boldsymbol{\beta}}^*} \left[\rho(u/s_n) \right] = 1/2 \, .$$

Since this equation is equivalent to

$$E_{F_{n\boldsymbol{\beta}}^*} \left[\rho(u/s_n) \right] = 1/2 \, ,$$

we have that $s_n = \text{median}_{H_{n\boldsymbol{\beta}}^*}(|u|) = \text{median}_{F_{n\boldsymbol{\beta}}^*}(|u|)$ and thus iterative algorithms are not required.

The following theorem shows that $H_{n\beta}^*$ is consistent to the true joint distribution function $H(u, \mathbf{x}) = F(u)G(\mathbf{x})$ when $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Moreover, Theorem 7 in the Appendix, shows that if $\boldsymbol{\beta}_n \xrightarrow{P} \boldsymbol{\beta}_0$, then $H_{n\boldsymbol{\beta}_n}^* \xrightarrow{P} H(u, \mathbf{x})$.

Theorem 1 *Let $(y_i^*, \mathbf{x}_i, \delta_i)$, $i = 1, \ldots, n$ be observations from a censored linear regression model as in Section 2, and assume that the errors and censoring variables satisfy R1 and R2 on page 5. Let $H^*_{n\boldsymbol{\beta}}$ be defined as above. Then*

$$H^*_{n\boldsymbol{\beta}_0}(u, \mathbf{x}) \;\to\; H(u, \mathbf{x}) \;\; a.s.$$

3 Properties

In general, for a sample $\mathbf{Z}_n$ of size n, the finite-sample breakdown point (Donoho and Huber, 1983) of an estimator $\mathbf{T}_n = \mathbf{T}_n(\mathbf{Z}_n)$ is defined as

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

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 $\|\mathbf{T}\|$ 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 $\mathbf{z}_i = (y_i^*, \mathbf{x}_i, \delta_i)$, $\mathbf{x}_i \in R^p$. Assume that the rank of $\{\mathbf{x}_1, \ldots \mathbf{x}_n\}$ is p and let $q = \max_{||\theta||=1} \#\{i : \theta'\mathbf{x}_i = 0\}$. Let m be the number of censored observations in the sample, $m = \sum_{i=1}^n \delta_i$. The following theorems show that a lower bound for the breakdown point of S-, MM- and τ-regression estimators is

$$\gamma = k_0/n \,, \tag{3.1}$$

where

$$k_0 = \min\left(n\left(1 - \frac{b}{a}\right) - q - m \,,\, n\,\frac{b}{a} - m \right) \,, \tag{3.2}$$

b is the right-hand side of equation (2.11) and $a = \sup_u \rho(u)$.

Theorem 2 – **Breakdown Point of S-estimators** – *Let S be a scale estimating functional based on a function ρ satisfying P1-P4. Let $\hat{\boldsymbol{\beta}}_n$ be the S-estimator defined in Sub-section 2.2, then*

$$\epsilon_n^*(\hat{\boldsymbol{\beta}}_n, \mathbf{Z}) \,\ge\, \gamma \,. \tag{3.3}$$

Theorem 3 – **Breakdown Point of MM-estimators** *Let $\widehat{\beta}_n$ be the MM estimator defined in Sub-section 2.3 with loss functions ρ_1 and ρ_2 satisfying P1-P4, $\rho_2 \le \rho_1$ and $a = \sup \rho_2 = \sup \rho_1$. Then*

$$\epsilon_n^*(\hat{\boldsymbol{\beta}}_n, \mathbf{Z}) \,\ge\, \gamma \,.$$

Theorem 4 – **Breakdown Point of τ-estimators** *Let $\hat{\boldsymbol{\beta}}_n$ be the τ-estimator defined in Subsection 2.4 with loss functions ρ_1 and ρ_2 satisfying P1-P6. Then*

$$\epsilon_n^*(\hat{\boldsymbol{\beta}}_n, \mathbf{Z}) \geq \gamma\,.$$

Note that the lower bound in (3.1) is maximized when

$$\frac{b}{a} = \frac{1 - q/n}{2}\,.$$

The smallest possible value of q is $p-1$, and in this case the sample is said to be in general position (Rousseeuw and Leroy, 1987). Using the optimal b/a we have

$$\epsilon_n^*(\hat{\boldsymbol{\beta}}_n, \mathbf{Z}_n) \;\geq\; \frac{1}{2}\left(\frac{n - p + 1 - 2\,m}{n}\right)\,.$$

Note that when $n \to \infty$ the right-hand side converges to $1/2 - \lambda$, where λ is the probability of censoring. This is in agreement with our discussion in the Introduction, where we mention that the breakdown point of the median may be as small as 1/2 - λ when there are censored observations. Although in linear regression models with uncensored response variables it is possible to obtain robust regression estimators with asymptotic breakdown point of 0.5, we believe that the loss in breakdown-point observed in the censored case is due to the use of the Kaplan-Meyer estimator that may convert censored observations into outliers. We conjecture that this loss can not be to reduced, at least when the estimate is defined using the Kaplan-Meyer estimate.

4 Asymptotic Properties

The next theorem shows a property related to the consistency of the S-estimator defined in Subsection 2.2.

Theorem 5 - *Let ρ satisfy regularity conditions P1-P4. Let the errors u and covariates $\mathbf{x}$ in the linear model (1.1) have joint distribution function $H_0(u, \mathbf{x}) = F_0(u)\,G(\mathbf{x})$ such that:*

(i) $F_0(u)$ is symmetric and has a unimodal density;

(ii) $G(\boldsymbol{\beta}'\mathbf{x} \neq 0) = t > b/a$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$.

Assume that R1 and R2 on page 5 hold and let $\boldsymbol{\gamma}_n(\boldsymbol{\beta}_0) = \arg\min_{\boldsymbol{\gamma}} S_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma})$, where $S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is defined in (2.11). Then

$$\boldsymbol{\gamma}_n(\boldsymbol{\beta}_0) \xrightarrow[n\to\infty]{a.s} \mathbf{0}\,. \tag{4.1}$$

The same kind of arguments used in the proof of Theorem 5 can be used to prove similar results for MM- and τ-estimators as defined in Sub-sections 2.3 and 2.4 respectively.

Note that a complete proof of consistency would require to show that if $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ then $\|\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})\|$ remains asymptotically away from zero. We have not been able to prove this. However, in all our numerical experiments this property seems to hold.

We can, however, prove the local consistency and asymptotic normality of the M-estimates defined in Section 2.1. The proof is based on Theorem 5.1 in Ritov (1990). In that paper the author studies M-estimates for censored regression which are defined as solutions of (2.6). Unfortunately, showing that there exists a sequence of consistent solutions of this equation seems to be very difficult. However, it can be shown that there exists a sequence $\boldsymbol{\beta}_n$ of approximate solutions to this equation which is $\sqrt{n}$-consistent and asymptotically normal. More precisely, under some regularity conditions there exists a sequence $\boldsymbol{\beta}_n$ such that

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} E_{F^*_{n\boldsymbol{\beta}_n}} [\psi(u)|\mathbf{w}_i(\boldsymbol{\beta}_n)]\, \mathbf{x}_i \xrightarrow{P} \mathbf{0}\,, \tag{4.2}$$

and such that $\sqrt{n}\,(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0) \xrightarrow{D} N(0,\ A_\psi^{-1}\, B_\psi\, A_\psi)$ where

$$A_\psi \; = \int E(\mathbf{x}\mathbf{x}'|C - \boldsymbol{\beta}_0'\mathbf{x} \geq u) W_\psi(u) W_{\psi_0}(u) P(C - \boldsymbol{\beta}_0'\mathbf{x} \geq u) dF_0(u), \tag{4.3}$$

where C is the censoring variable,

$$W_\psi(u) = \psi(u) - \frac{\int_u^\infty \psi(t) dF_0(t)}{1 - F_0(u)},$$

$$\psi_0(u) = -\frac{f_0'(u)}{f_0(u)}\,,$$

and

$$B_\psi \; = \int E(\mathbf{x}\mathbf{x}'|C - \beta_0'\mathbf{x} \geq u) W_\psi^2(u) P(C - \boldsymbol{\beta}_0'\mathbf{x} \geq u) dF_0(u)\,. \tag{4.4}$$

We will show a similar result for the estimates defined by (2.9). To simplify the proofs we will only consider the case where the error scale σ is known. More precisely, we have the following Theorem:

Theorem 6 *Assume that*

1. *ρ satisfies P1, P2 and P3 and P4 and is three times continuously differentiable with bounded derivatives. Moreover there exists c_0 such that $\rho(c_0) = \max_u \rho(u)$ and $P(\min(Y, C) - \boldsymbol{\beta}'\mathbf{x} < c_0) < 1$ for all $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$;*

2. *the errorrs u_i have a symmetric and a strictly unimodal density f_0 with finite information for location,i.e.* $\int_{-\infty}^{\infty} \left(\frac{f_0'(u)}{f_0(u_0)}\right)^2 f_0(u_0) < \infty$;

3. *the vector of explanatory variables* $\mathbf{x}$ *has compact support; and*

4. *the matrix A defined in (4.3) is non singular.*

Then, there exists a sequence $\boldsymbol{\beta}_n$ *such that (i)* $\sqrt{n}\,\boldsymbol{\gamma}_n(\boldsymbol{\beta}_n) \xrightarrow{P} 0$ *and (ii)* $\sqrt{n}\,(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0}, A_\psi^{-1}\, B_\psi\, A_\psi^{-1})$, *where* A_ψ *and* B_ψ *are defined in (4.3) and (4.4), respectively.*

Consider a differentiable function $\rho(u)$ satisfying P1-P4, and let $\rho' = \psi$ with $\psi(0) = a_0 > 0$. For $c > 0$ let $\rho_c(u) = (c/a_0)\rho(u/c)$ and $\psi_c(u) = \rho_c'(u) = (1/a_0)\psi(u/c)$. Then the functions ρ_c satisfy P1-P4 and $\lim_{c\to\infty} \psi_c(u) = u = \psi^*(u)$. It is possible to show that $A_{\psi_c} \to A_{\psi^*}$ and $B_{\psi_c} \to B_{\psi^*}$. Therefore, when $c \to \infty$ the relative asymptotic efficiency of the proposed M-estimate with respect to the Buckley and James estimate tends to 1. Choosing c large enough, this relative efficiency can be as close to 1 as desired. For example, this can be obtained using $\rho(u) = \rho_T(u)$ Tukey's bi-square function with derivative

$$\psi_T(u) = u(1-u^2)^2 I(|u| \leq 1)\,,$$

where $I(|u| \leq 1) = 1$ if $|u| \leq 1$ and 0 otherwise.

5 Computing algorithm

To compute the estimators proposed in Section 2 we have to solve highly complex optimization problems. For example, to obtain the S-estimators of Sub-section 2.2 for each vector $\boldsymbol{\beta}$ we have to find $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})$ in (2.12) minimizing a non-convex function for which there is no closed-form expression. Moreover, we then have to minimize $\|\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})\|$ over all vectors $\boldsymbol{\beta}$.

In this section we present an efficient algorithm to compute these S-estimators that is an improvement over a naive re-sampling-based approximation [Rousseeuw (1984)].

We will follow a widely used strategy to approximate the solution of complex optimization problems in robust statistics. This approach is based on generating a large number N of candidate vectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N$. The estimator is then approximated by the best candidate $\hat{\boldsymbol{\beta}}_n$. One way to generate these candidates is by drawing sub-samples of size p from the data and adjusting them. Intuitively, the idea behind this approach is that if N is large enough some of the sub-samples will not have outliers and that we can find a good solution among the corresponding candidates.

Let $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$ be a random sample satisfying the linear regression model (1.1). For each pair $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^p$ let $S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ be the M-scale estimator defined in (2.11). Let $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N$ be the resampling candidates described above. The estimator $\hat{\boldsymbol{\beta}}_n$ satisfies

$$\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}_k \,,$$

where

$$\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_k)' \, \mathbf{A}_n \, \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_k) = \min_{1 \le j \le N} \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_j)' \, \mathbf{A}_n \, \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_j) \,.$$

We now turn our attention to the calculation of $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_j)$ for each candidate $\boldsymbol{\beta}_j$. Recall that this requires to solve the minimization problem given by (2.12). We will use the same approach: for each $\boldsymbol{\beta}_j$ consider a large number of candidates for $\boldsymbol{\gamma}$ and set $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_j)$ to be the best of these candidates. Note that for each fixed $\boldsymbol{\beta}_j$ the vector $\boldsymbol{\beta}_r - \boldsymbol{\beta}_j$ is a natural candidate for $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_j)$ when $\boldsymbol{\beta}_r$ is a good approximation for the true $\boldsymbol{\beta}$. This follows by observing that in this case the residuals $r_i(\boldsymbol{\beta}_j)$ will follow a linear regression model with coefficients $\boldsymbol{\beta} - \boldsymbol{\beta}_j$ which are close to $\boldsymbol{\beta}_r - \boldsymbol{\beta}_j$. Then, we approximate $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_j)$ by the vector $\boldsymbol{\beta}_r - \boldsymbol{\beta}_j$ satisfying

$$S_n(\boldsymbol{\beta}_j, \boldsymbol{\beta}_r - \boldsymbol{\beta}_j) = \min_{1 \le i \le N} S_n(\boldsymbol{\beta}_j, \boldsymbol{\beta}_i - \boldsymbol{\beta}_j) \,.$$

Note that, in principle, this algorithm requires finding N^2 scales $S_n(\boldsymbol{\beta}_j, \boldsymbol{\beta}_i - \boldsymbol{\beta}_j)$, $i, j = 1, \ldots, n$. The number of re-sampling candidates N required to obtain a good approximation can be determined explicitly as in the uncensored case (see, for example, Rousseeuw and Leroy, 1987). This value increases exponentially with the number of covariates in the model. In what follows we present an algorithm that avoids the computation of all the N^2 scales.

Suppose that we have already computed $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_j)$ for $j = 1, \ldots, i$ and let

$$\kappa_i = \min_{1 \le j \le i} \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_j)' \, \mathbf{A}_n \, \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_j) \,,$$

the best value of the objective function obtained so far. We will need to compute $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_{i+1})$ only if

$$\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_{i+1})' \, \mathbf{A}_n \, \hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_{i+1}) < \kappa_i \,.$$

Divide the set of candidates for $\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_{i+1})$ into two sets: those with $(\boldsymbol{\beta}_k - \boldsymbol{\beta}_{i+1})' \, \mathbf{A}_n \, (\boldsymbol{\beta}_k - \boldsymbol{\beta}_{i+1}) \ge \kappa_i$ (call them $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{N_1}$) and those with $(\boldsymbol{\beta}_k - \boldsymbol{\beta}_{i+1})' \, \mathbf{A}_n \, (\boldsymbol{\beta}_k - \boldsymbol{\beta}_{i+1}) < \kappa_i$ (call them $\tilde{\boldsymbol{\gamma}}_1, \ldots, \tilde{\boldsymbol{\gamma}}_{N_2}$). Note that $\|\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_{i+1})\| < \kappa_i$ only if

$$\min_{1 \le j \le N_1} S_n(\boldsymbol{\beta}_{i+1}, \boldsymbol{\gamma}_j) > \min_{1 \le j \le N_2} S_n(\boldsymbol{\beta}_{i+1}, \tilde{\boldsymbol{\gamma}}_j) \,.$$

Hence, we first compute $\omega = \min_{1 \le j \le N_2} S_n(\boldsymbol{\beta}_{i+1}, \tilde{\boldsymbol{\gamma}}_j)$. Then we compare each $S_n(\boldsymbol{\beta}_{i+1}, \boldsymbol{\gamma}_m)$ for $m = 1, \ldots, N_1$ with ω. If for some m_0 we find $S_n(\boldsymbol{\beta}_{i+1}, \boldsymbol{\gamma}_{m_0}) < \omega$ then we stop and set $\kappa_{i+1} = \kappa_i$.

Since $\kappa_i \to 0$ we expect $E(N_1)$ to decrease as well. Our Monte Carlo experiments show that there is a substantial gain in speed with this modified algorithm.

6 Example

Consider the Heart dataset analyzed in Kalbfleisch and Prentice (1980). These data contain information on heart transplant recipients, including their age and their survival times, which are censored in some cases. In Figure 2 we plot Log(Survival time) versus Age for these patients. We indicate un-censored cases with the symbol "1" and censored ones with "0"s. In the same figure we also show the fitted lines corresponding to our modified extensions of the LS and MM-estimators. Note that the LS estimator is very much influenced by the early death of two young patients, that can be considered outliers. We used small diamonds around these points to identify them on the plot. We also plot the same LS fit with these two points removed. Note that this line is now close to the robust fit.

7 Monte Carlo study

To study the finite-sample properties of these estimators we performed a Monte Carlo study for the simple regression model:

$$y_i = \alpha + \beta x_i + u_i \,, \qquad i = 1, \ldots, n \,.$$

We considered 1000 samples of size $n = 100$, independent normal errors $u_i \sim \mathcal{N}(0, 1)$, random covariates $x_i \sim \mathcal{N}(0, 1)$ independent from the errors, $\alpha = 0$ and $\beta = 1.5$. We used censoring random variables $c_1, \ldots, c_n$ that were sampled from an independent random variable with distribution $\mathcal{N}(1, 1)$. With these choices we have $P(\delta = 0) = 0.32$.

We included the consistent versions under censoring proposed in this paper of the following estimators: the least squares estimator (LS), the least median of squares (LMS), an S-estimator (S) with 50% breakdown point when there is no censoring in the sample, an MM-estimator (MM) with 95% efficiency under normal errors and no censoring, the L1-estimator (L1) [an M-estimator with

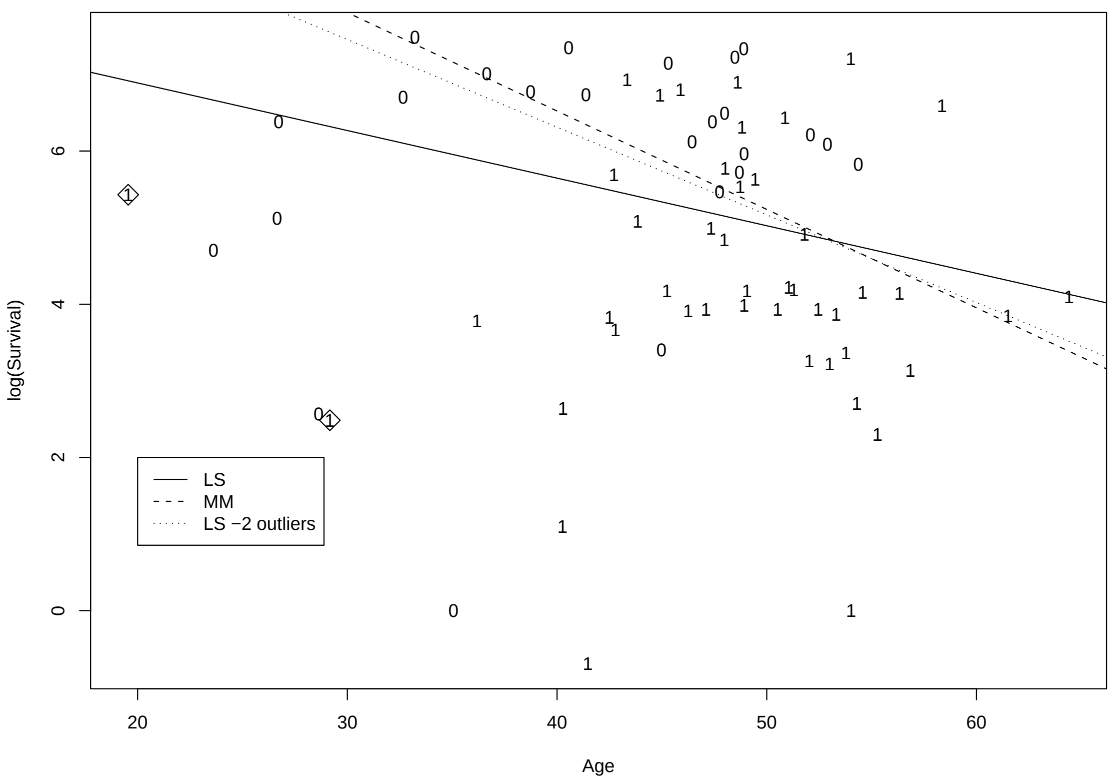


Figure 2: Heart transplant data. “1”s indicate deaths, “0”s indicate censored observations. The least squares estimator seems to be influenced by the two young patients that die early in the study.

$\psi(x) = \text{sign}(x)]$, and the GM estimator defined by

$$\sum_{i=1}^{n} E_{F^*_{n\,\beta}}[\psi_1(u - \alpha(\beta))|\mathbf{w}_{\beta i}]\, \psi_2\,(x_i - m_x) = \mathbf{0}\,,$$

where $\psi_1(x) = \psi_2(x) = \text{sgn}(x)$, $\alpha(\beta) = \text{median}(F^*_{n\,\beta})$, and $m_x = \text{median}(x_1, \ldots, x_n)$. This is the analogous to the Mood-Brown estimator with breakdown point 1/4. Both the S- and the MM-estimators used ρ functions in the bi-square family.

The samples were contaminated with 10% of outliers (10 observations). These 10 observations were changed to the points $(x_0, m\,x_0)$ where x_0 was set at 1 and 10 (resulting in low and high leverage outliers respectively), and m ranged between 2 and 5.

In Table 1 we report the MSE for β when there are no outliers in the sample. Tables 2 and 3 contain the MSE's for β for the cases $x_0 = 1$ and $x_0 = 10$ respectively.

From Table 1, we see that, as expected, the most efficient estimator is the LS, followed by the L1 and the MM with efficiencies of 76% and 70% respectively.

For low leverage contaminations (Table 2) the two estimators that perform better, from a maximum MSE point of view, are the L1 and the MM. These two estimators have a similar behavior with a small advantage of the MM. The other estimators are notably worse.

Table 3 shows that for high-leverage outliers the MM estimator had the smallest MSE, followed by the S-estimator. Not surprisingly, both the LS and L1 estimators have noticeably worse MSE's than all the other estimators considered here.

Based on these results, we may conclude that the MM-estimators have the best overall performance.

Estimates	MSE
S	0.060
LMS	0.164
LS	0.019
MM	0.027
GM	0.046
L1	0.025

Table 1: MSE's without outliers.

Estimator	Slopes						
	2	2.5	3	3.5	4	4.5	5
S	0.10	0.27	0.38	0.30	0.20	0.13	0.10
LMS	0.14	0.30	0.54	0.69	0.79	0.76	0.78
LS	0.03	0.05	0.10	0.15	0.23	0.33	0.43
MM	0.04	0.11	0.17	0.18	0.18	0.19	0.20
GM	0.09	0.25	0.40	0.52	0.62	0.71	0.78
L1	0.07	0.16	0.20	0.21	0.21	0.21	0.21

Table 2: MSE's with 10% of outliers at $x_0 = 1$.

Estimator	Slopes						
	2	2.5	3	3.5	4	4.5	5
S	0.25	0.50	0.34	0.20	0.11	0.08	0.10
LMS	0.31	0.45	0.58	0.65	0.49	0.40	0.38
LS	0.24	0.90	1.98	3.44	5.09	6.61	7.61
MM	0.23	0.45	0.30	0.17	0.08	0.06	0.07
GM	0.15	0.39	0.56	0.69	0.79	0.92	1.08
L1	0.25	0.93	2.04	3.59	5.63	8.08	11.03

Table 3: MSE's with 10% of outliers at $x_0 = 10$.

8 Appendix – Proofs

8.1 Consistency of $H^*_{n\beta_0}$

Proof of Theorem 1: Fix $(a, \mathbf{v}')' \in \mathbb{R}^{p+1}$ and note that $H^*_{n\boldsymbol{\beta}_0}(a, \mathbf{v}) = E_{H^*_{n\boldsymbol{\beta}_0}}[I(u \le a, \mathbf{x} \le \mathbf{v})]$ where $I(A)$ denotes the indicator function of the event A. Let $r_i^* = y_i^* - \boldsymbol{\beta}_0'\mathbf{x}_i$ for $i = 1, \ldots, n$. Using (2.24) we have

$$H^*_{n\boldsymbol{\beta}_0}(a, \mathbf{v}) = E_{H^*_{n\boldsymbol{\beta}_0}}[I(u \le a, \mathbf{x} \le \mathbf{v})] = \frac{1}{n}\sum_{i=1}^{n}\Big\{\delta_i\, I(r_i^* \le a, \mathbf{x}_i \le \mathbf{v}) + (1-\delta_i)\, E_{F^*_{n,\boldsymbol{\beta}_0}}[I(u \le a, \mathbf{x}_i \le \mathbf{v})|u > r_i^*]\Big\}\ .$$

Adding and substracting

$$\sum_{i=1}^{n}(1-\delta_i)\, E_H[I(u \le a, \mathbf{x} \le \mathbf{v})|u > r_i^*]$$

we obtain

$$H^*_{n\boldsymbol{\beta}_0}(a, \mathbf{v}) - H(a, \mathbf{v}) = \frac{1}{n}\sum_{i=1}^{n}[\tilde{g}(r_i^*, \mathbf{x}_i) - H(a, \mathbf{v})] + \frac{1}{n}\sum_{i=1}^{n}\Big[(1-\delta_i)\,\Big(E_{F^*_{n,\boldsymbol{\beta}_0}}[g(u, \mathbf{x}_i)|\, u > r_i^*] - E_H(g(u, \mathbf{x})|\, u > r_i^*)\Big)\Big]\ , \quad (8.1)$$

where

$$\tilde{g}(r_i^*, \mathbf{x}_i) = \delta_i\, I(r_i^* \le a, \mathbf{x}_i \le \mathbf{v}) + (1-\delta_i)\, E_H[I(u \le a, \mathbf{x} \le \mathbf{v})|u > r_i^*]\,,$$

H denotes the joint distribution of the vector $(\mathbf{x}', u)'$ and $g(u, \mathbf{x}) = I(u \le a, \mathbf{x} \le \mathbf{v})$. Also note that $E_H[\tilde{g}(r_i^*, \mathbf{x}_i)] = H(a, \mathbf{v})$. Since $\tilde{g}$ is bounded, Kolmogorov's Law of Large numbers yields

$$\frac{1}{n}\sum_{i=1}^{n}[\tilde{g}(r_i^*, \mathbf{x}_i) - H(a, \mathbf{v})] \xrightarrow[n\to\infty]{a.s.} 0\,.$$

Moreover, note since $g(u, \mathbf{x}) = I(u \le a)\, I(\mathbf{x} \le \mathbf{v})$ we have

$$E_{F^*_{n,\boldsymbol{\beta}_0}}[g(u, \mathbf{x}_i)|\, u > r_i^*] = I(\mathbf{x}_i \le \mathbf{v})\, E_{F^*_{n,\boldsymbol{\beta}_0}}[d(u)|\, u > r_i^*]\ ,$$

where $d(u) = I(u \le a)$. Also, because of the independence between u and $\mathbf{x}$ we have $E_H(\,g(u,\mathbf{x})|\,u > r_i^*) = P(\mathbf{x} \le \mathbf{v})\,E_F(\,d(u)|\,u > r_i^*)$. Hence, the second term in (8.1) equals

$$\frac{1}{n}\sum_{i=1}^{n}\Big[(1-\delta_i)\,I(\mathbf{x}_i \le \mathbf{v})\,E_{F^*_{n,\boldsymbol{\beta}_0}}\,[d(u)|\,u > r_i^*]$$
$$-P(\mathbf{x} \le \mathbf{v})\,E_F(\,d(u)|\,u > r_i^*)\Big]$$
$$= \frac{1}{n}\left[\sum_{i=1}^{n}(1-\delta_i)\,I(\mathbf{x}_i \le \mathbf{v})\,\left(E_{F^*_{n,\boldsymbol{\beta}_0}}\,[d(u)|\,u > r_i^*] - E_F(\,d(u)|\,u > r_i^*)\right)\right.$$
$$\left. +(1-\delta_i)\,E_F(\,d(u)|\,u > r_i^*)\,(I(\mathbf{x}_i \le \mathbf{v}) - P(\mathbf{x} \le \mathbf{v}))\right].$$

Let

$$R_n = \sum_{i=1}^{n}\left[(1-\delta_i)\,I(\mathbf{x}_i \le \mathbf{v})\right]/n\,,$$

and note that $|R_n| \le 1$. Since d is bounded, by Kolmogorov's Law of Large Numbers we have

$$\sum_{i=1}^{n}\left[(1-\delta_i)\,E_F(\,d(u)|\,u > u_i)\,(I(\mathbf{x}_i \le \mathbf{v}) - P(\mathbf{x} \le \mathbf{v}))\right]/n \xrightarrow[n\to\infty]{a.s.} 0\,.$$

Thus, we only need to show that

$$\sup_{b\in\mathbb{R}}\left|E_{F^*_{n,\boldsymbol{\beta}_0}}\,[d(u)|\,u > b] - E_F\,[d(u)|\,u > b]\right| \xrightarrow[n\to\infty]{a.s.} 0\,. \tag{8.2}$$

First, note that we only need to consider the supremum over the set $b \le a$, since

$$E_{F^*_{n,\boldsymbol{\beta}_0}}\,[d(u)|\,u > b] = E_F\,[d(u)|\,u > b] = 0 \qquad \text{for } b > a\,.$$

Next, note that $E_F\,[d(u)|\,u > b] = (F(a) - F(b))/(1 - F(b))$. Thus, we need to bound

$$\sup_{b\le a}\left|\frac{F^*_{n,\boldsymbol{\beta}_0}(a) - F^*_{n,\boldsymbol{\beta}_0}(b)}{1 - F^*_{n,\boldsymbol{\beta}_0}(b)} - \frac{F(a) - F(b)}{1 - F(b)}\right|$$
$$= \sup_{b\le a}\left|\frac{(F^*_{n,\boldsymbol{\beta}_0}(a) - F(a)) - (F^*_{n,\boldsymbol{\beta}_0}(b) - F(b))}{(1 - F^*_{n,\boldsymbol{\beta}_0}(b))(1 - F(b))}\right.$$
$$\left. + \frac{F(a)(F^*_{n,\boldsymbol{\beta}_0}(b) - F(b)) + F(b)(F(a) - F^*_{n,\boldsymbol{\beta}_0}(a))}{(1 - F^*_{n,\boldsymbol{\beta}_0}(b))(1 - F(b))}\right|$$
$$\le 4\,\frac{\sup_b\left|F^*_{n,\boldsymbol{\beta}_0}(b) - F(b)\right|}{(1 - F^*_{n,\boldsymbol{\beta}_0}(a))(1 - F(a))}\,. \tag{8.3}$$

Since we are assuming R1 and R2 on page 5, Corollary 1.3 of Stute and Wang (1993) implies

$$\lim_{n\to\infty} \sup_b \left|F^*_{n,\beta_0}(b) - F(b)\right| = 0 \text{ a.s.}$$

This completes the proof. $\Box$

Theorem 7 *Let $(y_i^*, \mathbf{x}_i, \delta_i)$, $i = 1, \ldots, n$ be observations from a censored linear regression model as in Section 2, and assume that the errors and censoring variables satisfy R1 and R2 on page 5. Furthermore, assume that $\boldsymbol{\beta}_n \xrightarrow{P} \boldsymbol{\beta}_0$ and let $H^*_{n\boldsymbol{\beta}}$ be defined as above. Then*

$$H^*_{n\boldsymbol{\beta}_n}(u, \mathbf{x}) \ \xrightarrow{P}\ H(u, \mathbf{x}) \, .$$

Proof: The proof follows the same steps as that of the previous theorem replacing $H^*_{n\boldsymbol{\beta}_0}$ by $H^*_{n,\hat{\boldsymbol{\beta}}_n}$. The only difference is that now we need to show that

$$\sup_b \left|F^*_{n,\hat{\boldsymbol{\beta}}_n}(b) - F(b)\right| \xrightarrow[n\to\infty]{P} 0 \, .$$

Lemmas 7.1 and 7.2 in Ritov (1990) show that

$$\sup_b \left|F^*_{n,\hat{\boldsymbol{\beta}}_n}(b) - F(b)\right| \le O_p\left(n^{-1/2}\right) + O\left(\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|\right) = o_p(1) \, ,$$

because $\hat{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}_0$. $\Box$

8.2 Breakdown point of the S-estimator

Regularity conditions: Define the M-scale estimator $S(F)$ for any arbitrary distribution function F by

$$S(F) = \inf\left\{s > 0 \, : \, E_F\left[\rho\left(x/s\right)\right] < b\right\} \, , \tag{8.4}$$

where $b \ge 0$ and $\rho : \mathbb{R} \to \mathbb{R}_+$ satisfies P1-P4 in Section 2.

The following lemma is needed to find the breakdown point of the S-estimators for censored observations.

Lemma 1 *Let $S(F)$ be a scale estimator defined by (8.4) where ρ satisfies properties P1-P4. Then we have:*

(a) Given any $K > 0$, and $C > b/a$ there exists K' such that if

$$P_F\{ \; |x| > K'\} > C \,, \tag{8.5}$$

then $S(F) > K$

(b) Given any $M > 0$ and $C < b/a$, there exist M' such that if

$$P_F\{ \; |x| > M) < C \,, \tag{8.6}$$

then $S(F) < M'$

Proof: Consider any $K > 0$. By assumption, there exists A large enough such that $C\,\rho(A) > b$. Put $K' = 2AK$. Let F be such that satisfies (8.5) and put $D = \{|x| > K'\}$. Then

$$\begin{aligned} E_F(\rho(x/2K)) &\geq E_F\left[\rho(x/2K) I\{D\}\right] \\ &\geq \rho(K'/2K) P_F(|x| > K') \\ &> \rho(A) C > b \,. \end{aligned}$$

Then $S(F) \geq 2K > K$ and (a) is proved.

To prove (b), let $M > 0$ be arbitrary, and let $D = \{|x| > M\}$. By assumption there exists $A > 0$ such that $\rho(A) < b - a\,C$, and put $M_1 = M/A$. Let F be such that (8.6) holds. Then

$$\begin{aligned} E_F\left[\rho(x/M_1)\right] &= E_F\left[\rho(x/M_1)\, I\{D^c\}\right] + E_F\left[\rho(x/M_1)\, I\{D\}\right] \\ &\leq \rho(M/M_1) + a\,C \\ &< b - a\,C + a\,C = b \end{aligned}$$

and therefore $S(F) \leq M_1$. Then (b) holds taking $M' = 2\,M_1$. $\square$

Given a distribution function H and a Borel set B, in the rest of the paper we will denote by $H(B)$ the probability of B under H, i.e. $H(B) = P_H(B)$.

Proof of Theorem 2: Observe that $S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ can be defined by

$$E_{H^*_{n,\boldsymbol{\beta}}}(\rho((r - \boldsymbol{\gamma}'\mathbf{x})/S_n(\boldsymbol{\beta}, \boldsymbol{\gamma}))) = b \,, \tag{8.7}$$

and $S_n(\boldsymbol{\beta}, \mathbf{0})$ by

$$E_{F^*_{n,\boldsymbol{\beta}}}(\rho(r/S_n(\boldsymbol{\beta}, \mathbf{0}))) = b \,. \tag{8.8}$$

Assume that (3.3) is not true. Then there exists a sequence of samples $\mathbf{Z}^{(j)} = (\mathbf{z}_1^{(j)}, \ldots, \mathbf{z}_n^{(j)})$, $1 \le j < \infty$, $\mathbf{z}_i^{(j)} = (y_i^{*(j)}, \mathbf{x}_i^{(j)}, \delta_i^{(j)})$ such that each $\mathbf{Z}^{(j)}$ differs from $\mathbf{Z}$ in t observations where t satisfies $t < k_0$, and such that if we call $\boldsymbol{\beta}_n^{(j)} = \hat{\boldsymbol{\beta}}_n(\mathbf{Z}^{(j)})$, then

$$\lim_{j\to\infty} ||\boldsymbol{\beta}_n^{(j)}|| = \infty \,. \tag{8.9}$$

Let $\boldsymbol{\gamma}_j(\boldsymbol{\beta})$ denote the function $\boldsymbol{\gamma}(\boldsymbol{\beta})$ defined in (2.8) when the sample is $\mathbf{Z}^{(j)}$. We will show that (8.9) is not possible by proving that

$$\lim_{j\to\infty} \|\boldsymbol{\gamma}_j(\boldsymbol{\beta}_n^{(j)})\| = \infty \,, \tag{8.10}$$

and that

$$\sup_j ||\boldsymbol{\gamma}_j\,(\mathbf{0})\,|| < \infty \,. \tag{8.11}$$

Let us start by proving (8.11). Assume that it is not true. Then without loss of generality we can assume that

$$\lim_{j\to\infty} ||\boldsymbol{\gamma}_j\,(\mathbf{0})\,|| = \infty \,, \tag{8.12}$$

and that

$$\lim_{j\to\infty} \frac{\boldsymbol{\gamma}_j\,(\mathbf{0})}{||\boldsymbol{\gamma}_j\,(\mathbf{0})\,||} = \boldsymbol{\lambda} \,. \tag{8.13}$$

We will show that this is not possible by proving that

$$\lim_{j\to\infty} S_n^{(j)}(\mathbf{0}, \boldsymbol{\gamma}_j(\mathbf{0})) = \infty \,, \tag{8.14}$$

and

$$\sup_j S_n^{(j)}(\mathbf{0}, \mathbf{0}) < \infty \,, \tag{8.15}$$

where $S_n^{(j)}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ denotes the function $S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ when the sample is $\mathbf{Z}^{(j)}$.

Let $F_{n,\beta,\gamma}^{*(j)}$ denote the distribution of $r - \boldsymbol{\gamma}'\mathbf{x}$ when $(r, \mathbf{x})$ has distribution $H_{n,\boldsymbol{\beta}}^*$ and the sample is $\mathbf{Z}^{(j)}$. Let

$$M = \max_{1\le i\le n} |y_i^*| + 1 \,. \tag{8.16}$$

Then the $y_i^{(j)*\prime}s$ in $\mathbf{Z}^{(j)}$ that are neither contaminated nor censored will have absolute value smaller than M. Moreover, $F_{n,\mathbf{0},\mathbf{0}}^{*(j)}$ gives at least mass $1/n$ to each of these points. Therefore $F_{n,\mathbf{0},\mathbf{0}}^{*(j)}\,(|y| < M) \ge (n-m-t)/n$. Since $t < k_0$, using (3.2) it follows that $(n-m-t)/n > 1 - b/a$. Thus, from Lemma 1 (b) there exists M' such that $S_n^{(j)}(\mathbf{0}, \mathbf{0}) < M'$ for all j, and (8.15) holds.

We now turn out attention to (8.14). Let $\xi_i = |\boldsymbol{\lambda}'\mathbf{x}_i|$, $1 \leq i \leq n$, where $\boldsymbol{\lambda}$ is defined in (8.13), and let

$$\xi = \min\{\xi_i \,:\, \xi_i > 0\}/\,2\,. \tag{8.17}$$

Then, for all the elements of the original sample, except at most q, we have $|\boldsymbol{\lambda}'\mathbf{x}_i| > \xi$. All the contaminated samples $\mathbf{Z}^{(j)}$ have at least $n-q-m-t$ non censored observations from the original sample $\mathbf{Z}$ such that $|\lambda'\mathbf{x}_i^{(j)}| > \xi$. Then, for j large enough, at least $n-q-m-t$ observations in $\mathbf{Z}^{(j)}$ satisfy

$$\left|y_i^{(j)} - \boldsymbol{\gamma}_j(\mathbf{0})'\mathbf{x}_i\right| \geq \left| ||\boldsymbol{\gamma}_j(\mathbf{0})|| \left| \left(\frac{\boldsymbol{\gamma}_j(\mathbf{0})}{||\boldsymbol{\gamma}_j(\mathbf{0})||}\right)' \mathbf{x}_i \right| - M \right|. \tag{8.18}$$

Fix $K > 0$ arbitrary and let K' be as in Lemma 1 (a) with C any real number satisfying

$$\frac{h_0}{n} > C > \frac{b}{a}\,, \tag{8.19}$$

where h_0 is the smallest integer larger than nb/a. Since $t < k_0$, by (3.2) we have $(n-q-m-t)/n > b/a$, and then

$$(n-q-m-t)/n > C\,. \tag{8.20}$$

Because of (8.12) and (8.13) we can always find j_0 large enough so that the right hand side of (8.18) is larger than K' for all $j > j_0$. Moreover, $F^{*(j)}_{n,\mathbf{0},\boldsymbol{\gamma}_j(\mathbf{0})}$ gives at least mass $1/n$ to those residuals $y_i^{(j)} - \boldsymbol{\gamma}_j(\mathbf{0})'\mathbf{x}_i$. Hence, by (8.20), for $j > j_0$ we have

$$F^{*(j)}_{n,\mathbf{0},\boldsymbol{\gamma}_j(\mathbf{0})}(|y| > K') \geq (n-q-m-t)/n > C\,.$$

From Lemma 1 (a) it follows that $S_n^{(j)}(\mathbf{0}, \boldsymbol{\gamma}_j(\mathbf{0})) > K$ for all $j > j_0$ and this proves (8.14).

We now prove (8.10). Assume that it is not true. Then we would have

$$\sup_j ||\boldsymbol{\gamma}_j\left(\boldsymbol{\beta}_n^{(j)}\right)|| = L < \infty\,. \tag{8.21}$$

To show that this is not possible we will prove that

$$\lim_{j\to\infty} S_n(\boldsymbol{\beta}_n^{(j)}, \boldsymbol{\gamma}_j(\boldsymbol{\beta}_n^{(j)})) = \infty\,, \tag{8.22}$$

and

$$\sup_j S_n(\boldsymbol{\beta}_n^{(j)}, -\boldsymbol{\beta}_n^{(j)}) < \infty\,. \tag{8.23}$$

To show (8.23) let M be as in (8.16) and observe that there are at least $n-m-t$ observations in $\mathbf{Z}^{(j)}$ with $\left|y_i^{(j)*}\right| < M$. It is easy to see that $F^{*(j)}_{n,\boldsymbol{\beta}_n,-\boldsymbol{\beta}_n}$ gives mass at least $1/n$ to these observations, and the proof follows as that of (8.15) above.

We will now prove (8.22). Without loss of generality assume that

$$\lim_{j\to\infty} \frac{\boldsymbol{\beta}_n^{(j)}}{||\boldsymbol{\beta}_n^{(j)}||} = \boldsymbol{\lambda}\,. \tag{8.24}$$

Let ξ be as defined in (8.17). Then for all the elements of the original sample, except at most q, we have $|\lambda'\mathbf{x}_i| > \xi$. All the contaminated samples $\mathbf{Z}^{(j)}$ have at least $n-q-m-t$ non censored observations from the original sample $\mathbf{Z}$ with $|\boldsymbol{\lambda}'\mathbf{x}_i^{(j)}| > \xi$. Then, for j large enough, at least $n-q-m-t$ observations in $\mathbf{Z}^{(j)}$ satisfy

$$\left|y_i^{(j)} - \alpha^{(j)\prime}\mathbf{x}_i\right| \geq \left| ||\beta_n^{(j)}|| \left| \left(\frac{\alpha^{(j)}}{||\beta_n^{(j)}||}\right)' \mathbf{x}_i \right| - M \right|\,, \tag{8.25}$$

where

$$\boldsymbol{\alpha}^{(j)} = \boldsymbol{\beta}_n^{(j)} + \boldsymbol{\gamma}_j(\boldsymbol{\beta}_n^{(j)})\,.$$

From (8.9), (8.21) and (8.24) it is easy to see that

$$\lim_{j\to\infty} \frac{\boldsymbol{\alpha}^{(j)}}{||\boldsymbol{\beta}_n^{(j)}||} = \boldsymbol{\lambda}\,.$$

Observing that $F^{*(j)}_{n,\boldsymbol{\beta}_n^{(j)},\boldsymbol{\gamma}(\boldsymbol{\beta}_n^{(j)})}$ gives at least mass $1/n$ to these $n-m-q-t$ residuals of the form $y_i^{(j)} - \boldsymbol{\alpha}^{(j)\prime}\mathbf{x}_i$, and that the right hand side of (8.25) can be made arbitrarily large, the rest of the proof follows the same lines as that of (8.14). $\square$

8.3 Breakdown point of MM-estimators

The following theorem is needed to find the breakdown point of MM-estimators when the response variable can be censored in Theorem 3 below.

Theorem 8 *Let* $\mathbf{Z} = (\mathbf{z}_1, ..., \mathbf{z}_n)$ *with* $\mathbf{z}_i = (y_i^*, \mathbf{x}_i, \delta_i)$ *and* $\mathbf{x}_i \in R^p$ *be a sample from a censored linear regression model. Let* $\widehat{\beta}_{1n}$ *be any regression estimator, and let* $\widehat{F}_n^* = F^*_{\beta_{1n},n}$ *the KM estimator of the corresponding residual distribution. Let* ρ_1 *and* ρ_2 *two functions satisfying P1-P4, and such that* $\rho_2 \leq \rho_1$ *and* $a = \sup \rho_2 = \sup \rho_1$*. Define* $s_n = S(\widehat{F}_n^*)$*, where* S *is a M-scale functional based on* ρ_1 *and* $0 < b < a$*. Let* $\widehat{\beta}_{2n}$ *be another estimator satisfying*

$$E_{H_n^*}(\rho_2((u + (\widehat{\beta}_{1n} - \widehat{\beta}_{2n})'\mathbf{x})/s_n)) \leq E_{H_n^*}(\rho_2(u/s_n))\,. \tag{8.26}$$

Assume that the rank of $\{\mathbf{x}_1, ...\mathbf{x}_n\}$ *is* p*, let* $q = \max_{||\theta||=1} \#\{i : \theta'\mathbf{x}_i = 0\}$ *and* $m = \sum_{i=1}^n \delta_i$*. Then*

$$\epsilon_n^*(\widehat{\beta}_{2n}, \mathbf{Z}) \geq \min(\epsilon_n^*(\widehat{\beta}_{1n}, \mathbf{Z}), (1 - b/a) - (q+m)/n, b/a - m/n)\,. \tag{8.27}$$

Proof: Let ε_0 be the right hand side of (8.27) and assume that the theorem is not true. Then there exists a sequence of samples $\mathbf{Z}^{(j)} = (\mathbf{z}_1^{(j)}, ..., \mathbf{z}_n^{(j)}),\ 1 \leq j < \infty,\ \mathbf{z}_i^{(j)} = (y_i^{*(j)}, \mathbf{x}_i^{(j)}, \delta_i^{(j)})$ such that each $\mathbf{Z}^{(j)}$ differs from $\mathbf{Z}$ in $t < \varepsilon_0 \, n$ observations and such that

$$\lim_{j\to\infty} ||\beta_{2n}^{(j)}|| = \infty \, .$$

Since $t < \epsilon_n^*(\widehat{\beta}_{1n}, \mathbf{Z})\, n$ we have $\sup_j ||\widehat{\beta}_{1n}^{(j)}|| < \infty$. Hence, if we call $\gamma_n^{(j)} = \beta_{1n}(\mathbf{Z}^{(j)}) - \beta_{2n}(\mathbf{Z}^{(j)})$ then

$$\lim_{j\to\infty} ||\gamma_n^{(j)}|| = \infty \, . \tag{8.28}$$

Moreover, in all the samples $\mathbf{Z}^{(j)}, 1 \leq j \leq n$, there are at least $n - t - m > (1 - b/a)n$ non-censored observations from the original sample. Since $\sup_j ||\widehat{\beta}_{1n}^{(j)}|| < \infty$ we have that the residuals $r_i^*(\beta_{1n}^{(j)})$ for these $n - t - m$ observations remain bounded uniformly in j. Let $\widehat{F}_n^{*(j)}$ be $\widehat{F}_n^*$ when the sample is $\mathbf{Z}^{(j)}$. Then it is clear that $\widehat{F}_n^{*(j)}$ assigns probability at least $1/n$ to these residuals, and hence by Lemma 1 (b) we have $\sup_j S(\widehat{F}_n^{*(j)}) = S^+ < \infty$.

Without loss of generality assume that

$$\lim_{j\to\infty} \frac{\gamma_n^{(j)}}{||\gamma_n^{(j)}||} = \lambda \, . \tag{8.29}$$

Let $M = \max_{1\leq i\leq n} |y_i^*| + 1$, $\delta_i = |\lambda' \mathbf{x}_i|,\ 1 \leq i \leq n$, and $\delta = \min\{\delta_i > 0\}/2$. Note that all the contaminated samples $\mathbf{Z}^{(j)}$ have at least $n-q-m-t$ non censored observations $\mathbf{z}_i^{(j)} = (y_i^{(j)}, \mathbf{x}_i^{(j)}, \delta_i^{(j)})$ from the original sample $\mathbf{Z}$ which have $|\lambda' \mathbf{x}_i^{(j)}| > \delta$. Then, since for j large enough

$$\left| y_i^{(j)} - \gamma_n^{(j)\prime} \mathbf{x}_i \right| \geq \left| ||\gamma_n^{(j)}|| \left| \left(\frac{\gamma_n^{(j)}}{||\gamma_n^{(j)}||} \right)' \mathbf{x}_i \right| - M \right| , \tag{8.30}$$

by (8.29) and (8.28), there are at least $n - q - m - t$ observations in $\mathbf{Z}^{(j)}$ such that

$$\left| y_i^{(j)} - \gamma_n^{(j)\prime} \mathbf{x}_i \right| \to \infty \, .$$

Since $n_0 = n - q - m - t > nb/a$ we can choose $bn/n_0 < \mu < a$ and let $M = \rho_2^{-1}(\mu)$. There exists a j_0 sufficiently large such that for $j \geq j_0$ these n_0 observations satisfy

$$\left| y_i^{(j)} - \gamma_n^{(j)\prime} \mathbf{x}_i \right| / S^+ > M \, .$$

Noting that the distribution function H_n^* assigns at least mass $1/n$ to each of these n_0 observations, we can conclude that

$$E_{H_n^*}(\rho_2((u + (\widehat{\beta}_{1n} - \widehat{\beta}_{2n})'\mathbf{x})/s_n)) > \frac{n_0}{n}\rho_2(M) > \frac{n_0}{n}\mu > \frac{n_0}{n}\frac{bn}{n_0} = b \, . \tag{8.31}$$

On the other hand, by the definition of s_n we have

$$E_{H_n^*}(\rho_2(u/s_n)) \leq E_{H_n^*}(\rho_1(u/s_n)) = b\,. \tag{8.32}$$

Finally, note that (8.31) and (8.32) contradict (8.26). $\square$

Proof of Theorem 3: Follows immediately from Theorem 8 $\square$

8.4 Breakdown point of τ-estimators

The following lemmas are needed to find the breakdown point of τ-estimators when the response variable can be censored.

Lemma 2 *Let ρ_1 and ρ_2 be two functions that satisfy P1-P6. Let S be the M-scale functional based on ρ_1, and T the τ functional defined in Section 2.4 based on ρ_1 and ρ_2. Then there exist finite and positive constants c_1 and c_2 such that, for all distributions F we have*

$$c_1 S(F) \leq T(F) \leq c_2 S(F)\,.$$

Proof: Take $c_2 = \sqrt{\sup_u \rho_2(u)}$, then, clearly $T(F) \leq c_2 S(F)$ for any F. To show the other inequality take $\varepsilon = \rho_1^{-1}(b/2)$ and let

$$B = \{\ u : |u|/S(F) \geq \varepsilon\}\,.$$

Then we have $b = E_F(\rho_1(u/S(F))) \leq (1 - F(B))b/2 + F(B)a$. Thus

$$F(B) \geq \frac{b/2}{a - b/2} \geq \frac{b}{2a}\,,$$

and it follows that

$$T^2(F) \geq S(F)^2 E_F(\rho_2(u/S(F))I\{B\}) \geq S(F)^2 F(B)\rho_2(\varepsilon) \geq S(F)^2 \frac{b}{2a}\rho_2(\varepsilon)\,.$$

Hence, we can take $c_1 = \sqrt{b\rho_2(\varepsilon)/(2a)}$. $\square$

Lemma 3 *Assume that ρ_1 and ρ_2 are two functions that satisfy P1-P6 and let T be the τ-scale functional based on ρ_1 and ρ_2. Then Lemma 1 holds replacing the S scale functional by T.*

Proof: Follows immediately from Lemma 2. □

Proof of Theorem 4: Follows the same lines of Theorem 2 using Lemma 3 instead of Lemma 1. □

8.5 Consistency of the S-regression estimator

The following results are needed to prove our main result in this section (Theorem 5).

Lemma 4 *Let $H_n(u)$ with $u \in \mathbb{R}^p$ be a sequence of stochastic processes such that, for each n and each element of the underlying probability space where the processes are defined, $H_n(u)$ is a distribution function. Assume that $H_n(u) \xrightarrow[n\to\infty]{a.s.} H(u)$ for each $u \in \mathbb{R}^p$, where $H(u)$ is a distribution function on $\mathbb{R}^p$. Let $G \subset \mathbb{R}^p$ such that G can be written as a finite union of disjoint rectangles (possibly unbounded). Then $H_n(G) \xrightarrow[n\to\infty]{a.s} H(G)$.*

The proof of this lemma is straightforward.

Theorem 9 *Let H_n and H be as in Lemma 4. Let $g : \mathbb{R}^p \to \mathbb{R}$ be bounded and continuous, then $E_{H_n}[g(u)] \xrightarrow[n\to\infty]{a.s.} E_H[g(u)]$.*

Proof: Let $\varepsilon > 0$, $J > 0$ and put $M = \sup|g|$. Let G be a rectangle such that $H(G) > 1 - \varepsilon/JM$. Let G_i, $1 \leq i \leq k$ be disjoint rectangles that cover G and such that if u y v are points in G_i then $|g(u) - g(v)| < \varepsilon/J$. Let $D = \{H_n(G) \to H(G)\} \bigcap \{\cap_{i=1}^k H_n(G_i) \to H(G_i)\}$. We have $P(D) = 1$. Let $u_i \in G_i$. In the event D we have

$$\left| \int_{R^p} g dH - \sum_{i=1}^k g(u_i) H(G_i) \right| \leq \sum_{i=1}^k \int_{G_i} |g(u) - g(u_i)| dH$$
$$+ \int_{G^c} |g(u)| dH \leq \varepsilon/J \sum_{i=1}^k H(G_i) + M \int_{G^c} | dH$$
$$\leq \varepsilon/J + M\varepsilon/MJ = 2\varepsilon/J\,.$$

Similarly

$$\left| \int_{R^p} g dH_n - \sum_{i=1}^K g(u_i) H_n(G_i) \right| \leq \varepsilon/J + M H_n(G^c)\,.$$

Since $H_n(G^c) \to H(G^c) \le \varepsilon/MJ$ a.s., we have

$$\limsup_{n\to\infty} \left| \int_{R^p} g dH_n - \sum_{i=1}^{K} g(u_i) H_n(G_i) \right| \le 2\varepsilon/J \text{ a.s.}$$

And then, for $n > n_0$ we have

$$\left| \int_{R^p} g dH - \int_{R^p} g dH_n \right| \le \left| \int_{R^p} g dH - \sum_{i=1}^{K} g(u_i) H(G_i) \right| + \left| \sum_{i=1}^{K} g(u_i)(H(G_i) - H_n(G_i)) \right| + \left| \int_{R^p} g dH_n - \sum_{i=1}^{K} g(u_i) H_n(G_i) \right| .$$

Since $H(G_i) \to H_n(G_i)$ a.s.

$$\limsup_{n\to\infty} \left| \int_{R^p} g dH - \int_{R^p} g dH_n \right| \le 4\varepsilon/J \, .$$

The result follows choosing $J = 4$. □

Lemma 5 *Let H_n and H be as in Lemma 4 and let $G \subset \mathbb{R}^p$ be an open set. Then $\liminf_n H_n(G) \ge H(G)$ a.s.*

Proof: Let $\varepsilon > 0$ be arbitrary. We can find a sequence $G_k \subset G$ such that G_k is a finite union of rectangles and such that $H(G_k) > H(G) - 1/k$ and V_k with $P(V_k) = 1$ and such that $\lim_{n\to\infty} H_n(G_k) = H(G_k)$ in V_k. Let $V = \cap_{k=1}^{\infty} V_k$, then $P(V) = 1$. We will show now that in V we have $\liminf H_n(G) \ge H(G)$.

Let $\varepsilon > 0$ and let j such that $1/j < \varepsilon/2$. Then there exists n_0 such that

$$H_n(G_j) > H(G) - \varepsilon \text{ for all } n > n_0 \, ,$$

and therefore

$$H_n(G) > H(G) - \varepsilon \text{ for all } n > n_0 \, .$$

This implies

$$\liminf_{n} H_n(G) \ge H(G) - \varepsilon \, .$$

Since this hold for all $\varepsilon > 0$, we have that in the set V $\liminf H_n(G) \ge H(G)$.

Lemma 6 *Let G be a probability measure in $\mathbb{R}^p$ such that $G(\beta'\mathbf{x} \neq 0) > t$ for all $\beta \in \mathbb{R}^p$. Then given $\varepsilon > 0$ there exist a finite number of sets C_i, and numbers $\delta_i > 0$, $1 \leq i \leq k$, such that $\cup_{i=1}^k C_i = \{\lambda \in R^p : ||\lambda|| = 1\}$ and*

$$G(\inf_{\beta \in C_i} |\beta' x| > \delta_i) > t - \varepsilon\,.$$

Proof: Let $\varepsilon > 0$ be given. By a standard compacity argument, it is enough to show that for any fixed $\boldsymbol{\beta} \in \mathbb{R}^p$ there exist $\eta > 0$ and $\delta > 0$ such that

$$G\Big(\inf_{\|\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\| \leq \eta} |\tilde{\boldsymbol{\beta}}'\mathbf{x}| > \delta\Big) > t - \varepsilon\,. \tag{8.33}$$

Fix $\boldsymbol{\beta}$. By assumption we have $G(\beta'\mathbf{x} \neq 0) > t$ and hence there exists $\delta = \delta(\boldsymbol{\beta}, \varepsilon) > 0$ such that

$$G\Big(|\beta'\mathbf{x}| > 2\,\delta\Big) > t - \varepsilon/2\,.$$

Let $\eta > 0$ be a small positive constant to be chosen later. Now note that for any $\tilde{\boldsymbol{\beta}}$ such that $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\| \leq \eta$ we have:

$$\begin{aligned} |\boldsymbol{\beta}'\mathbf{x}| &\leq |\boldsymbol{\beta}'\mathbf{x} - \tilde{\boldsymbol{\beta}}'\mathbf{x}| + |\tilde{\boldsymbol{\beta}}'\mathbf{x}| \\ &\leq \eta\,\|\mathbf{x}\| + |\tilde{\boldsymbol{\beta}}'\mathbf{x}|\,. \end{aligned}$$

Hence it follows that for all $\tilde{\boldsymbol{\beta}}$ such that $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\| \leq \eta$ we have the following

$$\Big\{|\boldsymbol{\beta}'\mathbf{x}| > 2\,\delta\Big\} \subseteq \Big\{|\tilde{\boldsymbol{\beta}}'\mathbf{x}| > \delta\Big\} \bigcup \Big\{\eta\,\|\mathbf{x}\| > \delta\Big\}\,,$$

and then

$$\Big\{|\boldsymbol{\beta}'\mathbf{x}| > 2\,\delta\Big\} \subseteq \Big\{\inf_{\|\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\| \leq \eta} |\tilde{\boldsymbol{\beta}}'\mathbf{x}| > \delta\Big\} \bigcup \Big\{\eta\,\|\mathbf{x}\| > \delta\Big\}\,.$$

It follows that

$$\begin{aligned} G\Big(\inf_{\|\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\| \leq \eta} |\tilde{\boldsymbol{\beta}}'\mathbf{x}| > \delta\Big) &\geq G\Big(|\boldsymbol{\beta}'\mathbf{x}| > 2\,\delta\Big) - G\Big(\eta\,\|\mathbf{x}\| > \delta\Big) \\ &\geq t - \varepsilon/2 - G\Big(\eta\,\|\mathbf{x}\| > \delta\Big)\,. \end{aligned} \tag{8.34}$$

Now choose $\eta = \eta(\delta, \varepsilon) > 0$ small enough such that

$$G\Big(\|\mathbf{x}\| > \delta/\eta\Big) < \varepsilon/2\,. \tag{8.35}$$

We have that (8.34) and (8.35) imply

$$G\Big(\inf_{\|\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\| \leq \eta} |\tilde{\boldsymbol{\beta}}'\mathbf{x}| > \delta\Big) > t - \epsilon\,,$$

which proves (8.33) and the lemma. $\square$

Lemma 7 *Let ρ satisfy regularity conditions P1-P4. Let $H_n(u, \mathbf{x}) \to F_0(u)G_0(\mathbf{x}) = H_0$ a.s. where*

(i) F_0 is symmetric and has a unimodal density;

(ii) $G(\boldsymbol{\beta}'\mathbf{x} \neq 0) \geq t$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$.

Then for any $s > 0$ and any $b^* < ta$ there exists K such that

$$\lim_{n\to\infty} \inf_{||\boldsymbol{\beta}||>K} E_{H_n}(\rho((u - \boldsymbol{\beta}'\mathbf{x})/s)) > b^* \ \ a.s..$$

Proof: Let us choose $\varepsilon > 0$ such that $(a-\varepsilon)(t-\varepsilon) > b^*$ and let M_1 such that $\rho(M_1) > a - \varepsilon$. By Lemma 6 we can find sets $C_1, ...C_k$, constants $\delta_1 > 0, \ldots, \delta_k > 0$ and M_2 such that $\bigcup_{i=1}^k C_i = \{\boldsymbol{\lambda} \in R^p : ||\boldsymbol{\lambda}|| = 1\}$ and if we call

$$A_i = \left\{ \inf_{\boldsymbol{\beta}\in C_i} |\boldsymbol{\beta}'\mathbf{x}| > \delta_i \right\} \bigcap \left\{ |u| < M_2 \right\},$$

then

$$H(A_i) > t - \varepsilon/2\,, \quad 1 \leq i \leq k\,.$$

Let $\delta = \min_{1\leq i\leq k} \delta_i$ and take $K = (sM_1 + M_2)/\delta$. Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\beta}) = \boldsymbol{\beta}/||\boldsymbol{\beta}||$. We have

$$E_{H_n}(\rho((u - \boldsymbol{\beta}'\mathbf{x})/s) \geq E_{H_n}\left[\rho((u - \|\boldsymbol{\beta}\|\, \boldsymbol{\lambda}'\mathbf{x})/s)\, I\{A_i\}\right]\,,$$

where $\boldsymbol{\lambda} \in C_i$. But, if $\|\boldsymbol{\beta}\| > K$ and $(u, \mathbf{x}) \in A_i$ we have $|u - \|\boldsymbol{\beta}\|\, \boldsymbol{\lambda}'\mathbf{x}|/s > M_1$. Then, for $\|\boldsymbol{\beta}\| > K$ we have

$$E_{H_n}(\rho((u - \boldsymbol{\beta}'\mathbf{x})/s) \geq \rho(M_1)H_n(A_i) \geq (a-\varepsilon) \min_{1\leq i\leq k} H_n(A_i)\,.$$

Since the A_i's are open sets, by Lemma 5 we have

$$\liminf_{n\to\infty} \min_{1\leq i\leq k} H_n(A_i) \geq \min_{1\leq i\leq k} \liminf_{n\to\infty} H_n(A_i) \geq t - \varepsilon\,,$$

and therefore

$$\liminf_{n\to\infty} \inf_{||\boldsymbol{\beta}||>K} E_{H_n}(\rho((u - \boldsymbol{\beta}'\mathbf{x})/s) \geq (a-\varepsilon)(t-\varepsilon) > b^*\,.$$

This proves the Lemma. $\square$

Lemma 8 *(Yohai, 1985). Let ρ satisfy regularity conditions P1-P4. Let $H_0(u, \mathbf{x}) = F_0(u)\, G_0(\mathbf{x})$ be the joint distribution of the errors and the covariates and assume that F_0 and G_0 satisfy:*

(i) F_0 *is symmetric and has a unimodal density;*

(ii) $G(\beta'\mathbf{x} \neq 0) > 0$ *for all* $\beta \in \mathbb{R}^p$.

Let $g(\boldsymbol{\beta}) = E_{H_0}[\rho(y - \boldsymbol{\beta}'\mathbf{x})]$. *Then* $g(\boldsymbol{\beta})$ *has a unique minimum at* $\boldsymbol{\beta} = \mathbf{0}$.

Lemma 9 *Let* ρ *satisfy regularity conditions P1-P4. Let* $H_n(u, \mathbf{x}) \to F_0(u)G_0(\mathbf{x}) = H_0$ *a.s. where*

(i) F_0 *is symmetric and has a unimodal density;*

(ii) $G(\beta'\mathbf{x} \neq 0) > b/a$ *for all* $\boldsymbol{\beta} \in \mathbb{R}^p$.

Let s_0 *be defined by* $E_{F_0}(\rho(u)/s_0) = b$. *Then given* $\varepsilon > 0$ *and* K *there exist* $s_1 > s_0$ *and* $b_1 > b$ *such that*

$$\lim_{n\to\infty} \inf_{\varepsilon \leq ||\boldsymbol{\beta}|| \leq K} E_{H_n}(\rho((u - \boldsymbol{\beta}'\mathbf{x})/s_1) > b_1 \,.$$

Proof: Let

$$m(\boldsymbol{\beta}, s) = E_H(\rho((u - \boldsymbol{\beta}'\mathbf{x})/s))\,.$$

It is easy to see that the function m es continuous and $m(0, s_0) = b$. Thus, by Lemma 8 $m(\boldsymbol{\beta}, s_0) > b$. Then, using the Dominated Convergence Theorem, for any $\boldsymbol{\beta} \in D = \{\varepsilon \leq ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq K\}$ there exists $\varepsilon_{\boldsymbol{\beta}} > 0$, $s_{\boldsymbol{\beta}} > s_0$ and $b_{\boldsymbol{\beta}} > b$ such that

$$E_H(\min_{\boldsymbol{\gamma}\in S(\boldsymbol{\beta},\varepsilon_{\boldsymbol{\beta}})\cap D} \rho((u - \boldsymbol{\gamma}'\mathbf{x})/s_{\boldsymbol{\beta}})) > b_{\boldsymbol{\beta}}\,.$$

Since D is compact, we can find a finite covering $D_i = S(\boldsymbol{\beta}_i, \varepsilon_{\boldsymbol{\beta}_i}) \cap D$, $i = 1, ..., k$, such that $\cup_{i=1}^k D_i = D$. Let b_1 be such that $b < b_1 < \min_{1\leq i\leq k} b_{\boldsymbol{\beta}_i}$ and $s_1 = \min_{1\leq i\leq k} s_{\boldsymbol{\beta}_i} > s_0$. By Theorem 9 we have

$$\lim_{n\to\infty} E_{H_n}(\min_{\boldsymbol{\gamma}\in D_i} \rho((u - \boldsymbol{\gamma}'\mathbf{x})/s_1)) = E_H(\min_{\boldsymbol{\gamma}\in D_i} \rho((u - \boldsymbol{\gamma}'\mathbf{x})/s_1)) > b_1 \quad \text{a.s.}\,, \quad i = 1, ..., k\,,$$

and then

$$\lim_{n\to\infty} \min_i E_{H_n}(\min_{\boldsymbol{\gamma}\in D_i} \rho((u - \boldsymbol{\gamma}'\mathbf{x})/s_1)) > b_1 \text{ a.s.}\,. \tag{8.36}$$

Take any $\boldsymbol{\beta} \in D$, then

$$\begin{aligned} E_{H_n}(\ \rho((u - \boldsymbol{\beta}'\mathbf{x})/s_1)) &\geq E_{H_n}(\min_{\boldsymbol{\gamma}\in D_i} \rho((u - \boldsymbol{\gamma}'\mathbf{x})/s_1)) \\ &\geq \min_i E_{H_n}(\min_{\boldsymbol{\gamma}\in D_i} \rho((u - \boldsymbol{\gamma}'\mathbf{x})/s_1))\,, \end{aligned}$$

and then

$$\inf_{\boldsymbol{\beta}\in D} E_{H_n}(\ \rho((u-\boldsymbol{\beta}'\mathbf{x})/s_1)) \geq \min_i E_{H_n}(\min_{\boldsymbol{\gamma}\in D_i}\ \rho((u-\boldsymbol{\gamma}'\mathbf{x})/s_1))\,.$$

Using (8.36) we get

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\beta}\in D} E_{H_n}(\ \rho((u-\boldsymbol{\beta}'\mathbf{x})/s_1)) > b_1,$$

and then the Lemma is proved. □

Lemma 10 *Let ρ satisfy regularity conditions P1-P4. Let $H_n(u,\mathbf{x}) \to H_0(u,\mathbf{x}) = F_0(u)\,G_0(\mathbf{x})$ a.s. where F_0 and G_0 satisfy:*

(i) F_0 is symmetric and has a unimodal density;

(ii) $G(\beta'\mathbf{x} \neq 0) = t > b/a$ for all $\beta \in \mathbb{R}^p$.

Let s_0 defined by $E_{F_0}(\rho(u/s_0)) = b$, then if $s_1 > s_0$

$$\lim_{n\to\infty} E_{H_n}(\rho(u/s_1)) < b\,.$$

Proof: From P1-P4 we have that $E_{H_0}(\rho(u/s_1)) < b$, and then the result follows from Theorem 9. □

Proof of Theorem 5: Observe that $S_n(\boldsymbol{\beta}_0,\boldsymbol{\gamma})$ is the value s satisfying

$$E_{H^*_{n,\boldsymbol{\beta}_0}}(\rho((y-\boldsymbol{\gamma}'\mathbf{x})/s) = b\,.$$

We know by Theorem 1 that $H^*_{n,\boldsymbol{\beta}_0}(u,\mathbf{x}) \to H_0(u,\mathbf{x}) = F_0(u)\,G_0(\mathbf{x})$ a.s. for all u and $\mathbf{x}$. Define s_0 by

$$E_{H_0}(\rho(u/s_0)) = b\,.$$

Then using Lemma 7 with $s = s_0 + 1$ we can find K such that

$$\liminf_{n\to\infty} \inf_{||\boldsymbol{\gamma}||>K} S_n(\boldsymbol{\beta}_0,\boldsymbol{\gamma}) \geq s_0 + 1 \text{ a.s.}\,.$$

Let $\varepsilon > 0$ be arbitrary. For this ε and the K found above, by Lemma 9, we can find $s_1 > s_0$ such that

$$\liminf_{n\to\infty} \inf_{\varepsilon\leq||\boldsymbol{\gamma}||\leq K} S_n(\boldsymbol{\beta}_0,\boldsymbol{\gamma}) \geq s_1 \text{ a.s.}\,.$$

Take s_2 such that $s_0 < s_2 < \min(s_0 + 1, s_1)$. By Lemma 10 we have that

$$\lim_n S_n(\boldsymbol{\beta}_0, 0) \leq s_2 \text{ a.s.}$$

This implies that, with probability 1, there exists n_0 such that for all $n \geq n_0$ we have $||\boldsymbol{\gamma}_n(\boldsymbol{\beta}_0)|| < \varepsilon$. This proves the theorem. $\square$

8.6 Asymptotic distribution

The following lemmas are needed to prove Theorem 6.

Lemma 11 *Let $H_n(u)$ with $u \in \mathbb{R}^p$ be a sequence of stochastic processes such that, for each n and each element of the underlying probability space where the processes are defined, $H_n(u)$ is a distribution function. Assume that $H_n(u) \xrightarrow{p} H(u)$ for each $u \in \mathbb{R}^p$, where $H(u)$ is a distribution function on $\mathbb{R}^p$. Let $G \subset \mathbb{R}^p$ such that G can be written as a finite union of disjoint rectangles (possibly unbounded). Then $H_n(G) \xrightarrow{P} H(G)$.*

The proof of this lemma is straightforward.

Lemma 12 *Let H_n and H be as in Lemma 11. Let $g : \mathbb{R}^p \to \mathbb{R}$ be bounded and continuous, then $E_{H_n}[g(u)] \xrightarrow{P} E_H[g(u)]$.*

Proof: Let $\varepsilon > 0$ and put $M = \sup |g|$. Let G be a rectangle such that $H(G) > 1 - \varepsilon/(6M)$. Let G_i, $1 \leq i \leq k$, be disjoint rectangles that cover G and such that if u y v are points in G_i then $|g(u) - g(v)| < \varepsilon/6$. For each integer m and $i = 1, \ldots, k$ let

$$A_{m\,i} = \left\{ |H_m(G_i) - H(G_i)| > \frac{\epsilon}{6k\,M} \right\}.$$

For each $i = 1, \ldots, k$ there exist m_i such that for $m \geq m_i$

$$P\left(A_{m\,i}\right) < \frac{\epsilon}{2k}.$$

Let $m_0 = \max(m_1, \ldots, m_k)$, then for all $m \geq m_0$ we have

$$P\left(\bigcap_{i=1}^{k} A_{m\,i}^c\right) \geq 1 - \frac{\epsilon}{2}. \tag{8.37}$$

Take $u_i \in G_i, 1 \le i \le k$. Then

$$\left| \int_{R^p} g dH - \sum_{i=1}^{k} g(u_i) H(G_i) \right| \le \sum_{i=1}^{k} \int_{G_i} |g(u) - g(u_i)| dH + \int_{G^c} |g(u)| dH$$
$$\le \frac{\varepsilon}{6} \sum_{i=1}^{k} H(G_i) + M \int_{G^c} dH$$
$$\le \frac{\varepsilon}{6} + \frac{M\varepsilon}{6M} = \frac{2\varepsilon}{6} . \tag{8.38}$$

Similarly,

$$\left| \int_{R^p} g\, dH_n - \sum_{i=1}^{k} g(u_i) H_n(G_i) \right| \le \sum_{i=1}^{k} \int_{G_i} |g(u) - g(u_i)| dH_n + \int_{G^c} |g(u)| dH_n$$
$$\le \frac{\varepsilon}{6} \sum_{i=1}^{k} H_n(G_i) + M \int_{G^c} dH_n$$
$$\le \frac{\epsilon}{6} + M\, H_n(G^c) . \tag{8.39}$$

Since $H_n(G) \xrightarrow{P} H(G)$ it follows that there exists $n_0 \ge m_0$ such that for all $n \ge n_0$ we have

$$P\left(|H_n(G) - H(G)| \le \frac{\epsilon}{6M} \right) \ge 1 - \frac{\epsilon}{2}.$$

Moreover, note that

$$\Big\{ H_n(G) > 1 - 2\,\epsilon/(6M) \Big\} \supseteq \left\{ |H_n(G) - H(G)| \le \frac{\epsilon}{6M} \right\} ,$$

and hence

$$P\left(H_n(G) > 1 - \frac{2\,\epsilon}{6M} \right) \ge 1 - \frac{\epsilon}{2} \quad \forall\, n \ge n_0.$$

Thus, with probability at least $1 - \epsilon/2$, for $n \ge n_0$ we have

$$H_n(G^c) \le \frac{2\,\epsilon}{6M},$$

and from (8.39) we get

$$\left| \int_{R^p} g\, dH_n - \sum_{i=1}^{k} g(u_i) H_n(G_i) \right| \le \frac{3\epsilon}{6} . \tag{8.40}$$

Finally, by (8.37), for $n \ge m_0$, with probability at least $1 - (\epsilon/2)$

$$\left| \sum_{i=1}^{k} g(u_i)(H_n(G_i) - H(G_i)) \right| \le M \sum_{i=1}^{k} \Big| H_n(G_i) - H(G_i) \Big| \le \frac{\epsilon}{6}. \tag{8.41}$$

Since

$$\left|\int_{R^p} gdH - \int_{R^p} gdH_n\right| \leq \left|\int_{R^p} gdH - \sum_{i=1}^{k} g(u_i)H(G_i)\right|$$
$$+\left|\sum_{i=1}^{k} g(u_i)(H(G_i) - \ H_n(G_i))\right| + \left|\int_{R^p} gdH_n - \sum_{i=1}^{k} g(u_i)H_n(G_i)\right|,$$

from (8.38), (8.40) and (8.41), for $n \geq n_0$ with probability at least $(1-\varepsilon)$ we have

$$\left|\int_{R^p} gdH - \int_{R^p} gdH_n\right| \leq \varepsilon\,.$$

□

Lemma 13 *Let H_n and H be as in Lemma 11 and let $G \subset \mathbb{R}^p$ be an open set. Then for all $\varepsilon > 0$ we have* $\lim_{n\to\infty} P(H_n(G) \leq H(G) - \varepsilon) = 0$

Proof: Let $\varepsilon > 0$ be arbitrary and find G^* a finite union of rectangles such that $G^* \subset G$ and $H(G^*) > H(G) - \varepsilon/2$. Since

$$\{H_n(G) \leq H(G) - \varepsilon\} \subset \{H_n(G^*) \leq H(G) - \varepsilon\} \subset \{H_n(G^*) \leq H(G^*) - \varepsilon/2\}$$

the result follows by Lemma 11

Lemma 14 *Let ρ satisfy regularity conditions P1-P4. Let $H_n(u,\mathbf{x}) \xrightarrow{P} F_0(u)G_0(\mathbf{x}) = H_0$ where*

(i) F_0 is symmetric and has a unimodal density;

(ii) $G(\beta'\mathbf{x} \neq 0) > t$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$.

Assume that $t > E_{F_0}\left[\rho\left(u/\sigma\right)\right]/a$ where $a = \sup_u \rho(u)$. For all $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\lim_{n\to\infty} P\left(\inf_{||\boldsymbol{\alpha}||>\varepsilon} C_n(\boldsymbol{\beta}_n, \boldsymbol{\alpha}) < E_{F_0}\left(\rho\left(\frac{u}{\sigma}\right)\right) + \delta\right) = 0$$

where

$$C_n(\boldsymbol{\beta}, \boldsymbol{\alpha}) = E_{H^*_{n\boldsymbol{\beta}}}\left(\rho\left(\frac{u - \boldsymbol{\alpha}'\mathbf{x}}{\sigma}\right)\right)\,.$$

Proof: It will be enough to prove that there exist M and $\delta_1 > 0$ such that

$$\lim_{n\to\infty} P\left(\inf_{||\boldsymbol{\alpha}||>M} C_n(\boldsymbol{\beta}_n, \boldsymbol{\alpha}) < E_{F_0}\left(\rho\left(\frac{u}{\sigma}\right)\right) + \delta_1\right) = 0 \tag{8.42}$$

and that for all $\varepsilon > 0$ and M, there exists $\delta_2 > 0$ such that

$$\lim_{n\to\infty} P\left(\inf_{\varepsilon\leq||\boldsymbol{\alpha}||\leq M} C_n(\boldsymbol{\beta}_n, \boldsymbol{\alpha}) < E_{F_0}\left(\rho\left(\frac{u}{\sigma}\right)\right) + \delta_2\right) = 0 \tag{8.43}$$

We start proving (8.42). Since $d = E_{F_0}\left(\rho\left(\frac{u}{\sigma}\right)\right) < ta$ we can find $\varepsilon > 0$ such that $(a-\varepsilon)(t-\varepsilon) > d$ and let M_1 such that $\rho(M_1) > a-\varepsilon$. By Lemma 6 we can find sets $C_1, ...C_k$, constants $\delta_1 > 0, \ldots, \delta_k > 0$ and M_2 such that $\bigcup_{i=1}^k C_i = \{\boldsymbol{\lambda} \in R^p : ||\boldsymbol{\lambda}|| = 1\}$ and if we call

$$A_i = \left\{\inf_{\boldsymbol{\beta}\in C_i} |\boldsymbol{\beta}'\mathbf{x}| > \delta_i\right\} \bigcap \{|u| < M_2\}\ ,$$

then

$$H(A_i) > t - \varepsilon/2\,, \quad 1 \leq i \leq k\,.$$

Let $\delta = \min_{1\leq i\leq k} \delta_i$ and take $K = (\sigma M_1 + M_2)/\delta$. Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\beta}) = \boldsymbol{\beta}/||\boldsymbol{\beta}||$. We have

$$E_{H_n}(\rho((u - \boldsymbol{\beta}'\mathbf{x})/\sigma) \geq E_{H_n}\left[\rho((u - \|\boldsymbol{\beta}\|\,\boldsymbol{\lambda}'\mathbf{x})/\sigma)\, I\{A_i\}\right]\ ,$$

where $\boldsymbol{\lambda} \in C_i$. But, if $\|\boldsymbol{\beta}\| > K$ and $(u, \mathbf{x}) \in A_i$ we have $|u - \|\boldsymbol{\beta}\|\,\boldsymbol{\lambda}'\mathbf{x}|/\sigma > M_1$. Then, for $\|\boldsymbol{\beta}\| > K$ we have

$$E_{H_n}(\rho((u - \boldsymbol{\beta}'\mathbf{x})/s) \geq \rho(M_1)H_n(A_i) \geq (a-\varepsilon)\min_{1\leq i\leq k} H_n(A_i)\,. \tag{8.44}$$

Since the A_i's are open sets, by Lemma 13 for any $\eta > 0$ there exists n_i such that for $n > n_i$

$$P(H_n(A_i) > H(A_i) - \varepsilon/2) \geq 1 - \eta/k \tag{8.45}$$

Let $n_0 = \max_{1\leq i\leq k} n_i$, we have $P(\min_{1\leq i\leq k} H_n(A_i) > t - \varepsilon) \geq 1 - \eta$ for all $n > n_0$. Hence, from (8.44) and (8.45) we have

$$\lim_{n\to\infty} P\left(\inf_{\|\boldsymbol{\beta}\|>K} E_{H_n}(\rho((u - \boldsymbol{\beta}'\mathbf{x})/\sigma) \geq (a-\varepsilon)(t-\varepsilon)\right) = 1$$

Since $(a-\varepsilon)(t-\varepsilon) > d$ there exists $\delta_1 > 0$ such that

$$\lim_{n\to\infty} P\left(\inf_{\|\boldsymbol{\beta}\|>K} E_{H_n}(\rho((u - \boldsymbol{\beta}'\mathbf{x})/\sigma) \geq d + \delta_1\right) = 1$$

proving (8.42).

We now turn our attention to (8.43). Let

$$m(\beta) = E_H(\rho((u - \beta'\mathbf{x})/\sigma)) \,.$$

It is easy to see that the function m es continuous and $m(0) = d$. Thus, by Lemma 8, $m(\beta) > d$ for $\boldsymbol{\beta} \neq 0$. Then, using the Dominated Convergence Theorem, for any $\boldsymbol{\beta} \in D = \{\varepsilon \leq ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq M\}$ there exists $\varepsilon_{\boldsymbol{\beta}} > 0$, and $d_{\boldsymbol{\beta}} > d$ such that

$$E_H \left(\min_{\boldsymbol{\gamma} \in S(\boldsymbol{\beta}, \varepsilon_{\boldsymbol{\beta}}) \cap D} \rho\left((u - \boldsymbol{\gamma}'\mathbf{x})/\sigma\right) \right) > d_{\boldsymbol{\beta}}$$

where $S(\boldsymbol{\beta}, \varepsilon_{\boldsymbol{\beta}})$ denotes the sphere of radius $\varepsilon_{\boldsymbol{\beta}}$ centered at $\boldsymbol{\beta}$. Since D is compact, we can find a finite covering $D_i = S(\boldsymbol{\beta}_i, \varepsilon_{\boldsymbol{\beta}_i}) \cap D$, $i = 1, ..., k$, such that $\cup_{i=1}^k D_i = D$. Let d_1 be such that $d < d_1 < \min_{1 \leq i \leq k} \; d_{\boldsymbol{\beta}_i}$. By Lemma 12 we have

$$E_{H_n}(\min_{\boldsymbol{\gamma} \in D_i} \; \rho((u - \boldsymbol{\gamma}'\mathbf{x})/\sigma)) \xrightarrow{P} E_H(\min_{\boldsymbol{\gamma} \in D_i} \; \rho((u - \boldsymbol{\gamma}'\mathbf{x})/\sigma)) > d_1, \quad i = 1, ..., k \,,$$

and then there exists $\delta_2 > 0$ such that

$$P(\min_i E_{H_n}(\min_{\boldsymbol{\gamma} \in D_i} \; \rho((u - \boldsymbol{\gamma}'\mathbf{x})/\sigma)) > d + \delta_2) \rightarrow 1 \qquad (8.46)$$

For any $\boldsymbol{\beta} \in D_i$ we have

$$\begin{aligned} E_{H_n}(\; \rho((u - \boldsymbol{\beta}'\mathbf{x})/\sigma)) &\geq E_{H_n}(\min_{\boldsymbol{\gamma} \in D_i} \; \rho((u - \boldsymbol{\gamma}'\mathbf{x})/\sigma)) \\ &\geq \min_i E_{H_n}(\min_{\boldsymbol{\gamma} \in D_i} \; \rho((u - \boldsymbol{\gamma}'\mathbf{x})/\sigma)) \,, \end{aligned}$$

and then

$$\inf_{\boldsymbol{\beta} \in D} E_{H_n}(\; \rho((u - \boldsymbol{\beta}'\mathbf{x})/\sigma)) \geq \min_i E_{H_n}(\min_{\boldsymbol{\gamma} \in D_i} \; \rho((u - \boldsymbol{\gamma}'\mathbf{x})/\sigma)) \,.$$

Using (8.46) we get

$$P\left(\inf_{\boldsymbol{\beta} \in D} E_{H_n}(\; \rho((u - \boldsymbol{\beta}'\mathbf{x})/s_1)) > d + \delta_2 \right) \rightarrow 1$$

which proves (8.43).

Proof of Theorem 6: Let

$$\begin{aligned} C_n(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \frac{1}{n} \sum_{i=1}^n E_{F^*_{n\boldsymbol{\beta}}} \left[\rho\left(\frac{u - \boldsymbol{\alpha}'\mathbf{x}_i}{\sigma} \right) \middle| \mathbf{w}_i(\boldsymbol{\beta}) \right] = E_{H^*_{n\boldsymbol{\beta}}} \left(\rho\left(\frac{u - \boldsymbol{\alpha}'\mathbf{x}}{\sigma} \right) \right) \\ D_n(\boldsymbol{\beta}) &= \frac{1}{n\sigma} \sum_{i=1}^n E_{F^*_{n\boldsymbol{\beta}}} \left[\psi\left(\frac{u}{\sigma} \right) \mathbf{x}_i \middle| \mathbf{w}_i(\boldsymbol{\beta}) \right] = \frac{1}{\sigma} E_{H^*_{n\boldsymbol{\beta}}} \left(\psi\left(\frac{u}{\sigma} \right) \right) \\ L_n(\boldsymbol{\beta}) &= \frac{1}{n\sigma^2} \sum_{i=1}^n E_{F^*_{n\boldsymbol{\beta}}} \left[\psi'\left(\frac{u}{\sigma} \right) \mathbf{x}_i\mathbf{x}_i' \middle| \mathbf{w}_i(\boldsymbol{\beta}) \right] = \frac{1}{\sigma^2} E_{H^*_{n\boldsymbol{\beta}}} \left(\psi'\left(\frac{u}{\sigma} \right) \mathbf{x}\mathbf{x}' \right) \end{aligned}$$

By Theorem 5.1 in Ritov (1990) there exists a sequence $\boldsymbol{\beta}_n$ such that

$$n^{1/2} D_n(\boldsymbol{\beta}_n) \to_p 0 \,, \tag{8.47}$$

and $n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0) \to_D \mathcal{N}(\mathbf{0}, A_\psi^{-1}\, B_\psi\, A_\psi^{-1})$. Then we only have to prove that $n^{1/2}\gamma_n(\beta_n) \to_p 0$.

Using a second order Taylor expansion around $\alpha = \mathbf{0}$ we obtain

$$C_n(\beta_n, \alpha) = C_n(\beta_n, \mathbf{0}) + D_n'(\beta_n)\alpha + \frac{1}{2}\alpha' L_n(\beta_n)\alpha + ||\alpha||^3 K_n(\alpha) \tag{8.48}$$

where there exists ε_0 and K_0 such that

$$p \lim_{n\to\infty} \sup_{||\alpha|| \le \varepsilon_0} K_n(\alpha)| \le K_0 \tag{8.49}$$

Using Theorem 7 we have that $H^*_{n\beta_n}(u, \mathbf{x}) \to F_0(u) G_0(\mathbf{x})$ in probability for any u and $\mathbf{x}$, and therefore by Lemma 14 we have that for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\lim_{n\to\infty} P\left(\inf_{||\alpha|| > \varepsilon} C_n(\beta_n, \alpha) < E_{F_0}\left(\rho\left(\frac{u}{\sigma}\right)\right) + \delta \right) = 0 \tag{8.50}$$

On the other hand by Lemma 12

$$C_n(\boldsymbol{\beta}_n, 0) \xrightarrow{P} E_{F_0}\left(\rho\left(\frac{u}{\sigma}\right)\right) = d\,, \tag{8.51}$$

$$D_n(\boldsymbol{\beta}_n) \xrightarrow{P} E_{H_0}\left(\psi\left(u/\sigma\right)\right)\mathbf{x} = \mathbf{0} \tag{8.52}$$

and

$$L_n(\boldsymbol{\beta}_n) \xrightarrow{P} L_0\,, \tag{8.53}$$

where

$$L_0 = \frac{1}{\sigma^2} E_{F_0}\left[\psi'\left(\frac{u}{\sigma}\right)\right] E_{G_0}(\mathbf{x}\mathbf{x}') \tag{8.54}$$

The next step is to prove that $\boldsymbol{\gamma}_n(\boldsymbol{\beta}_n) \xrightarrow{P} 0$. We have

$$\{||\gamma_n(\beta_n)|| > \varepsilon\} \subset \{\inf_{||\alpha|| > \varepsilon} C_n(\beta_n, \alpha) < d + 2\delta/3\} \cup \{C_n(\beta_n, 0) > d + \delta/3\}$$

and therefore (8.50) and (8.51) imply $P\{||\gamma_n(\beta_n)|| > \varepsilon\}) \to 0$.

Finally we will prove that $n^{1/2}||\gamma_n(\beta_n)|| = o_p(1)$. Then if we denote

$$J_n = \{n^{1/2}||\gamma_n(\beta_n)|| > \varepsilon\}\,,$$

we have to prove that for any $\varepsilon > 0$ we have

$$\lim_{n\to\infty} P(J_n) = 0\,. \tag{8.55}$$

According to (8.48) we have

$$J_n \subset \left\{ \inf_{\varepsilon_0 > ||\alpha|| > \varepsilon n^{-1/2}} [D_n'(\beta_n)\alpha + \frac{1}{2}\alpha' L_n(\beta_n)\alpha + ||\alpha||^3 K_n(\alpha)] \leq 0 \right\} \bigcup \{||\gamma_n(\beta_n)|| \geq \varepsilon_0\}$$

Since $P\{||\gamma_n(\beta_n)|| > \varepsilon_0\} \to 0$, in order to prove that (8.55) it is enough to show that

$$P\left(\inf_{\varepsilon_0 > ||\alpha|| > \varepsilon n^{-1/2}} \left[\frac{\alpha' D_n(\beta_n)}{||\alpha||^2} + \frac{1}{2}\frac{\alpha'}{||\alpha||} L_n(\beta_n) \frac{\alpha}{||\alpha||} + ||\alpha|| K_n(\alpha) \right] > 0 \right) \to 1\,, \tag{8.56}$$

and since (8.52),(8.53) and (8.54) hold, it is enough to prove that for all ε

$$p \lim_{n\to\infty} \sup_{||\alpha|| > \varepsilon n^{-1/2}} \frac{\alpha' D_n(\beta_n)}{||\alpha||^2} = 0,$$

This follows from

$$\sup_{||\alpha|| > \varepsilon n^{-1/2}} \frac{|\alpha' D_n(\beta_n)|}{||\alpha||^2} \leq \frac{n^{1/2}|}{\varepsilon} ||D_n(\beta_n)||$$

and (8.47).

References

[1] Breslow, N.E. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, **2**, 437-453.

[2] Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, **66**, 429-436.

[3] Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown-point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L Hodges, Jr., eds.) 157-184. Wadsworth, Belmont, California.

[4] Dutter, R. (1975). Robust regression: different approaches to numerical solutions and algorithms, Res. Rep. no. 6, Fachgruppe für Statistik, Eidgen. Technische Hochschule, Zurich.

[5] Efron, B. (1967). The two sample problem with censored data. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, **4**, 831-853.

[6] Hampel, F.R. (1975). Beyond location parameters: robust concepts and methods. *Proceedings of the 40th Session of the ISI*, volume 46, Book 1; *Bulletin of the International Statistical Institute*, **46**, 375-382.

[7] Hössjer, O. (1992). On the optimality of S-estimators. *Statistics and Probability Letters*, **14**, 413-419.

[8] Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, **1**, 799-821.

[9] Huber, P.J. (1981). *Robust Statistics*. Wiley, New York.

[10] James, I.R. and Smith, P. J. (1984). Consistency results for linear regression with censored data. *The Annals of Statistics*, **12**, 590-600.

[11] Kalbfleisch, J.D. and Prentice, R.L. (1980). *The statistical analysis of failure time data*. Wiley : New York.

[12] Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation for incomplete observations. *Journal of the American Statistical Association*, **53**, 457-481.

[13] Krasker, W.S. (1980). Estimation in linear regression models with disparate data points. *Econometrica*, **48**, 1333-1346.

[14] Krasker, W.S. and Welsch, R.E. (1982). Efficient bounded influence regression estimation. *Journal of the American Statistical Association*, **77**, 595-604.

[15] Lai, T.L. and Ying, Z. (1991). Large sample theory of a modified Bukley-James estimator for regression analysis with censored data. *The Annals of Statistics*, **19**, 1370-1402.

[16] Lai, T.L. and Ying, Z. (1994). A missing information principle and M-estimators in regression analysis with censored and truncated data. *The Annals of Statistics*, **22**, 1222-1255.

[17] Maronna, R.A., Bustos, O.H. and Yohai, V.J. (1979). Bias and efficiency robustness of general M-estimators for regression with random carriers. In *Smoothing Techinques for Curve Estimation*, (T. Gasser and M. Rosenblatt, eds.). Springer, New York.

[18] Maronna, R.A. and Yohai, V.J. (1993). Bias-robust estimates of regression based on projections. *Annals of Statistics*, **21**, 965-990.

[19] Miller, R.G. (1976). Least squares regression with censored data. *Biometrika*, **63**, 449-464.

[20] Ritov, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics*, **18**, 303-328.

[21] Rousseeuw, P.J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, **79**, 871-880.

[22] Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust regression and outlier detection*, Wiley : New York.

[23] Rousseeuw, P.J. and Yohai, V.J. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series* (Lecture Notes in Statistics No. 26), eds. J. Franke, W. Hardle and R. D. Martin, New York: Springer-Verlag, pp. 256-272.

[24] Sellero, C.S., Manteiga, W.G. and Van Keilegom, I. (2005). Uniform representation of product-limit integrals with applications. *Scandinavian Journal of Statistics*, **32**, Number 4, 563-581.

[25] Stute, W. (1993). Consistent Estimation under random censorship when covariates are present, *Journal of Multivariate Analysis*, **45**, 89-103.

[26] Stute, W. (1996). Distributional convergence under random censorship when covariates are present. *Scandinavian Journal of Statistics*, **23**, 461-471.

[27] Stute, W. and Wang, J.L. (1993). The strong law under random censoring. *The Annals of Statistics*, **21**, 1591-1607.

[28] Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, **15**, 642-656.

[29] Yohai, V.J. (1985). High breakdown-point and high efficiency robust estimates for regression. Technical Report # 66. Department of Statistics, University of Washington, Seattle, WA.

[30] Yohai, V.J. and Zamar, R.H. (1988). High breakdown point and high efficiency estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, **83**, 406-413.

[31] Yohai, V.J. and Zamar, R.H. (1993). A minimax-bias property of the least α-quantile estimates. *The Annals of Statistics*, **21**, 1824-1842.