Consistency of the robust residual autocorrelation estimate of a transformation parameter

Alfio Marazzi[1] and Victor J. Yohai[2]

Key words. S-estimates, MM-estimates, RAC-estimate.
AMS classification. Primary 62J05; secondary 62F35.

Abstract. The linear regression models for a transformed response is considered. S- and MM-estimates depending on the transformation parameter λ are defined and asymptotic results for these estimates are obtained. Using these results, consistency of the robust residual autocorrelation estimate of λ based on S- and MM-estimates is proved in the simple regression case.

1 Introduction

We consider a random sample $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ of the random variable $(\mathbf{x}, y)$, where $\mathbf{x}$ is a vector of p explanatory variables and y is a positive response variable. They are assumed to be linked by the linear relationship

$$g(y, \lambda_0) = \mathbf{x}^{\mathrm{T}} \boldsymbol{\beta}_0 + q(\mathbf{x})u, \tag{1}$$

where $g(y, \lambda_0)$ denotes a response transformation, λ_0 is a parameter that takes values in $\Lambda \subset \mathbb{R}$, $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ is a parameter vector (the first component of $\boldsymbol{\beta}_0$ being an intercept term), and $q(\mathbf{x})$ is an unknown scale function. We assume that u is independent of $\mathbf{x}$. In Section 2, we define $\tilde{\boldsymbol{\beta}}_n(\lambda)$, $\tilde{S}_n(\lambda)$, and $\hat{\boldsymbol{\beta}}_n(\lambda)$ as the S-estimate of $\boldsymbol{\beta}$, the associated S-estimate of scale, and the MM-estimate of $\boldsymbol{\beta}$ when the responses are $g(y_i, \lambda)$ and y_i is distributed according to (1). Two results are proved: First, these estimates are asymptotically uniformly bounded (for all λ in a compact subset of $\mathbb{R}$); second, for n sufficiently large and λ sufficiently close to λ_0, $\tilde{\boldsymbol{\beta}}_n(\lambda)$ and $\hat{\boldsymbol{\beta}}_n(\lambda)$ are arbitrarily close to $\boldsymbol{\beta}_0$. In Section 3, consistency of the robust residual autocorrelation estimate of the transformation parameter defined in Marazzi and Yohai (2005) is proved, for the case $p = 2$, under conditions which are insured by the results of Section2. The following assumptions concerning the model are required:

H1. The distribution F of u_i is continuous and symmetric.
H2. For each λ, the function $g(y, \lambda)$ is continuous and strictly monotone with respect to y and $\Lambda \subset \mathbb{R}$ is compact.

[1]University of Lausanne, Institute for social and preventive medicine, Bugnon 17, CH 1005 Lausanne, Switzerland.
[2]Universidad de Buenos Aires, Departamento de Matematica, Ciudad Universitaria, Pabellon 1, 1428 Buenos Aires, Argentina.

2 S- and MM-estimates

We consider a sample $\mathbf{u} = (u_1,\ u_2, ..., u_n)$ of size n of a univariate distribution. Huber (1964) defines the *M-estimate of scale* as a solution $s_n(\mathbf{u})$ of the equation

$$\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{u_i}{s_n(\mathbf{u})}\right) = b, \tag{2}$$

where b is a given positive real number and χ_1 a given function $\mathbb{R} \to \mathbb{R}^+$. In the following, we require that χ_1 satisfies the following properties:

H3. (i) $\chi_1(0) = 0$; (ii) χ_1 is even; (iii) if $|u| < |v|$, then $\chi_1(u) \leq \chi_1(v)$; (iv) χ_1 is bounded; (v) χ_1 is continuous at 0.

We now consider a sample $Z = \{(\mathbf{x}_1, z_1), ..., (\mathbf{x}_n, z_n)\}$ of a regression model $z = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + u$. Rousseeuw and Yohai (1984) define the *S-estimate of* $\boldsymbol{\beta}$ by

$$\tilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} s_n(\mathbf{r}(\boldsymbol{\beta})), \tag{3}$$

where $\mathbf{r}(\boldsymbol{\beta})= (r_1(\boldsymbol{\beta}), ..., r_n(\boldsymbol{\beta}))$ and $r_i(\boldsymbol{\beta})=y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$. An associated *S-estimate of the error scale* is then given by

$$\tilde{S}_n = s_n(\mathbf{r}(\tilde{\boldsymbol{\beta}}_n)). \tag{4}$$

Rousseeuw and Yohai (1984) show that the asymptotic breakdown point of the S-estimates is $\varepsilon^* = \min(b/a, 1 - b/a)$, where $a = \max \chi_1$. Therefore, if $b = a/2$ we have $\varepsilon^* = 1/2$. Unfortunately, the S-estimate of $\boldsymbol{\beta}$ cannot simultaneously attain a breakdown point of 0.5 and a high efficiency under normal errors (Hossjer, 1992). In order to obtain both these properties, Yohai (1987) propose the *MM-estimate of* $\boldsymbol{\beta}$ defined by

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n}\chi_2\left(\frac{r_i(\boldsymbol{\beta})}{\tilde{S}_n}\right). \tag{5}$$

where $\tilde{S}_n$ is the error scale estimate (4) and χ_2 is a second function satisfying H3 and such that:

H4. $\chi_2 \leq \chi_1$ for all u.

The MM-estimate $\hat{\boldsymbol{\beta}}_n$ has the same breakdown point as the S-estimate $\tilde{\boldsymbol{\beta}}_n$. The most frequent examples of functions χ_1 and χ_2 are taken in the Tukey's biweight family

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

where k is a given constant. Taking $\chi_1(z) = \chi(z, 1.548)$ and $b = 0.5$ in (2), the asymptotic breakdown point of the S-estimates $\tilde{\boldsymbol{\beta}}_n$ and $\tilde{S}_n$ equal to 0.5. Taking $\chi_2(z) = \chi(z, 4.687)$ in (5), the MM-estimate $\hat{\boldsymbol{\beta}}_n$ attains an asymptotic efficiency of 0.95 under normal errors while preserving breakdown point 0.5.

We now define $\tilde{\boldsymbol{\beta}}_n(\lambda)$, $\tilde{S}_n(\lambda)$, and $\hat{\boldsymbol{\beta}}_n(\lambda)$ as the S-estimate of $\boldsymbol{\beta}$, the associated S-estimate of scale, and the MM-estimate of $\boldsymbol{\beta}$ when $z_i = g(y_i, \lambda)$ and y_i is distributed according to (1). The following additional assumptions are required:

H5. $P(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} \neq 0) > b/a$ for all $\boldsymbol{\beta} \neq \mathbf{0}$.
H6. $P(g(y, \lambda) - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} \neq 0) > b/a$ for all $\boldsymbol{\beta} \neq \mathbf{0}$ and all $\lambda \in \Lambda$.
H7. The distribution of the error u has a strictly unimodal density.

Theorem 1. Let $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)$ be a random sample of model (1). Assume H1–H6. Then, there exist K, s_0, and s_1 such that:

(i) $\overline{\lim}_{n\to\infty} \sup_{\lambda\in\Lambda} ||\tilde{\boldsymbol{\beta}}_n(\lambda)|| \leq K$ a.s.,
(ii) $\overline{\lim}_{n\to\infty} \sup_{\lambda\in\Lambda} \tilde{S}_n(\lambda) \leq s_1$ a.s.,
(iii) $\underline{\lim}_{n\to\theta} \inf_{\lambda\in\Lambda} \tilde{S}_n(\lambda) \geq s_0$ a.s.,
(iv) $\overline{\lim}_{n\to\infty} \sup_{\lambda\in\Lambda} ||\hat{\boldsymbol{\beta}}_n(\lambda)|| \leq K$ a.s..

Theorem 2. Let $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)$ be a random sample of model (1). Assume H1–H7. Then, given $\delta > 0$, there exists $\varepsilon > 0$ such that

$$\overline{\lim}_{n\to\infty} \sup_{|\lambda-\lambda_0|\leq\varepsilon} ||\tilde{\boldsymbol{\beta}}_n(\lambda) - \boldsymbol{\beta}_0|| \leq \delta \text{ a.s.,} \tag{6}$$

$$\overline{\lim}_{n\to\infty} \sup_{|\lambda-\lambda_0|\leq\varepsilon} ||\tilde{S}_n(\lambda) - s_0|| \leq \delta \text{ a.s.,} \tag{7}$$

and

$$\overline{\lim}_{n\to\infty} \sup_{|\lambda-\lambda_0|\leq\varepsilon} ||\hat{\boldsymbol{\beta}}_n(\lambda) - \boldsymbol{\beta}_0|| \leq \delta \text{ a.s..} \tag{8}$$

3 The robust residual autocorrelation estimate

In this section we consider the simple regression model

$$g(y, \lambda_0) = \beta_{01} + \beta_{02}x + q(x)u, \tag{9}$$

which is a special case of (1) for $p = 2$. Suppose that $\boldsymbol{\beta}_n = (\beta_{n1}, \beta_{n2})$ is a consistent robust estimator of the coefficients for a simple regression model with homoscedastic errors and S_n a measure of error scale. Specific examples are provided by the MM-estimate $\hat{\boldsymbol{\beta}}_n$ and the associated S-estimate of scale $\tilde{S}_n$ defined in Section 3. Let $\boldsymbol{\beta}_n(\lambda) = (\beta_{n1}(\lambda), \beta_{n2}(\lambda))$ and that $S_n(\lambda)$ be the results of applying these estimators to the transformed sample $(x_1, g(y_1, \lambda)), \ldots, (x_n, g(y_n, \lambda))$ and $\beta(\lambda)$, $S(\lambda)$ their asymptotic values. Since $(x_1, g(y_1, \lambda_0)), \ldots, (x_n, g(y_n, \lambda_0))$ is a sample of a linear model, where the coefficient vector is $\boldsymbol{\beta}_0$, we have $\boldsymbol{\beta}(\lambda_0) = \boldsymbol{\beta}_0$. We define

$$r(\lambda, \boldsymbol{\beta}, \mathbf{x}, y) = g(y, \lambda) - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$$

and denote by h the inverse of $g(y, \lambda_0)$. If the residuals $r(\lambda, \boldsymbol{\beta}(\lambda), x_i, y_i)$ are computed using the true parameter λ_0, their conditional mean is close to zero for all values of the covariates. On the other hand, when the residuals are computed using a $\lambda \neq \lambda_0$, there is a functional relationship between the residual conditional mean and the fitted values. A suitable value of λ should therefore minimize a measure of non-linearity for this relationship. One such measure is the *robust residual autocorrelation* $\rho_n^*(\lambda)$ defined as follows.

We first suppose that all the values $x_1, \ldots, x_n$ are distinct. In this case, we sort sort x_i, $i = 1, \ldots, n$ in ascending order and define $j_1, \ldots j_n$ as the corresponding permuted indices, i.e., $x_{j_1} < x_{j_2} < ... < x_{j_n}$.Then, for any $\lambda \in \Lambda$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\mathrm{T}}$ and $s > 0$ we define

$$\rho_n(\lambda, \boldsymbol{\beta}, s) = \frac{1}{n-1} \sum_{i=1}^{n-1} \psi\left(\frac{r(\lambda, \boldsymbol{\beta}, x_{j_i}, y_{j_i})}{s}\right) \psi\left(\frac{r(\lambda, \boldsymbol{\beta}, x_{j_{i+1}}, y_{j_{i+1}})}{s}\right) \tag{10}$$

and

$$\rho_n^*(\lambda) = \rho_n(\lambda, \boldsymbol{\beta}(\lambda), S_n(\lambda)). \tag{11}$$

In the case of ties among the values $x_1, \ldots, x_n$, one can modify (10) by arbitrarily permuting the tied values and computing the correlations by averaging over the permutations. Details of this modification can be found in Marazzi and Yohai (2004). The robust *residual autocorrelation estimate (RAC-estimate)* of λ_0 is defined by

$$\lambda_n = \arg\min_{\lambda \in \Lambda} \rho^*(\lambda).$$

The following Theorem 3 provides a consistency result for the RAC-estimate in the simple regression case. Unfortunately, consistency has not yet been proved for the multiple regression case. The following assumptions are required.

H8. The distribution H of x_i in model (9) is continuous.

H9. q is continuous.

H10. ψ is continuous, odd, monotone non decreasing and bounded.

H11. The estimates $\boldsymbol{\beta}_n(\lambda)$ and $S_n(\lambda)$ have the following properties:

(i) Given $\delta > 0$, there exists $\varepsilon > 0$ such that

$$\overline{\lim} \sup_{|\lambda - \lambda_0| \leq \varepsilon} ||\boldsymbol{\beta}_n(\lambda) - \boldsymbol{\beta}_0|| \leq \delta \text{ a.s.},$$

where $\|\cdot\|$ denotes the Euclidean norm.

(ii) There exists $K > 0$ such that

$$\overline{\lim}_{n\to\infty} \sup_{\lambda \in \Lambda} ||\boldsymbol{\beta}_n(\lambda)|| \leq K \text{ a.s.}.$$

(iii) There exists $s_2 > 0$ such that

$$\overline{\lim}_{n\to\infty} \sup_{\lambda \in \Lambda} s_n(\lambda) \leq s_2 \text{ a.s.}.$$

(iv) There exists $s_1 > 0$ such that

$$\underline{\lim}_{n\to\infty} \inf_{\lambda \in \Lambda} s_n(\lambda) \geq s_1, \text{ a.s.}.$$

H12. Robust Identifiability Condition. In model (9) let

$$d(\lambda,\boldsymbol{\beta},s,x)=E\left(\psi\left(\frac{r(\lambda,\boldsymbol{\beta},x,y)}{s}\right)\Bigg|\,x\right). \qquad (12)$$

Then, for any $\lambda \neq \lambda_0$, $\boldsymbol{\beta} \in \mathbb{R}^2$ and $s>0$, $P(d(\lambda,\boldsymbol{\beta},s,x)\neq 0)>0$.

Theorem 3. Let $(x_1,y_1),...,(x_n,y_n)$ be a random sample of model (9). Assume H1–H2, H8–H12. Then
(i) $\hat{\lambda}_n \rightarrow_P \lambda_0$,
(ii) $\boldsymbol{\beta}_n(\hat{\lambda}_n) \rightarrow_P \boldsymbol{\beta}_0$,
where $\rightarrow_P$ denotes convergence in probability.

Remark. Theorem 1, Section 2, gives sufficient conditions for the S- and MM-estimates to satisfy assumptions H11 (ii)–(iv) (i.e., conditions A6 (ii)–(iv) in Marazzi and Yohai (2005, Section 4). Theorem 2 gives sufficient conditions for the S- and MM-estimates to satisfy assumption H11 (i) (i.e., condition A6 (i) in Marazzi and Yohai (2005, Section 4). Marazzi and Yohai (2004) proved consistency of the RAC-estimates under stronger conditions for which no result similar to Theorem 2 was available.

4 Proofs

The proof of Theorem 1 is structured as follows. First, given $\delta>0$, $M>0$, and $\zeta\in(0.5,1)$, we define $\mathcal{Z}_{\delta,M,\zeta,n}$ as the set of all the samples $Z=\{((\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n))\}$ of size n such that $\#\{i:|y_i|\leq M,\ |\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}|\geq\delta\}/n\geq\zeta$ for all $||\boldsymbol{\beta}||=1$. Lemma 1 shows that S-estimates are uniformly bounded for $Z\in\mathcal{Z}_{\delta,M,\boldsymbol{\zeta},n}$ independently of n and Lemma 2 proves a similar result for MM-estimates. Second, given $\tau>0$, $K>0$, and $\zeta\in(0.5,1)$, we define $\mathcal{Z}^*_{\tau,K,\zeta,n}$ as the set of all the samples Z of size n, such that $\#\{i:|r_i(\boldsymbol{\beta})|>\tau\}/n>\zeta$ for all $\boldsymbol{\beta}$ such that $||\boldsymbol{\beta}||\leq K$. Lemma 3 shows that S-estimates of scale are uniformly bounded away from 0 for $Z\in\mathcal{Z}^*_{\tau,K,\zeta,n}$ independently of n. Finally, we show in Lemma 5 and Lemma 6 that, for large n, the samples $(\mathbf{x}_1,g(y_1,\lambda)),.....,(\mathbf{x}_n,g(y_n,\lambda))$ belong almost surely to some $\mathcal{Z}_{\delta,M,\zeta,n}$ and some $\mathcal{Z}^*_{\tau,K,\zeta,n}$ for all λ. To prove this result, we use the general Lemma 4 , which follows from standard compacticity arguments.

Lemma 1. Let $\tilde{\boldsymbol{\beta}}_n$ be an S-estimate of $\boldsymbol{\beta}$ and $\tilde{S}_n$ the associated S-estimate of scale as defined by (3) and (4). Assume that χ_1 satisfies H3, and let $a=\max\chi_1$. Given $\delta>0$, $M>0$, and ζ such that $0.5\leq b/a<\zeta<1$, there exist positive K and s_1 (independent of n) such that, for any sample $Z\in Z_{\delta,M,\zeta,n}$, we have:

(i) $\tilde{S}_n\leq s_1$ for all n;
(ii) $||\tilde{\boldsymbol{\beta}}_n||\leq K$ for all n.

Proof. We first prove (i). Clearly $\boldsymbol{\zeta}>1-b/a$, i.e., $b-a(1-\boldsymbol{\zeta})>0$. Therefore, by H3, there exists $\mu>0$ such that $\chi_1(\mu)<b-a(1-\boldsymbol{\zeta})$. Let $s_1=M/\mu$. Then,

if $Z \in \mathcal{Z}_{\delta,M,\boldsymbol{\zeta},n}$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\mathbf{0}}{s_1}\right)=\frac{1}{n}\sum_{|y_i|\leq M}\chi_1\left(\frac{y_i}{s_1}\right)+\frac{1}{n}\sum_{|y_i|>M}\chi_1\left(\frac{y_i}{s_1}\right)$$
$$\leq \chi_1(\mu)+(1-\boldsymbol{\zeta})a<b-a(1-\boldsymbol{\zeta})+(1-\boldsymbol{\zeta})a=b.$$

Therefore, $s_n(\mathbf{r}(\mathbf{0}))<s_1$ and hence $\tilde{S}_n<s_1$. To prove (ii), we take $R<a$ such that $\boldsymbol{\zeta}R>b$ and K such that $\chi_1((K\delta-M)/s_1)>R$. Suppose that $\boldsymbol{\beta}$ is such that $\|\boldsymbol{\beta}\|>K$ and set $\boldsymbol{\theta}=\boldsymbol{\beta}/\|\boldsymbol{\beta}\|$. Then,

$$\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\ \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}{s_1}\right)=\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\|\boldsymbol{\beta}\|\,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}}{s_1}\right)$$
$$\geq\frac{1}{n}\sum_{|y_i|\leq K,\ |\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}|\geq\delta}\chi_1\left(\frac{y_i-\|\boldsymbol{\beta}\|\,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}}{s_1}\right)$$
$$\geq\boldsymbol{\zeta}\chi_1((K\delta-M)/s_1)>\boldsymbol{\zeta}R>b.$$

Thus, $\|\boldsymbol{\beta}\|>K$ implies $s_n(\mathbf{r}(\boldsymbol{\beta}))>s_1$ and, since $\tilde{S}_n<s_1$, it follows that $\left\|\tilde{\boldsymbol{\beta}}_n\right\|\leq K$.

Lemma 2. Let $\hat{\boldsymbol{\beta}}_n$ be a MM-estimate of $\boldsymbol{\beta}$ as defined by (5). Assume that χ_1 and χ_2 satisfy H3 and H4. Given $\delta>0$, $M>0$, and ζ such that $0.5\leq b/a<\zeta<1$, there exists a positive K such that, for any sample $Z\in\mathcal{Z}_{\delta,M,\boldsymbol{\zeta},n}$, we have $\left\|\hat{\boldsymbol{\beta}}_n\right\|\leq K$ for all n.

Proof. By (5), we have

$$\frac{1}{n}\sum_{i=1}^{n}\chi_2\left(\frac{y_i-\ \mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}_n}{\tilde{S}_n}\right)\leq\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\ \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}}_n}{\tilde{S}_n}\right)=b. \tag{13}$$

Let s_1 be as in Lemma 1. We choose $R<a$ such that $\boldsymbol{\zeta}R>b$ and K such that $\chi_2((K\delta-M)/s_1)>R$. Suppose that $\boldsymbol{\beta}$ is such that $\|\boldsymbol{\beta}\|>K$ and set $\boldsymbol{\theta}=\boldsymbol{\beta}/\|\boldsymbol{\beta}\|$. Let $Z\in\mathcal{Z}_{\delta,M,\boldsymbol{\zeta},n}$. Then, since $\tilde{S}_n\leq s_1$ we have

$$\frac{1}{n}\sum_{i=1}^{n}\chi_2\left(\frac{y_i-\ \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}{\tilde{S}_n}\right)\geq\frac{1}{n}\sum_{|y_i|\leq K,\ |\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}|\geq\delta}\chi_2\left(\frac{y_i-\|\boldsymbol{\beta}\|\,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}}{\tilde{S}_n}\right)$$
$$\geq\boldsymbol{\zeta}\chi_2\left(\frac{K\delta-M}{\tilde{S}_n}\right)>\boldsymbol{\zeta}R>b.$$

Therefore, by (13), we obtain $\left\|\hat{\boldsymbol{\beta}}_n\right\|\leq K$.

Lemma 3. Assume that χ_1 satisfies H3, and let $a=\max\chi_1$. Given $\tau>0$, $K>0$, and ζ such $b/a<\zeta<1$, there exists $s_0>0$ such that, for any sample $Z\in\mathcal{Z}^{*}_{\tau,K,\zeta,n}$, we have $s_n(\mathbf{r}(\boldsymbol{\beta}))>s_0$ for all n and all $\boldsymbol{\beta}$ such that $\|\boldsymbol{\beta}\|\leq K$.

Proof. Take $R < a$ such that $R\zeta > b$ and $s_0 > 0$ such that $\chi_1(\tau/s_0) > R$. Then, for all $\boldsymbol{\beta}$ such that $\|\boldsymbol{\beta}\| \leq K$ and $Z \in \mathcal{Z}^*_{\tau,K,\zeta,n}$ we have

$$\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}{s_0}\right) \geq \frac{1}{n}\sum_{|r_i(\boldsymbol{\beta})|>\tau}^{n}\chi_1\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}{s_0}\right)$$
$$> \zeta\chi_1\left(\frac{\tau}{s_0}\right) > \zeta R > b,$$

and therefore $s_n(\mathbf{r}(\boldsymbol{\beta})) > s_0$ for all n.

Lemma 4. Suppose that $\mathbf{x}_1, ..., \mathbf{x}_n$ are i.i.d. random vectors of dimension p, Θ is a compact set in an Euclidean space and $t : \mathbb{R}^p \times \Theta \to \mathbb{R}$ is a continuous function. Suppose that there exists c such that $P(t(\mathbf{x}, \boldsymbol{\theta}) \neq 0) > c$ for all $\boldsymbol{\theta} \in \Theta$. Then there exist $\zeta > c$ and $\delta > 0$ such that

$$\underline{\lim}_{n\to\infty} \inf_{\boldsymbol{\theta}\in\Theta} \frac{\#\{i : (|t(\mathbf{x}_i, \boldsymbol{\theta})| > \delta\}}{n} > \zeta \text{ a.s..}$$

Proof. By the Dominated Convergence Theorem, given $\boldsymbol{\theta} \in \Theta$, we can find $\delta(\boldsymbol{\theta}) > 0, \zeta(\boldsymbol{\theta}) > c$ and $\varepsilon(\boldsymbol{\theta})$ such that

$$P\left(\inf_{\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| \leq \varepsilon(\boldsymbol{\theta})} |t(\mathbf{x}, \boldsymbol{\theta}^*)| > \delta(\boldsymbol{\theta})\right) > \zeta(\boldsymbol{\theta}). \tag{14}$$

By the Heine-Borel Theorem there exist $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_k$, such that

$$\Theta \subset \cup_{j=1}^{k}\left\{\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_j\| \leq \varepsilon(\boldsymbol{\theta}_j)\right\}.$$

Let $\delta = \min_{1\leq j\leq k}\delta(\boldsymbol{\theta}_j)$, then

$$\inf_{\boldsymbol{\theta}\in\Theta}\frac{\#\{i : |t(\mathbf{x}_i, \boldsymbol{\theta})| > \delta\}}{n} \geq \inf_{1\leq j\leq k}\frac{\#\left\{i : \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_j\|\leq\varepsilon(\boldsymbol{\theta}_j)} |t(\mathbf{x}_i, \boldsymbol{\theta})| > \delta(\boldsymbol{\theta}_j)\right\}}{n}.$$

Using (14) and the Strong Law of Large Numbers, we have that for $1 \leq j \leq k$

$$\underline{\lim}_{n\to\infty}\frac{\#\left\{i : \inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_j\|\leq\varepsilon(\boldsymbol{\theta}_j)} |t(\mathbf{x}_i, \boldsymbol{\theta})| > \delta(\boldsymbol{\theta}_j)\right\}}{n} > \zeta(\boldsymbol{\theta}_j) \text{ a.s..}$$

Then, putting $\zeta = \min_{1\leq j\leq k}\zeta(\boldsymbol{\theta}_j) > c$, the Lemma follows.

Lemma 5. Let $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)$ be a random sample of model (1). Assume that g is continuous, Λ is compact and that there exists c such that $P(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} \neq 0) > c$ for all $\boldsymbol{\beta} \neq \mathbf{0}$. Then, there exist $\zeta > c$, $\delta > 0$ and $M > 0$ such that

$$P\left(\cup_{n=1}^{\infty}\cap_{m>n}\cap_{\lambda\in\Lambda}\{((\mathbf{x}_1, g(y_1, \lambda)), ...(\mathbf{x}_m, g(y_m, \lambda))) \in \mathcal{Z}_{\delta,M,\zeta,m}\}\right) = 1. \tag{15}$$

Proof. By Lemma 4, there exists $\delta > 0$ and $\zeta_1 > c$ such that

$$\underline{\lim}_{n\to\infty} \inf_{\|\boldsymbol{\beta}\|=1} \frac{\#\left\{i : |\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}| > \delta\right\}}{n} > \zeta_1 \text{ a.s..} \tag{16}$$

Take ζ such $c < \zeta < \zeta_1$ and put $y_i^* = \max_{\lambda\in\Lambda} |g(y_i, \lambda)|$. Since the y_i^* are i.i.d. random variables, we can find M such that $P(y_i^* \leq M) > 1 - (\zeta_1 - \zeta)/2$. Therefore, by the Strong Law of Large Numbers

$$\lim_{n\to\infty} \frac{\#\{i : y_i^* \leq M\}}{n} > 1 - \frac{\zeta_1 - \zeta}{2} \text{ a.s..} \tag{17}$$

From (16) and (17) we get

$$\underline{\lim}_{n\to\infty} \inf_{\|\boldsymbol{\beta}\|=1} \frac{\#\left\{i : |\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}| > \delta \text{ and } y_i^* \leq M\right\}}{n} > \zeta \text{ a.s.}$$

and therefore the Lemma follows.

Lemma 6. Let $(\mathbf{x}_1, y_1), \ldots.., (\mathbf{x}_n, y_n)$ be a random sample of model (1). Assume that g is continuous, Λ is compact, and that there exist d such that $P\left(g(y,\lambda) - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} \neq 0\right) > d$ for all λ. Then, for any $K > 0$, there exist $\zeta > d$ and $\tau > 0$ such that

$$P\left(\cup_{n=1}^{\infty} \cap_{m>n} \cap_{\lambda\in\Lambda} \left\{((\mathbf{x}_1, g(y_1,\lambda)), \ldots..(\mathbf{x}_m, g(y_m,\lambda))) \in \mathcal{Z}^*_{\tau,K,\zeta,m}\right\}\right) = 1. \tag{18}$$

Proof. Using Lemma 4, we can find $\zeta > d$ and $\tau > 0$, such that

$$\underline{\lim}_{n\to\infty} \inf_{(\boldsymbol{\beta},\boldsymbol{\lambda})\in\{\|\boldsymbol{\beta}\|\leq K\}\times\Lambda} \frac{\#\left\{i : |g(y_i,\lambda) - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}| > \tau\right\}}{n} > \zeta \text{ a.s.}$$

and then the Lemma follows.

Proof of Theorem 2. Using H5 and Lemma 5 with $c = b/a$, we can find M and $\zeta > b/a$ such that (15) holds. By Lemma 1, we can find K and s_1 such that, for any sample $Z \in \mathcal{Z}_{\delta,M,\zeta,m}$, we have $\tilde{S}_n \leq s_1$, $||\tilde{\boldsymbol{\beta}}_n|| \leq K$ and $\left\|\hat{\boldsymbol{\beta}}_n\right\| \leq K$ for all n. Therefore,

$$\begin{aligned} P\left(\cup_{n=1}^{\infty} \cap_{m>n} \cap_{\lambda\in\Lambda}\{||\tilde{\boldsymbol{\beta}}_n(\lambda)|| \leq K\}\right) &= 1, \\ P\left(\cup_{n=1}^{\infty} \cap_{m>n} \cap_{\lambda\in\Lambda} \left\{\tilde{S}_n(\lambda) \leq s_1\right\}\right) &= 1, \\ P\left(\cup_{n=1}^{\infty} \cap_{m>n} \cap_{\lambda\in\Lambda} \left\{||\hat{\boldsymbol{\beta}}_n(\lambda)|| \leq K\right\}\right) &= 1. \end{aligned} \tag{19}$$

Then, parts (i), (ii), and (iv) of Theorem 2 follow. Using H6 and Lemma 6 with $d = b/a$, we can find $\tau > 0$ and $\zeta > b/a$ such that (18) holds. By Lemma 3, we can find s_0 such that $Z \in \mathcal{Z}^*_{\tau,K,\zeta,n}$ implies $s_n(\mathbf{r}(\boldsymbol{\beta})) > s_0$ for all $||\boldsymbol{\beta}|| \leq K$. Then, from (18) and 19 we get

$$P\left(\cup_{n=1}^{\infty} \cap_{m>n} \cap_{\lambda\in\Lambda} \left\{\tilde{S}_n(\lambda) \geq s_0\right\}\right) = 1,$$

and part (iii) of Theorem 2 follows.

The following Lemma is necessary to prove Theorem 2.

Lemma 7. Suppose that $y = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + u$, where $\mathbf{x}$ is a random vector of dimension p and u a random variable independent of $\mathbf{x}$ with a distribution satisfying H1 and H7. Let χ be a function satisfying H3. Then:
(a) $E(\chi(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}))$ has a unique minimum at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$;
(b) if $s(\boldsymbol{\beta})$ is the solution with respect to s of

$$E\left(\chi\left(\frac{y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}}{s}\right)\right) = b,$$

then $s(\boldsymbol{\beta})$ has a unique minimum at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

Proof. The proof of this Lemma can be found in Yohai and Zamar (1988).

Proof of Theorem 2. Let $s_n(\lambda, \boldsymbol{\beta})$ denote the M-scale defined by

$$\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{r\left(\lambda, \boldsymbol{\beta}, \mathbf{x}_i, y_i\right)}{s_n(\lambda, \boldsymbol{\beta})}\right) = b,$$

and let $S(\boldsymbol{\beta}, \lambda)$ denote the asymptotic version of $s_n(\lambda, \boldsymbol{\beta})$ defined by

$$E\left(\chi_1\left(\frac{r(\lambda, \boldsymbol{\beta}, \mathbf{x}, y)}{S(\lambda, \boldsymbol{\beta})}\right)\right) = b.$$

By Lemma 7,

$$\sigma_0 = S(\boldsymbol{\beta}_0, \lambda_0) < S(\boldsymbol{\beta}, \lambda_0)$$

for all $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ and, therefore,

$$E\left(\chi_1\left(\frac{r(\lambda_0, \boldsymbol{\beta}, \mathbf{x}, y)}{\sigma_0}\right)\right) > b.$$

for all $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. Let K be as in Theorem 1 and

$$C = \{\boldsymbol{\beta} : \delta \leq \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq K + \|\boldsymbol{\beta}_0\|\}.$$

By the Dominated Convergence Theorem, for any $\boldsymbol{\beta} \in C$ there exists $\eta(\boldsymbol{\beta}) \leq \boldsymbol{\delta}$ such that

$$E\left(\inf_{(\boldsymbol{\beta}^*, \lambda) \in D_{\boldsymbol{\beta}}} \chi_1\left(\frac{r(\lambda, \boldsymbol{\beta}, \mathbf{x}, y)}{s_0 + \eta(\boldsymbol{\beta})}\right)\right) > b,$$

where

$$D_{\boldsymbol{\beta}} = \{(\boldsymbol{\beta}^*, \lambda) : \|\boldsymbol{\beta}^* - \boldsymbol{\beta}\| \leq \eta(\boldsymbol{\beta}),\ |\lambda - \lambda_0| \leq \eta(\boldsymbol{\beta})\}.$$

Using the Heine Borel Theorem, we can find vectors $\boldsymbol{\beta}_i \in C$, $1 \leq i \leq k$ and sets

$$C_i = \{\boldsymbol{\beta}^* : \|\boldsymbol{\beta}^* - \boldsymbol{\beta}_i\| \leq \eta(\boldsymbol{\beta}_i)\},$$

such that $\bigcup_{i=1}^{k} C_i \supset C$. Then, putting $\eta_0 = \min_{1\leq i\leq k} \eta(\boldsymbol{\beta}_i) \leq \boldsymbol{\delta}$ and

$$D_i = \{(\boldsymbol{\beta},\lambda) : \|\boldsymbol{\beta}-\boldsymbol{\beta}_i\| \leq \eta_0,\ |\lambda-\lambda_0| \leq \eta_0\}$$

we have

$$E\left(\inf_{(\boldsymbol{\beta},\lambda)\in D_{\mathrm{i}}} \chi_1\left(\frac{r(\lambda,\boldsymbol{\beta},\mathbf{x},y)}{\sigma_0+\eta_0}\right)\right) > b. \tag{20}$$

By the definition of σ_0 we have

$$E\left(\chi_1\left(\frac{r(\lambda_0,\boldsymbol{\beta}_0,\mathbf{x},y)}{\sigma_0}\right)\right) = b.$$

Then, be the Dominated Convergence Theorem, we can find $0<\varepsilon<\eta_0$ such that

$$E\left(\sup_{|\lambda-\lambda_0|\leq\varepsilon} \chi_1\left(\frac{r(\lambda,\boldsymbol{\beta}_0,\mathbf{x},y)}{\sigma_0+\eta_0/2}\right)\right) < b. \tag{21}$$

By (20), (21), $\varepsilon<\eta_0$ and the Strong Law of Large Numbers we obtain

$$\begin{aligned}
&\underline{\lim}_{n\to\infty} \inf_{\boldsymbol{\beta}\in C,\ |\lambda-\lambda_0|\leq\varepsilon} \frac{1}{n}\sum_{i=1}^{n} \chi_1\left(\frac{r(\lambda,\boldsymbol{\beta},\mathbf{x}_i,y_i)}{\sigma_0+\eta_0}\right) \\
&\geq \min_{1\leq i\leq k} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \inf_{(\boldsymbol{\beta},\lambda)\in D_{\mathrm{i}}} \chi_1\left(\frac{r(\lambda,\boldsymbol{\beta},\mathbf{x}_i,y_i)}{\sigma_0+\eta_0}\right) > b \text{ a.s.}
\end{aligned} \tag{22}$$

and

$$\begin{aligned}
&\overline{\lim}_{n\to\infty} \sup_{|\lambda-\lambda_0|\leq\varepsilon} \frac{1}{n}\sum_{i=1}^{n} \chi_1\left(\frac{r(\lambda,\boldsymbol{\beta}_0,\mathbf{x}_i,y_i)}{\sigma_0+\eta_0/2}\right) \\
&\leq \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \sup_{|\lambda-\lambda_0|\leq\varepsilon} \chi_1\left(\frac{r(\lambda,\boldsymbol{\beta}_0,\mathbf{x}_i,y_i)}{\sigma_0+\eta_0/2}\right) < b \text{ a.s..}
\end{aligned} \tag{23}$$

Therefore, for $|\lambda-\lambda_0|\leq\varepsilon$ and n sufficiently large, we have

$$s_n(\lambda,\boldsymbol{\beta}) > \sigma_0+\eta_0 \text{ for any } \boldsymbol{\beta}\in C, \text{ and } s_n(\lambda,\boldsymbol{\beta}_0) < \sigma_0+\frac{\eta_0}{2}.$$

This implies that, with probability one, there exists n_0 such that

$$\left\|\tilde{\boldsymbol{\beta}}_n(\lambda)-\boldsymbol{\beta}_0\right\| \leq \delta \text{ or } \left\|\tilde{\boldsymbol{\beta}}_n(\lambda)-\boldsymbol{\beta}_0\right\| \geq \|\boldsymbol{\beta}_0\|+K$$

for $n\geq n_0$ and all λ such that $|\lambda-\lambda_0|\leq\varepsilon$. The last inequality implies $\left\|\tilde{\boldsymbol{\beta}}_n(\lambda)\right\| > K$. However, according to Theorem 1, (i), with probability one there exists n_1 such that for all $n\geq n_1$, $\sup_{\lambda\in\Lambda} |\tilde{\boldsymbol{\beta}}_n(\lambda)|| \leq K$ for all $\lambda\in\Lambda$. Thus, for $n\geq\max(n_0,n_1)$, $\left\|\tilde{\boldsymbol{\beta}}_n(\lambda)-\boldsymbol{\beta}_0\right\| \leq \delta$ for all λ such that $|\lambda-\lambda_0|\leq\varepsilon$. This proves (6). Since $\eta_0\leq\delta$, (23) implies (7). The proof of (8) is similar to the proof of (6) and it is omitted.

To prove Theorem 3 we need Lemmas 8–11. The proofs of Lemmas 8–10 can be found in Marazzi and Yohai (2004).

Lemma 8.
(i) Assume that ψ, g and q are continuous. Then, $d(\lambda, \boldsymbol{\beta}, s, x)$ defined in (12) is continuous.
(ii) Assume that ψ is even and u is symmetrically distributed. Let $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02})^{\mathrm{T}}$. Then, $d(\lambda_0, \boldsymbol{\beta}_0, s, x) = 0$ for all s.

Suppose that H2 holds and let $h(z)$ be the inverse of $g(y, \lambda_0)$. Then, for any $\lambda \in \Lambda$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\mathrm{T}}$, x and u we define

$$\tilde{r}(\lambda, \boldsymbol{\beta}, x, u) = r\left(\lambda, \boldsymbol{\beta}, x, h(\beta_{01} + \beta_{02} x + q(x) u)\right).$$

In addition, for given $\lambda \in \Lambda$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\mathrm{T}}$, s, u, u^*, x, and $\varepsilon > 0$ we define

$$t(\lambda, \boldsymbol{\beta}, s, u, u^*, x, \varepsilon) = \inf_{D(\lambda, \boldsymbol{\beta}, s, \varepsilon, x)} \psi\left(\frac{\tilde{r}(\lambda^*, \boldsymbol{\beta}^*, x, u)}{s^*}\right) \psi\left(\frac{\tilde{r}(\lambda^*, \boldsymbol{\beta}^*, x^*, u^*)}{s^*}\right),$$

where $D(\lambda, \boldsymbol{\beta}, s, \varepsilon, x)$ is the set of all λ^*, $\boldsymbol{\beta}^* = (\beta_{01}^*, \beta_{02}^*)^{\mathrm{T}}$, s^*, x^* satisfying the following conditions:

$$|\lambda^* - \lambda| \leq \varepsilon,\ ||\boldsymbol{\beta}^* - \boldsymbol{\beta}\,|| \leq \varepsilon,\ |s^* - s\,| \leq \varepsilon,\ |x^* - x| \leq \varepsilon.$$

Lemma 9. Assume H1–H2, H9–H12. Let u, u^*, and x be independent random variables such that u and u^* have distribution F and x has distribution H. Then, for given $\lambda \neq \lambda_0$, $\boldsymbol{\beta} = (\beta_1, \beta_1)^{\mathrm{T}}$, and $s > 0$, there exists $\varepsilon^* = \varepsilon^*(\lambda, \beta, s)$ such that $E(t(\lambda, \boldsymbol{\beta}, s, u, u^*, x, \varepsilon^*)) > 0$.

Lemma 10. Let $x_1, \ldots, x_n$ be a random sample of a continuous distribution and let $j_1, ..., j_n$ be such that $x_{j_1} \leq \ldots \leq x_{j_n}$. Define $d_i = x_{j_{i+1}} - x_{j_i}$, $i = 1, \ldots, n-1$ and $m_n(\varepsilon) = \#\{i : d_i > \varepsilon\}/n$. Then, for any $\varepsilon > 0$, $\lim_{n \to \infty} m_n(\varepsilon) = 0$ a.s..

Lemma 11. Let $(x_1, y_1), ..., (x_n, y_n)$ be a random sample of model (9). Assume H1–H2, H8–H12 and let $\rho_n(\lambda, \boldsymbol{\beta}, s)$ be as defined in (11). Then:
(a) Given positive numbers δ, K, s_0, and s_1, there exists $z > 0$ such that

$$\lim_{n \to \infty} P(\inf_{(\lambda, \boldsymbol{\beta}, s) \in L} \rho_n(\lambda, \boldsymbol{\beta}, s) \geq z) = 1,$$

where

$$L = \{(\lambda, \boldsymbol{\beta}, s) : \lambda \in \Lambda,\ \ |\lambda - \lambda_0| \geq \delta,\ ||\boldsymbol{\beta}|| \leq K,\ s_0 \leq s \leq s_1\}. \tag{24}$$

(b) Given positive numbers s_0, s_1, and η, there exists $\nu > 0$ such that

$$\lim_{n \to \infty} P(\sup_{(\boldsymbol{\beta}, s) \in L^*} |\rho_n(\lambda_0, \boldsymbol{\beta}, s)| \leq \eta) = 1,$$

where

$$L^* = \{(\boldsymbol{\beta}, s) : ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq \nu,\ s_0 \leq s \leq s_1\}. \tag{25}$$

Proof. We prove part (a). Since the distribution of x is continuous we can suppose that all the x_i's are different. For any $(\lambda, \boldsymbol{\beta}, s) \in L$, let $\varepsilon^*(\lambda, \boldsymbol{\beta}, s)$ be defined as in Lemma 9. Since L is compact, according to the Heine-Borel Theorem, we can find $(\lambda_k, \boldsymbol{\beta}_k, s_k) \in L$, $1 \leq k \leq m$, such that $L \subset \cup_{k=1}^m L_k$, where

$$L_k = \{(\lambda, \boldsymbol{\beta}, s) : |\lambda - \lambda_k| \leq \varepsilon_k^*, \ ||\boldsymbol{\beta} - \boldsymbol{\beta}_k|| \leq \varepsilon_k^* \ \ |s - s_k| \leq \varepsilon_k^*\}$$

and $\varepsilon_k^* = \varepsilon^*(\lambda_k, \boldsymbol{\beta}_k, s_k)$. It is therefore enough to show that there exist numbers $z_k > 0$, $1 \leq k \leq m$, such that

$$\lim_{n\to\infty} P\left(\inf_{(\lambda,\boldsymbol{\beta},s)\in L_k} \rho_n(\lambda, \boldsymbol{\beta}, s) > z_k\right) = 1. \tag{26}$$

Let $j_1, ..., j_n$ be defined as in Lemma 10, $I = \{i : | \, x_{j_{i+1}} - x_{j_i}| > \varepsilon_k^*\}$, and $m_n = \#I/n$. Then, if we call $l_1, \ldots, l_n$ the inverse permutation of $j_1, \ldots, j_n$ and $a = \sup \psi$, we have

$$\begin{aligned}
&\inf_{(\lambda,\boldsymbol{\beta},s)\in L_k} \rho_n(\lambda, \boldsymbol{\beta}, s) \\
&> \frac{1}{n} \sum_{i \notin I} t(\lambda_k, \boldsymbol{\beta}_k, s_k, u_{j_i}, u_{j_{i+1}}, x_{j_i}, \varepsilon_k^*) + \\
&\inf_{(\lambda,\boldsymbol{\beta},s)\in L_k} \frac{1}{n} \sum_{i \in I} \psi\left(\frac{\tilde{r}\left(\lambda, \boldsymbol{\beta}, x_{j_i}, u_{j_i}\right)}{s}\right) \psi\left(\frac{\tilde{r}\left(\lambda, \boldsymbol{\beta}, x_{j_{i+1}}, u_{\ j_{i+1}}\right)}{s}\right) \\
&\geq \frac{1}{n} \sum_{i \notin I} t(\lambda_k, \boldsymbol{\beta}_k, s_k, u_{j_i}, u_{j_{i+1}}, x_{j_i}, \varepsilon_k^*) - m_n a^2 \\
&\geq \frac{1}{n} \sum_{i=1}^{n-1} t(\lambda_k, \boldsymbol{\beta}_k, s_k, u_{j_i}, u_{j_{i+1}}, x_{j_i}, \varepsilon_k^*) - 2m_n a^2 \\
&= \frac{1}{n} \sum_{i=1}^{n-1} t(\lambda_k, \boldsymbol{\beta}_k, s_k, u_i, u_{l_{i+1}}, x_i, \varepsilon_k^*) - 2m_n a^2.
\end{aligned}$$

Using Lemma 10, we have $m_n \to 0$ a.s., and therefore

$$\begin{aligned}
&\underline{\lim}_{\ n\to\infty} \inf_{(\lambda,\boldsymbol{\beta},s)\in L_k} \rho_n(\lambda, \boldsymbol{\beta}, s, \boldsymbol{\beta}_0) \\
\geq \quad &\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n-1} t(\lambda_k, \boldsymbol{\beta}_k, s_k, u_i, u_{l_{i+1}}, x_i, \varepsilon_k^*) \text{ a.s..}
\end{aligned} \tag{27}$$

Since the permutation $j_1, \ldots j_n$ depends only on the x_i's but not on the u_i's, we have

$$E\left(t\left(\lambda_k, \boldsymbol{\beta}_k, s_k, u_i, u_{l_{i+1}}, x_i, \varepsilon_k^*\right)\right) = E\left(t\left(\lambda_k, \boldsymbol{\beta}_k, s_k, u, u^*, x, \varepsilon_k^*\right)\right),$$

where u, u^* and x are independent random variables, the first two with distribution F and the third with distribution H. Therefore, by Lemma 9, there exists

$z_k > 0$ such that

$$E\left(\frac{1}{n}\sum_{i=1}^{n-1} t\left(\lambda_k, \boldsymbol{\beta}_k, s_k, u_i, u_{l_{i+1}}, x_i, \varepsilon_k^*\right)\right) = z_k > 0. \tag{28}$$

Since $t(\lambda_k, \boldsymbol{\beta}_k, s_k, u_i, u_{l_{i+1}}, x_i, \varepsilon_k^*)$ and $t(\lambda_k, \boldsymbol{\beta}_k, s_k, u_j, u_{l_{j+1}}, x_j, \varepsilon_k^*)$ with $i \neq j$ are independent, except when $i = l_{j+1}$ or $j = l_{i+1}$, we also have

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n-1} t\left(\lambda_k, \boldsymbol{\beta}_k, s_k, u_i, u_{l_{i+1}}, x_i, \varepsilon_k^*\right)\right) \\
&= \frac{(n-1)}{n^2}\left[\operatorname{Var}\left(t(\lambda_k, \boldsymbol{\beta}_k, s_k, u, u^*, x, \varepsilon_k^*)\right)\right. \\
&\left. + \ 2\operatorname{Cov}(t\left(\lambda_k, \boldsymbol{\beta}_k, s_k, u, u^*, x, \varepsilon_k^*\right), t(\lambda_k, \boldsymbol{\beta}_k, s_k, u^*, u^{**}, x^*, \varepsilon_k^*))\right],
\end{aligned}$$

where u, u^*, u^{**}, x, x^* are independent and thus

$$\lim_{n\to\infty} \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n-1} t\left(\lambda_k, \boldsymbol{\beta}_k, s_k, u_i, u_{l_{i+1}}, x_i, \varepsilon_k^*\right)\right) = 0. \tag{29}$$

Therefore, using Chebychev's inequality, (28), and (29) we obtain

$$\frac{1}{n}\sum_{i=1}^{n-1} t(\lambda_k, \boldsymbol{\beta}_k, s_k, u_i, u_{l_{i+1}}, x_i, \varepsilon_k^*)) \to_P z_k > 0. \tag{30}$$

Finally, (26) follows from (27) and (30). The proof of part (b) is similar to the proof of part (a). The main difference is that we now use Lemma 8 (ii) instead of Lemma 9.

Proof of Theorem 3. By assumptions H11 (ii), (iii) and (iv) there exist K, s_0, s_1 and n_1 such that, if

$$\begin{aligned}
A_{1n} &= \left\{\sup_{\lambda\in\Lambda} ||\boldsymbol{\beta}_n(\lambda)|| \leq K\right\}, \\
A_{2n} &= \left\{\inf_{\lambda\in\Lambda} S_n(\lambda) \geq s_0\right\}, \\
A_{3n} &= \left\{\sup_{\lambda\in\Lambda} S_n(\lambda) \leq s_1\right\},
\end{aligned}$$

we have $P(A_{in}) \geq 1 - \varepsilon/6$, for $n > n_1$ and $1 \leq i \leq 3$. Let z be as in Lemma 11 (a), then there exists n_2 such that, if L is defined as in (24) and

$$A_{4n} = \left\{\inf_{(\lambda,\boldsymbol{\beta},s)\in L} \rho_n(\lambda, \boldsymbol{\beta}, s) > z\right\},$$

we have $P(A_{4n}) \geq 1 - \varepsilon/6$ for $n > n_2$. By Lemma 11, (b), there exist ν and n_3 such that, if L^* is defined as in (25) and

$$A_{5n} = \left\{\sup_{(\boldsymbol{\beta},s)\in L^*} |\rho_n(\lambda_0, \boldsymbol{\beta}, s)| \leq \frac{z}{2}\right\},$$

we have $P(A_{5n}) \geq 1 - \varepsilon/6$ for $n > n_3$. Finally, by H10 (i) there exists n_4 such that, if

$$A_{6n} = \{\|\boldsymbol{\beta}_n(\lambda_0) - \boldsymbol{\beta}_0\| \leq v\},$$

we have $P(A_{6n}) \geq 1 - \varepsilon/6$ for $n > n_4$. Since $\rho_n^*(\lambda) = \rho_n(\lambda, \boldsymbol{\beta}_n(\lambda), s_n(\lambda))$, we have

$$\left\{ \inf_{|\lambda - \lambda_0| > \delta} \rho_n^*(\lambda) \geq z \right\} \supset \cap_{i=1}^4 A_{in}$$

and

$$\{\rho_n^*(\lambda_0) \leq z/2\} \supset A_{2n} \cap A_{3n} \cap A_{5n} \cap A_{6n}.$$

Therefore, for $n \geq n_0 = \max_{1 \leq i \leq 4} n_i$

$$P\left(\left|\hat{\lambda}_n - \lambda_0\right| > \delta\right) \leq \sum_{i=1}^{6} P(A_{in}^c) \leq \varepsilon,$$

where A^c denotes the complement of A. The consistency of $\hat{\boldsymbol{\beta}}_n$ follows easily from H11 (i).

Acknowledgement

This work was completed with the support of Grant 2053-066895.01 from the Swiss National Science Foundation, Grant PICT–99 03–06277 from the Agencia Nacional de Promoción de la Ciencia y la Tecnología, Grant X-094 from the Universidad de Buenos Aires, Argentina, and a grant from Fundación Antorchas, Argentina.

References

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

Marazzi A., Yohai V. (2005). Robust Box-Cox transformations based on residual autocorrelation. *Computational Statistics & Data Analysis.* Accepted for publication.

Yohai V. J., Rousseeuw P. (1984). Robust regression by means of S-estimators. In: *Robust and nonlinear time series*, edited by Franke J., Härdle W., Martin D., Lecture Notes in Statistics No 26, Springer Verlag, 256-272.

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

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