

On the Consistency of a Class of Nonlinear Regression Estimators

Asheber Abebe
221 Parker Hall
Department of Mathematics and Statistics
Auburn University, AL 36849
abebeas@auburn.edu

Joseph W. McKean
Western Michigan University
joseph.mckean@wmich.edu

Huybrechts F. Bindele
Auburn University
hfb0001@auburn.edu

Abstract

In this paper, we study conditions sufficient for strong consistency of a class of estimators of parameters of nonlinear regression models. The study considers continuous functions depending on a vector of parameters and a set of random regressors. The estimators chosen are minimizers of a generalized form of the signed-rank norm. The generalization allows us to make consistency statements about minimizers of a wide variety of norms including the L_1 and L_2 norms. By implementing trimming, it is shown that high breakdown estimates can be obtained based on the proposed dispersion function.

Keywords and phrases: Nonlinear regression, Signed-rank, Order statistics, Strong consistency.

AMS 2000 subject classification: Primary 62J02, 62G05; Secondary 62F12, 62G20.

We congratulate Professor Mir Masoom Ali. His work in statistics has been commendable and his dedication to its teaching is praiseworthy.

1. Introduction

Over the last twenty five years considerable work has been done on robust procedures for linear models. Several classes of robust estimates have been proposed for these models. One such class is the generalized signed-rank class of estimates. This class uses an objective function which depends on the choice of a score function, φ^+ . If φ^+ is monotone then the objective function is a norm and the geometry of the resulting robust analysis, (estimation, testing, and confidence procedures), is similar to that of the geometry of the traditional least squares (LS) analysis; see McKean and Schrader (1980). Generally this robust analysis is highly efficient relative to the LS analysis; see the monograph by Hettmansperger and McKean (1998) for a discussion of this analysis. For

the simple location model, if Wilcoxon scores, $\varphi^+(u) = u$, are used then this estimate is the famous Hodges-Lehmann estimate while if sign scores are used, $\varphi^+(u) \equiv 1$, it is the sample median. If the monotonicity of φ^+ is relaxed then high breakdown estimates can be obtained; see Hössjer (1994). Thus the signed-rank family of robust estimates for the linear model contain estimates which range from highly efficient to those with high breakdown and they generalize traditional nonparametric procedures in the simple location problem.

Many interesting problems, though, are nonlinear in nature. Traditional procedures based on LS estimation have been used for years. Since these LS procedures for nonlinear models use the Euclidean norm they are as easily interpreted as their linear model counterparts. The asymptotic theory for nonlinear LS has been developed by Jennrich (1969) and Wu (1981), among others. In this paper, we propose a nonlinear analysis based on the signed-rank objective function. The objective function is a norm if φ^+ is monotone; hence, the estimates are easily interpretable. We keep our development quite general, though, to include nonlinear estimates based on Hössjer-type estimates also. Hence our estimates include the nonlinear extensions of the signed-rank Wilcoxon estimate and the L_1 estimate as well as the extensions of high breakdown linear model estimates. Thus we offer a rich family of estimates from which to select for nonlinear models.

Abebe & McKean (2007) studied the asymptotic properties of the Wilcoxon estimator for the general nonlinear model. Just as in linear models, this estimator was shown to be efficient but sensitive to local changes in the direction of \mathbf{x} . Jurečková (2008) studied the asymptotic properties of general rank tests using regression rank scores for the nonlinear model. Her approach uses the asymptotic equivalence of regression quantiles and regression rank scores. This limits the set score functions that can be used. In contrast, our proposed estimator allows for a set of scores generated by any nondecreasing bounded score function that has at most a finite number of discontinuities. In Section 2 we present our family of estimates for nonlinear models.

In Section 3, we show that these estimates are strongly consistent under certain assumptions. We discuss these assumptions, contrasting them with assumptions for current existing estimates. The same section contains a general discussion of interesting special cases such as the L_1 and the Wilcoxon. Section 4 discusses the conditions needed to achieve positive breakdown of our estimator. In Section 5 we provide the proofs of our theory.

2. Definition and Existence

Consider the following general regression model

$$y_i = f(\mathbf{x}_i, \theta_0) + e_i, \quad 1 \leq i \leq n \quad (2.1)$$

where $\theta_0 \in \Theta$ is a vector of parameters, $\mathbf{x}_i \in \mathbb{X}$ is a vector of independent variables, and f is a real-valued function defined on $\mathbb{X} \times \Theta$. Let $\mathbf{V} = \{(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)\}$ be the set of sample data points. Note that $\mathbf{V} \subset \mathbb{V} \equiv \mathbb{R} \times \mathbb{X}$.

We shall assume that Θ is compact, θ_0 is an interior point of Θ , and $f(\mathbf{x}, \theta)$ is a continuous function of θ for each $\mathbf{x} \in \mathbb{X}$ and a measurable function of \mathbf{x} for each $\theta \in \Theta$.

We define the estimator of θ_0 to be any vector θ minimizing

$$D_n(\mathbf{V}, \theta) = \frac{1}{n} \sum_{i=1}^n a_n(i) \rho(|z(\theta)|_{(i)}) \quad (2.2)$$

where $z_i(\theta) = y_i - f(\mathbf{x}_i, \theta)$ and $|z(\theta)|_{(i)}$ is the i th ordered value among $|z_1(\theta)|, \dots, |z_n(\theta)|$. The function $\rho: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is continuous and strictly increasing. The numbers $a_n(i)$ are scores generated as $a_n(i) = \varphi^+(i/(n+1))$, for some bounded score function $\varphi^+: (0,1) \rightarrow \mathbb{R}^+$ that has at most a finite number of discontinuities.

This estimator will be denoted by $\hat{\theta}_n$.

Because $D_n(\mathbf{V}, \theta)$ is continuous in θ , Lemma 2 of Jennrich (1969) implies the existence of a minimizer of $D_n(\mathbf{V}, \theta)$.

We adopt Doob's (1994) convention and denote by L^p , $1 \leq p \leq \infty$, the space of measurable functions $h: (0,1) \rightarrow \mathbb{R}$ for which $|h|^p$ is integrable for $1 \leq p < \infty$ and the space of essentially bounded measurable functions for $p = \infty$. The L^p norm of h is $\|h\|_p \equiv \{\int |h|^p\}^{1/p}$ for $1 \leq p < \infty$ and $\|h\|_\infty \equiv \text{ess sup } |h|$ for $p = \infty$. All integrals are with respect to Lebesgue measure on $(0,1)$. The range of integration will be assumed to be $(0,1)$ unless specified otherwise.

3. Consistency

Let (Ω, \mathcal{F}, P) be a probability space. For $i=1, \dots, n$, assume that \mathbf{x}_i and $e_i = y_i - f(\mathbf{x}_i; \theta_0)$ are independent random variables (carried by (Ω, \mathcal{F}, P)) with distributions H and G , respectively. We shall write \mathbf{x} , e and $|z(\theta)|$ for \mathbf{x}_1 , e_1 and $|z_1(\theta)|$ respectively. Let \tilde{G}_θ denotes the distribution of $|z(\theta)|$ and we will assume

A1: $P(f(\mathbf{x}; \theta) = f(\mathbf{x}; \theta_0)) < 1$ for any $\theta \neq \theta_0$;

A2: for $1 \leq q \leq \infty$, assume there exists a function h such that $|\rho(G_\theta^{-1}(y))| \leq h(y)$, $\forall \theta \in \Theta$ with $E[h^q(Y)] < \infty$ and,

A3: G has a density g that is symmetric about 0 and strictly decreasing on \mathbb{R}^+ .

As usual, we let *a.s.* convergence, denote almost sure convergence, i.e., pointwise convergence everywhere except for possibly an event in \mathcal{F} of probability 0.

Theorem 1.1 Under A1 - A3, $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$.

Remark 1.2 Assumption A1 is a very weak condition needed for θ_0 to be identified. The linear version of A1 was given by Hössjer (1994) as $P(|\theta' \mathbf{x}|=0) < 1$ under the assumption that $\theta_0 = 0$.

Remark 2.3 Since $\|\varphi^+\|_p < \infty$ for p such that $1/p + 1/q = 1$, then A2 puts h and φ^+ in conjugate spaces when $p \in (1, \infty)$. Hölder's inequality ensures that the product $(\varphi^+)(\rho \circ \tilde{G}_\theta^{-1})$ is integrable. Furthermore, if ρ is a convex function, an application of Minkowski's inequality yields

$$\{E[\rho(|z(\theta)|)]\}^{1/q} \leq \{E[\rho(|e|)^q]\}^{1/q} + \{E[\rho(|f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta_0)|)^q]\}^{1/q}.$$

Thus separate conditions on e and f are sufficient for $E[\rho(|z(\theta)|)^q] < \infty$.

Remark 3.4 Condition A3 admits a wide variety of error distributions examples of which are the normal, double exponential and Cauchy distributions with location parameter equal to 0.

Some Corollaries

Next some special cases of interest are considered. We consider the L_1 , least squares, signed-rank Wilcoxon, and their trimmed variations. All these cases involve a convex ρ and hence Remark 2 is directly applicable. Trimming is implemented by "chopping-off" the ends of the score generating function, φ^+ [cf Hössjer (1994)]. The proofs follow from Theorem 1 in a straightforward manner.

Least Squares, Least Trimmed Squares

Let $I_A(\omega)$ be a function such that $I_A(\omega) = 1$ if $\omega \in A$ and $I_A(\omega) = 0$ otherwise. Let $\varphi^+(u) = I_{(\alpha, \beta)}(u)$ for $0 \leq \alpha < \beta \leq 1$ and $\rho(w) = w^2$ for $w \geq 0$. In the case where $\alpha = 0$ and $\beta = 1$ the dispersion function given by (2.2) is the least squares dispersion function. If $0 < \alpha < \beta < 1$, then the dispersion function becomes the least trimmed squares dispersion. The following corollary gives the sufficient conditions for the strong consistency of the least squares estimator by taking $p = q = 2$ in Theorem 1.

Corollary 1.5 If

B1: $P(f(\mathbf{x}; \theta) = f(\mathbf{x}; \theta_0)) < 1$ for any $\theta \neq \theta_0$,

B2: $E(e^2) < \infty$ and $E([f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta_0)]^2) < \infty$ for all $\theta \in \Theta$, and

B3: G has a density g that is symmetric about 0 and strictly decreasing on \mathbb{R}^+ ,

then the least squares (least trimmed squares) estimator is strongly consistent for θ_0 .

Jennrich (1969) establishes the strong consistency of the least squares estimator under some assumptions. His assumptions in the notation of this paper are

J1: $E([f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta_0)]^2) = 0$ if and only if $\theta = \theta_0$,

J2: $E(e^2) < \infty$ and $E([f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta_0)]^2) < \infty$ for all $\theta \in \Theta$, and

J3: $E(e) = 0$.

Assumptions B2 and J2 are identical. B3 and J3, while not generally comparable, are identical in most practical situations where a symmetric, unimodal error density is assumed. Proceeding to compare B1 and J1, assume that B1 fails to hold, that is there exists a point $\theta' \neq \theta_0$ in Θ such that $P(f(\mathbf{x}; \theta') = f(\mathbf{x}; \theta_0)) = 1$. This implies that $E([f(\mathbf{x}; \theta') - f(\mathbf{x}; \theta_0)]^2) = 0$. Thus J1 fails. The converse is also immediate. Hence our assumptions reduce to the assumptions of Jennrich (1969) in the case of least squares.

For linear models, the consistency of the least trimmed squares estimator is established by Víšek (2006). He considers the estimator to be nonlinear, since a subset of the data is considered, and establishes consistency using two different approaches: (1) using an asymptotic linearity argument and (2) using the uniform law of large numbers of Andrews (1987). Čížek (2006) applied the approach used in Víšek (2006) and studied least trimmed squares estimators for nonlinear regression models. His study included models with certain types of dependence such as β -mixing. The conditions given in Víšek (2006) and Čížek (2006) are general; however, our approach establishes consistency for a much larger class of models and estimators.

L_1 , Trimmed Absolute Deviations

The L_1 estimator corresponds to the case where $\varphi^+ \equiv 1$ and $\rho(w) = w$ for $w \geq 0$. A situation similar to the least trimmed squares estimator holds for the trimmed absolute deviations estimator. The sufficient conditions for the strong consistency of the L_1 and trimmed absolute deviations estimators can be found from Theorem 1 by taking $p = \infty$ and $q = 1$. These are given in the following corollary.

Corollary 2.6

C1: $P(f(\mathbf{x}; \theta) = f(\mathbf{x}; \theta_0)) < 1$ for any $\theta \neq \theta_0$,

C2: $E(|e|) < \infty$ and $E(|f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta_0)|) < \infty$ for all $\theta \in \Theta$, and

C3: G has a density g that is symmetric about 0 and strictly decreasing on \mathbb{R}^+ , then the L_1 (trimmed absolute deviations) estimator is strongly consistent for θ_0 .

We next compare the result in Corollary 2 with the one given by Oberhofer (1982). Oberhofer proves the *weak* consistency by imposing the following conditions.

O1: If Θ^* is a closed set not containing θ_0 , then there exist numbers $\varepsilon > 0$ and n_0 such that for all $n \geq n_0$

$$\inf_{\theta \in \Theta^*} n^{-1} \sum_{i=1}^n |l_i(\theta)| \min\{G(|l_i(\theta)|/2) - 1/2, 1/2 - G(-|l_i(\theta)|/2)\} \geq \varepsilon.$$

for all such θ^* where $l_i(\theta) = f(\mathbf{x}_i; \theta) - f(\mathbf{x}_i; \theta_0)$.

O2: $E(|e|) < \infty$ and $E([f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta_0)]^2) < \infty$ for all $\theta \in \Theta$, and

O3: $G(0) = 1/2$.

Here O3 is weaker than C3. However, O2 is stronger than C2. Following similar contrapositive arguments as in the least squares case, we can easily show that O1 is also stronger than C1 (see also Oberhofer (1982) p. 318). For a detailed discussion of this and sufficient conditions for O1, the reader is referred to Oberhofer (1982).

Signed-Rank Wilcoxon

Set $\varphi^+(u) = u$ for $0 < u < 1$ and $\rho(w) = w$ for $w \geq 0$. The following corollary gives the sufficient conditions for the strong consistency of the signed-rank Wilcoxon estimator. The proof is analogous to the proof of Corollary 2 and thus omitted.

Corollary 3. 7 If

D1: $P(f(\mathbf{x}; \theta) = f(\mathbf{x}; \theta_0)) < 1$ for any $\theta \neq \theta_0$,

D2: for some $r \geq 1$, $E(|e|^r) < \infty$ and $E(|f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta_0)|^r) < \infty$ for all $\theta \in \Theta$, and

D3: G has a density g that is symmetric about 0 and strictly decreasing on \mathbb{R}^+ , then the signed-rank Wilcoxon estimator is strongly consistent for θ_0 .

Remark 4. 8 Normal Scores

The frequently used normal scores are generated by

$$\varphi^+(u) = \Phi^{-1}\left(\frac{u+1}{2}\right),$$

for $u \in (0,1)$ where Φ represents the standard normal distribution function. These scores were first proposed by Fraser (1957). Since φ^+ needs to be bounded for our approach to work, our results do not directly extend to the case of normal scores. However, we may use Winsorized normal scores such as

$$\varphi^+(u) = \begin{cases} \Phi^{-1}(-k), & \text{if } u < -2k-1; \\ \Phi^{-1}\left(\frac{u+1}{2}\right), & \text{if } -2k-1 \leq u < 2k-1; \\ \Phi^{-1}(k), & \text{if } u \geq 2k-1. \end{cases}$$

Usually we take $k = 4$.

4. Breakdown Point

One of the virtues of the estimators discussed in this paper is that they allow for trimming. This in turn provides us with estimates that are robust when one or more of the model assumptions are violated. In this section we will consider the breakdown point of our estimator as a measure of its robustness. Assuming that the true value of the

parameter to be estimated is in the interior of the parameter space Θ , breakdown represents a severe form of *inconsistency* in that the estimator converges to a point on the boundary of Θ instead of θ_0 .

Recall that $\mathbf{V} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\} \subset \mathbb{V}$ denotes the sample data points. Let \mathcal{V}^m be the set of all data sets obtained by replacing any m points in \mathbf{V} by arbitrary points. The finite sample breakdown point of an estimator $\hat{\theta}$ is defined as [see Donoho and Huber (1983)]

$$\varepsilon_n^*(\hat{\theta}, \mathbf{V}) = \min_{1 \leq m \leq n} \left\{ \frac{m}{n} : \sup_{\mathbf{Z} \in \mathcal{V}^m} |\hat{\theta}(\mathbf{Z}) - \hat{\theta}(\mathbf{V})| = \infty \right\}, \quad (4.1)$$

where $\hat{\theta}(\mathbf{V})$ is the estimate obtained based on the sample \mathbf{V} . In nonlinear regression, however, this definition of the breakdown point fails since ε^* is not invariant to nonlinear reparameterizations. For a discussion of this see Stromberg and Ruppert (1992). We will adopt the definition of breakdown point for nonlinear models given by Stromberg and Ruppert (1992). The definition proceeds by defining finite sample upper and lower breakdown points, ε_+ and ε_- , which depend on the regression model, f . For any $\mathbf{x}_0 \in \mathbb{X}$, the upper and lower breakdown points are defined as

$$\varepsilon_+(f, \hat{\theta}, \mathbf{V}, \mathbf{x}_0) = \begin{cases} \min_{0 \leq m \leq n} \left\{ \frac{m}{n} : \sup_{\mathbf{Z} \in \mathcal{V}^m} f(\mathbf{x}_0, \hat{\theta}(\mathbf{Z})) = \sup_{\theta} f(\mathbf{x}_0, \theta) \right\} \\ \quad \text{if } \sup_{\theta} f(\mathbf{x}_0, \theta) > f(\mathbf{x}_0, \hat{\theta}), \\ 1 \quad \text{otherwise,} \end{cases} \quad (4.2)$$

and

$$\varepsilon_-(f, \hat{\theta}, \mathbf{V}, \mathbf{x}_0) = \begin{cases} \min_{0 \leq m \leq n} \left\{ \frac{m}{n} : \inf_{\mathbf{Z} \in \mathcal{V}^m} f(\mathbf{x}_0, \hat{\theta}(\mathbf{Z})) = \inf_{\theta} f(\mathbf{x}_0, \theta) \right\} \\ \quad \text{if } \inf_{\theta} f(\mathbf{x}_0, \theta) < f(\mathbf{x}_0, \hat{\theta}), \\ 1 \quad \text{otherwise.} \end{cases} \quad (4.3)$$

Let

$$\varepsilon(f, \hat{\theta}, \mathbf{V}, \mathbf{x}_0) = \min\{\varepsilon_+(f, \hat{\theta}, \mathbf{V}, \mathbf{x}_0), \varepsilon_-(f, \hat{\theta}, \mathbf{V}, \mathbf{x}_0)\}.$$

The finite sample breakdown point is now defined as

$$\varepsilon(f, \hat{\theta}, \mathbf{V}) = \inf_{\mathbf{x}_0 \in \mathbb{X}} \{\varepsilon(f, \hat{\theta}, \mathbf{V}, \mathbf{x}_0)\}. \quad (4.4)$$

The finite sample upper and lower breakdown points are defined analogously by replacing ε by ε_+ and ε_- , respectively, in the above definition. Stromberg and Ruppert (1992) also show that $\varepsilon = \varepsilon^*$ in the case of a linear regression (i.e. $f(\mathbf{x}, \theta) = \mathbf{x}'\theta$) and $\varepsilon = n^{-1}$ for nonlinear least squares regression as expected.

Assume the scores $a_n(i)$ are nonnegative and

$$k = \max \{i : a_n(i) > 0\}$$

where $k \geq [n/2] + 1$. Here $[b]$ stands for the greatest integer less than or equal to b . This forces at least the first half of the ordered absolute residuals to contribute to the dispersion function. In light of this, the dispersion function may be written as

$$D_n(\mathbf{V}, \theta) = \frac{1}{n} \sum_{i=1}^k a_n(i) \rho(|z(\theta)|_{(i)})$$

The following theorem is a version of Theorem 3 of Stromberg and Ruppert (1992). We impose the same conditions but give the result in terms of k . The results given are for upper breakdown. Analogues for lower breakdown are straightforward. The proof is obtained by replacing $med_{1 \leq i \leq n}$ with $n^{-1} \sum_{i=1}^k$ and m with $n-k$ in Stromberg and Ruppert's (1992) proof of Theorem 3. In the following, $\#(A)$ denotes the cardinality of the set A .

Theorem 2.9 Assume for some fixed \mathbf{x}_0 there exist $\tau_k \subset \{i : 1 \leq i \leq n\}$ where $\#(\tau_k) = 2n - [n/2] - k$ such that

$$\lim_{M \uparrow \infty} \left\{ \inf_{\{\theta : f(\mathbf{x}, \theta) > M\}} \left\{ \inf_{i \in \tau_k} f(\mathbf{x}_i, \theta) \right\} \right\} = \sup_{\theta} f(\mathbf{x}_0, \theta)$$

Then

$$\varepsilon_+(f, \hat{\theta}, \mathbf{V}, \mathbf{x}_0) \geq \frac{n-k+1}{n}.$$

Theorem 2 establishes that even when the regression function f lies on the boundary for a portion of the data, the bias of the estimator of θ_0 remains within reasonable bounds if trimming is implemented. The following corollary gives the asymptotic (as $n \rightarrow \infty$) breakdown point of $\hat{\theta}_n$.

Corollary 4.10 Let $\alpha = \sup\{u : \varphi^+(u) > 0\}$. The asymptotic breakdown point of $\hat{\theta}_n$ is at least $1 - \alpha$.

This is reminiscent of the breakdown point of a linear function of order statistics which is equal to the smaller one of the two fractions of mass at either ends of the distribution which receive weights equal to zero (Hampel, 1971). The same result obtained in Corollary 4 was given by Hampel (1971) for one-sample location estimators based on linear functions of order statistics (see sec. 7 (i) of Hampel (1971)).

Consider the class of models with the form $f(x, \theta) = g(\beta_0 + \beta_1 x)$, where $(\beta_0, \beta_1) \in \mathbb{R}^2$ and $g(t)$ is monotone increasing in t . This class of models is considered by Stromberg and Ruppert (1992) and contains popular models like the logistic regression model $g(\beta_0, \beta_1 x) = \{1 + \exp(-(\beta_0 + \beta_1 x))\}^{-1}$. A breakdown point of $1 - \alpha$ can be achieved if $\hat{\theta}_n$

is obtained via a minimization of (2.2) with $a_n(i) = \varphi^+(i/(n+1))$ such that $\alpha = \sup\{u : \varphi^+(u) > 0\}$.

Remark 5.11A *A definition of breakdown based on 'badness measures' which includes the definition given by Stromberg and Ruppert (1992) was given by Sakata and White (1995). Under our assumptions this definition reduces to the one used in the current paper as shown in Theorem 2.3 of Sakata and White (1995).*

5. Proofs

Let $\xi_{(1)}, \dots, \xi_{(n)}$ be order statistics from a sample of n i.i.d. uniform(0,1) random variables. Let $J_n : (0,1) \rightarrow \mathbb{R}$, $n=1,2,\dots$ be Lebesgue measurable functions and let $g : (0,1) \rightarrow \mathbb{R}$ be a Borel measurable function. Define $g_n(t) \equiv g(\xi_{([nt]+1)})$. In the defining expression for the function $D_n(\mathbf{V}, \theta)$, (2.2), let \tilde{G}_θ denote the cdf of $|z(\theta)|$. Then we can express $D_n(\mathbf{V}, \theta)$ as

$$D_n(\mathbf{V}, \theta) = \frac{1}{n} \sum_{i=1}^n a_n(i) (\rho \circ \tilde{G}_\theta^{-1})(\xi_{(i)}). \quad (5.1)$$

The following is Corollary 2.1 of van Zwet (1980) in the notation of this paper and is given for completeness.

Lemma 1 (van Zwet) **12** *Let $1 \leq p \leq \infty$, $1/p + 1/q = 1$, and suppose that $J_n \in L^p$ for $n=1,2,\dots$, $g \in L^q$, and there exists a function $J \in L^p$ such that $\lim_{n \rightarrow \infty} \int_0^t J_n = \int_0^t J$ for all $t \in (0,1)$. If either*

(i) $1 < p \leq \infty$ and $\sup_n \|J_n\|_p < \infty$, or

(ii) $p=1$ and $\{J_n : n=1,2,\dots\}$ is uniformly integrable,

then $\int J_n g_n \xrightarrow{a.s.} \int J g$.

For our purposes let $J_n(t) = \sum_{i=1}^n \varphi^+(i/(n+1)) I_{((i-1)/n, i/n]}(t)$ for $i=1,\dots,n$ where I_A is the indicator of the set A and take $J = \varphi^+$. Notice that J_n is a step function and thus the uniform integrability condition in assumption (ii) of Lemma 1 becomes

$$\limsup_{\alpha \rightarrow \infty} \frac{1}{n} \sum_{i \in A_\alpha} |\varphi^+(i/(n+1))| = 0,$$

where $A_\alpha = \{i : |\varphi^+(i/(n+1))| > \alpha\}$. This condition is satisfied if we have convergence in L^1 of J_n [cf. also Doob (1994), Theorem VI.18]. To this end, we will marginally violate assumption (ii) of Lemma 1 and assume that

$$\sup_n \|J_n\|_p \equiv \sup_n \left\{ \frac{1}{n} \sum_{i=1}^n |\varphi^+(i/(n+1))|^p \right\}^{1/p} < \infty \quad (5.2)$$

for $1 \leq p \leq \infty$. Notice also that

$$\frac{1}{n} \sum_{i=1}^{[nt]} \varphi^+(i/(n+1)) \leq \int_0^t J_n \leq \frac{1}{n} \sum_{i=1}^{[nt]+1} \varphi^+(i/(n+1)).$$

Taking the limit as $n \rightarrow \infty$ we obtain that $\lim_{n \rightarrow \infty} \int_0^t J_n = \int_0^t \varphi^+$ for all $t \in (0,1)$ provided that φ^+ has at most a finite number of discontinuities. Thus if φ^+ satisfies (5.2) and $g \in L^q$ all the conditions of Lemma 1 hold. The following corollary is a special case of this result.

Corollary 5.13 *Let W_1, \dots, W_n be a random sample from a distribution F with support on \mathbb{R}^+ . Let $\rho: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be a continuous Borel measurable function. Suppose, for $1 \leq p, q \leq \infty$ with $1/p + 1/q = 1$, $E[\rho(W)]^q < \infty$ and $\|\varphi^+\|_p < \infty$. Then*

$$T_n \equiv n^{-1} \sum_{i=1}^n \varphi^+(i/(n+1)) \rho(W_{(i)}) \xrightarrow{a.s.} \int (\varphi^+)(\rho \circ F^{-1}) < \infty.$$

A formal proof of Corollary 5 may be constructed along the lines described in the paragraph preceding it with the function g defined as $\rho \circ F^{-1}$. It will not be included here for the sake of brevity.

Lemma 2.14 *Under assumptions A1 - A3*

$$D_n(\mathbf{V}, \theta) \xrightarrow{a.s.} \mu(\theta) \quad a.e. \mathbb{V}, \text{ uniformly for all } \theta \in \Theta, \quad (5.3)$$

where $\mu: \Theta \rightarrow \mathbb{R}$ is a function satisfying

$$\inf_{\theta \in \Theta^*} \mu(\theta) > \mu(\theta_0), \quad (5.4)$$

for any Θ^* a closed subset of Θ not containing θ_0 .

Proof. The *a.s.* pointwise convergence of $D_n(\mathbf{V}, \theta)$ follows from expression (5.1) and Corollary 5, which also furnishes the function

$$\mu(\theta) \equiv \int (\varphi^+)(\rho \circ \tilde{G}_\theta^{-1}) < \infty. \quad (5.5)$$

Then under A1 - A3, Theorem 2 of Jennrich (1969) gives (5.3).

To establish (5.4) we follow a similar strategy as in Hössjer (1994). Under A1 and A3 for any $s > 0$, for $\theta \neq \theta_0$,

$$\begin{aligned} \tilde{G}_\theta(s) &= P(|e - \{f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta_0)\}| \leq s) \\ &= E_{\mathbf{x}} \{P_e(|e - \{f(\mathbf{x}; \theta) - f(\mathbf{x}; \theta_0)\}| \leq s | \mathbf{x})\} \\ &< E_{\mathbf{x}} \{P_e(|e| \leq s)\} = \tilde{G}_{\theta_0}(s) \end{aligned}$$

Since μ is a continuous function depending on θ only through $\rho \circ \tilde{G}_\theta^{-1}$ and since ρ is a strictly increasing function, it follows that $\mu(\theta) > \mu(\theta_0)$ whenever $\theta \neq \theta_0$. Thus for any $\theta \in \Theta^*$, we have a $\mu^* \in \mathbb{R}$ such that $\mu(\theta) > \mu^* > \mu(\theta_0)$. Then it follows from the compactness of Θ^* that $\inf_{\theta \in \Theta^*} \mu(\theta) > \mu(\theta_0)$.

Lemma 3.15 *Let $\{h_n\}$ be a sequence of continuous functions defined on a compact set $\Theta \subset \mathbb{R}^p$ and that converges uniformly to h . Then $\{h_n\}$ is equicontinuous on Θ .*

Proof. Since $\{h_n\}$ converges uniformly to h , for any $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that $|h_n(\theta) - h(\theta)| < \varepsilon/3$ for all $n \geq N$. The function h being continuous on a compact set, it is uniformly continuous. Thus there exists some $\delta > 0$ such that for all $\theta, \theta' \in \Theta$ such that $\|\theta - \theta'\| < \delta$, we have $|h(\theta) - h(\theta')| < \varepsilon/3$. Then for all $n \geq N$ and for all $\theta, \theta' \in \Theta$ such that $\|\theta - \theta'\| < \delta$, we have

$$|h_n(\theta) - h_n(\theta')| \leq |h_n(\theta) - h(\theta)| + |h_n(\theta') - h(\theta')| + |h(\theta) - h(\theta')| < \varepsilon.$$

Also, by uniform continuity of $\{h_n\}$, for any fixed $n \in \{1, \dots, N-1\}$, there exists a $\delta_n > 0$ such that for all $\theta, \theta' \in \Theta$ with $\|\theta - \theta'\| < \delta_n$, we have $|h_n(\theta) - h_n(\theta')| < \varepsilon$. Now set $\delta' = \min\{\delta_1, \dots, \delta_{N-1}\}$. Then for all $n \in \{1, \dots, N-1\}$ and all $\theta, \theta' \in \Theta$ with $\|\theta - \theta'\| < \delta'$, we have $|h_n(\theta) - h_n(\theta')| < \varepsilon$.

Therefore, setting $\Delta = \min\{\delta, \delta'\}$, for all $n \in \mathbb{N}$ and all $\theta, \theta' \in \Theta$ with $\|\theta - \theta'\| < \Delta$, we have $|h_n(\theta) - h_n(\theta')| < \varepsilon$.

Proof of Theorem 1. By Lemma 1 of Wu (1981), to establish the consistency of $\hat{\theta}_n$, it is sufficient to show that

$$\liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta^*} (D_n(\mathbf{V}, \theta) - D_n(\mathbf{V}, \theta_0)) > 0 \quad a.s. \quad (5.6)$$

for any Θ^* a closed subset of Θ not containing θ_0 . But

$$\begin{aligned} \liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta^*} (D_n(\mathbf{V}, \theta) - D_n(\mathbf{V}, \theta_0)) &\geq \liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta^*} A_n(\mathbf{V}, \theta) + \\ &\inf_{\theta \in \Theta^*} B(\theta, \theta_0) + \liminf_{n \rightarrow \infty} C_n(\mathbf{V}, \theta_0), \end{aligned} \quad (5.7)$$

where $A_n(\mathbf{V}, \theta) = D_n(\mathbf{V}, \theta) - \mu(\theta)$, $B(\theta, \theta_0) = \mu(\theta) - \mu(\theta_0)$, and $C_n(\mathbf{V}, \theta_0) = \mu(\theta_0) - D_n(\mathbf{V}, \theta_0)$.

As a result of Corollary 5, $\liminf_{n \rightarrow \infty} C_n(\mathbf{V}, \theta_0) = 0$ a.s. Due to Lemma 2 we have $\inf_{\theta \in \Theta^*} B(\theta, \theta_0) > 0$. For the statement given in (5.6) to hold, it suffices to show is that $\liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta^*} A_n(\mathbf{V}, \theta) = 0$ a.s. Again by Lemma 2, $A_n(\mathbf{V}, \theta) \xrightarrow{a.s.} 0$ uniformly for all $\theta \in \Theta^*$. Also $A_n(\mathbf{V}, \theta)$, being continuous on a compact set Θ^* , is uniformly

continuous on Θ^* . Then $A_n(\mathbf{V}, \theta)$ is equicontinuous on Θ^* a.e. \mathbb{V} by Lemma 3. Thus $\forall \varepsilon > 0$ there exists a $\delta > 0$ such that $\forall \theta, \theta' \in \Theta^*$,

$$\text{if } \|\theta - \theta'\| < \delta \text{ then } |A_n(\mathbf{V}, \theta) - A_n(\mathbf{V}, \theta')| < \varepsilon, \text{ a.e. } \mathbb{V}, \forall n \in \mathbb{N}. \quad (5.8)$$

Let $D_{\theta'} = \{\theta : \|\theta - \theta'\| < \delta\}$, for $\theta' \in \Theta^*$. Then $D_{\theta'}$, $\theta' \in \Theta^*$, forms an open covering of Θ^* . But Θ^* is compact, hence there is a finite subcovering $D_{\theta'_j}$, $j = 1, \dots, m$ which covers Θ^* . Let θ^* be an arbitrary point in Θ^* . Then for some $j = 1, \dots, m$, $\theta^* \in D_{\theta'_j}$.

Hence, $\|\theta^* - \theta'_j\| < \delta$. Thus by (5.8)

$$|A_n(\mathbf{V}, \theta^*) - A_n(\mathbf{V}, \theta'_j)| < \varepsilon, \text{ a.e. } \mathbb{V}, \forall n \in \mathbb{N}.$$

That is,

$$A_n(\mathbf{V}, \theta'_j) - \varepsilon < A_n(\mathbf{V}, \theta^*) < A_n(\mathbf{V}, \theta'_j) + \varepsilon, \text{ a.e. } \mathbb{V}, \forall n \in \mathbb{N}$$

which implies that

$$\min_{1 \leq j \leq m} A_n(\mathbf{V}, \theta'_j) - \varepsilon < A_n(\mathbf{V}, \theta^*) < \max_{1 \leq j \leq m} A_n(\mathbf{V}, \theta'_j) + \varepsilon, \text{ a.e. } \mathbb{V}, \forall n \in \mathbb{N}.$$

Since, θ^* is arbitrary, we have

$$\min_{1 \leq j \leq m} A_n(\mathbf{V}, \theta'_j) - \varepsilon < \inf_{\theta^* \in \Theta^*} \{A_n(\mathbf{V}, \theta^*)\} < \max_{1 \leq j \leq m} A_n(\mathbf{V}, \theta'_j) + \varepsilon, \text{ a.e. } \mathbb{V}, \forall n \in \mathbb{N}.$$

Now take \liminf of all three parts as $n \rightarrow \infty$. Since the functions \min and \max are continuous, we have

$$0 - \varepsilon \leq \liminf_{n \rightarrow \infty} \inf_{\theta^* \in \Theta^*} \{A_n(\mathbf{V}, \theta^*)\} \leq 0 + \varepsilon \text{ a.s.}$$

Since ε was arbitrary, we have $\liminf_{n \rightarrow \infty} \inf_{\theta^* \in \Theta^*} \{A_n(\mathbf{V}, \theta^*)\} = 0$ a.s. The proof is complete.

References

1. Abebe, A. and McKean, J. W. (2007). Highly efficient nonlinear regression based on the Wilcoxon norm. In D. Umbach, editor, *Festschrift in Honor of Mir Masoom Ali*, pages 340--357.
2. Andrews, D. W. K. (1987). Consistency in nonlinear econometric models: a generic uniform law of large numbers. *Econometrica*. 55 1465--1471.
3. Čížek, P. (2006). Least trimmed squares in nonlinear regression under dependence. *J. Statist. Plann. Inference* 136 3967--3988.
4. Donor, D. L. and Huber, P. J. (1983). The notion of breakdown point. *A Festschrift for Erich L. Lehmann*, eds. P. J. Bickel, K. A. Doksum, and J. L. Hodges, Jr., Belmont, CA : Wadsworth, 157--184.
5. Doob, J. L. (1994). *Measure Theory*. Springer-Verlag, New York.

6. Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics*. John Wiley and Sons, New York.
7. Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* 42 1887--1896.
8. Hettmansperger, T. P. and McKean, J. W. (1998). *Robust Nonparametric Statistical Methods*. Arnold, London.
9. Hössjer, O. (1994). Rank-based estimates in the linear model with high breakdown point. *J. Amer. Statist. Assoc.* 89 149--158.
10. Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* 40 633--643.
11. Jurečková, J. (2008). Regression rank scores in nonlinear models. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, 173--183, Inst. Math. Stat. Collect., 1, Inst. Math. Statist., Beachwood, OH, 2008.
12. McKean, J. W. and Schrader, R. (1980). The geometry of robust procedures in linear models. *J. Roy. Statist. Soc. Ser. B* 42 366--371.
13. Oberhofer, W. (1982). The consistency of nonlinear regression minimizing the L_1 -norm. *Ann. Statist.* 10 316--319.
14. Sakata, S. and White, H. (1995). An alternative definition of finite-sample breakdown point with applications to regression model estimators. *J. Amer. Statist. Assoc.* 90 1099--1106.
15. Stromberg, A. J. and Ruppert, D. (1992). Breakdown in nonlinear regression. *J. Amer. Statist. Assoc.* 87 991--997.
16. van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes. With applications to statistics*. Springer-Verlag, New York.
17. van Zwet, W. (1980). A strong law for linear functions of order statistics. *Ann. Probability* 8 986--990.
18. Višek, J. Ā. (2006). The least trimmed squares. I. Consistency. *Kybernetika (Prague)*. 42 1--36.
19. Wu, C. F. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* 9 501--513.