

NONLINEAR LEAST TRIMMED SQUARES

PAVEL ČÍŽEK

Abstract. The estimation of regression models is often based on the least squares method, which is very sensitive to misspecification and data errors. An alternative estimation approach is based on the theory of robust statistics, which builds upon parametric specification, but provides a methodology for designing misspecification-proof estimators. However, this concept, developed in statistics, has so far been applied almost exclusively to linear regression models. Therefore, I adapt the least trimmed squares estimator to nonlinear regression models. This paper presents the adapted robust estimator and the proof of its consistency. Additionally, I derive the asymptotic distribution of the nonlinear least trimmed squares including its asymptotic variance.

Zusammenfassung. In der Regressionsanalyse werden hufig die Methoden, die auf dem Prinzip der kleinsten Quadrate beruhen, eingesetzt, obwohl sie sehr empfindlich gegen Datenfehler und Fehlspezifikation des Modells sind. Im Gegensatz dazu ermöglicht es die Theorie der robusten Statistik, robuste parametrische Schätzer zu entwerfen (zum Beispiel, die Least Trimmed Squares (LTS) Methode). Diese robuste Methoden sind aber fast ausschlieslich nur fur lineare Modelle geeignet. Um den LTS Schätzer fur die nichtlineare Regression zu adaptieren, verallgemeinere ich den LTS fur nichtlineare Modelle und beweise, dass der LTS Schutzer in diesen Follen konsistent ist. Schlieslich leite ich auch die asymptotische Verteilung des LTS inklusive der asymptotischen Varianz ab.

1. Introduction

Least trimmed squares (LTS) is a statistical technique for estimation of unknown parameters of a linear regression model. It was proposed by Rousseeuw (1985) as a robust alternative to the classical least squares method, which, while being frequently used in regression analysis, is quite sensitive to data contamination and model misspecification. Although the asymptotic and robust properties of this estimator were already studied by Rousseeuw and Leroy (1987), LTS was not widely used until recently. There are several reason for this, but the main one is computational: it is possible to compute LTS only approximately and even obtaining an approximation was relatively time consuming; moreover, a good approximation algorithm did not previously exist. However, availability of a good and fast approximation algorithm (see, for example, Rousseeuw and Van Driessen (1999)) and faster computers make LTS more attractive recently, of course, hand in hand with the presence of these algorithm in some widely-spread statistical packages.

Still, the LTS estimator has several shortcomings concerning especially its applicability in nonlinear regression. The only existing results cover its robust properties (Stromberg (1992)) and consistency (Chen, Stromberg, Zhou (1997)) in nonlinear regression. In this paper, I aim to extend the asymptotic results concerning the LTS estimation of nonlinear regression models, namely to prove its $\sqrt{n}$-consistency

2000 *Mathematics Subject Classification.* 62F35 62J02 62F12.
Key words and phrases. Least trimmed squares, nonlinear regression, robust statistics.

and asymptotic normality. This provides not only the rate of convergence, but most importantly the asymptotic variance of LTS.

Why is it useful to think about nonlinear models at all? Let me exemplify this. It is sometimes not clear, for instance, which functional form describes best the dependence on an explanatory variable. To resolve this point, the Box-Cox transformation can be used (see Box and Cox (1964)), that is a transformation of a random variable Z parameterized by $\lambda \in \mathbb{R}$ that represents various functions of Z for different values of λ: linear ($\lambda = 1$), logarithmic ($\lambda = 0$), inversely proportional ($\lambda = -1$), and so on. Applying the transformation to dependent and independent variables provides then a parameterized choice between different regression models (linear, log-linear, semi-logarithmic, reciprocal, etc.). Another example of an intrinsically nonlinear model can be a model with an exponential regression function but an additive error term (instead of a multiplicative one). Finally, time series models with state-dependent regression function are becoming more widely used (see Tong (1990) for summary of these models) and they are typically estimated with nonlinear least squares as well. Thus, they represent another class where nonlinear LTS can be applied.

Let me now precise the goal of the work. In this paper, I study the behavior of the LTS estimator applied in the nonlinear regression model

$$y_i = h(x_i, \beta) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $y_i \in \mathbb{R}$ represents the dependent variable and $h(x, \beta)$ is a function of $x_i \in \mathbb{R}^k$, a vector of explanatory variables, and of $\beta \in \mathbb{R}^p$, a vector of unknown parameters. The LTS estimator used within this framework is further referred to as the *nonlinear least trimmed squares* (NLTS) estimator in order to differentiate it from the LTS estimator used within the linear-regression framework. In the paper, I first review important facts about LTS and define NLTS itself (Section 2). Later, I discuss necessary assumptions for the asymptotic properties of NLTS (Section 3.2), and finally, I derive the asymptotic linearity, consistency, and asymptotic normality of the proposed NLTS (Sections 3.4 and 3.5).

2. Definition of nonlinear least trimmed squares

To assure easy understanding, it is beneficial to describe first the least trimmed squares estimator, introduced by Rousseeuw (1985), and its properties (Section 2.1). The nonlinear LTS estimator is introduced in Section 2.2.

2.1. Least trimmed squares.

Let us consider a linear regression model $y_i = \beta^T x_i + \varepsilon_i$ for $i = 1, \ldots, n$. The least trimmed squares estimator $\hat{\beta}_n^{(\mathrm{LTS})}$ is then defined as

$$\hat{\beta}_n^{(\mathrm{LTS})} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{h} r_{[i]}^2(\beta), \tag{2}$$

where $r_{[i]}^2(\beta)$ represents the ith order statistics of squared residuals $r_1^2(\beta), \ldots, r_n^2(\beta)$; $r_i(\beta) = y_i - \beta^T x_i$ and $\beta \in \mathbb{R}^p$. The *trimming constant* h have to satisfy $\frac{n}{2} < h \leq n$. This constant determines robustness of the LTS estimator, since definition (2) implies that $n - h$ observations with the largest residuals do not have a direct influence on the estimator. The highest level of robustness is achieved for $h = [n/2] + [(p+1)/2]$ (Rousseeuw and Leroy (1987, Theorem 6)), whereas the LTS robustness is lowest for $h = n$, which corresponds to the least squares estimator. There is, of course, a trade-off: lower values of h lead to a higher degree of robustness, while higher values of

h improve efficiency (if the data are not too contaminated) since more (presumably correct) information in the data is utilized.

2.2. **Definition of nonlinear least trimmed squares.** Let us consider model (1), $y_i = h(x_i, \beta) + \varepsilon_i$, where y_i is the dependent variable and $h(x_i, \beta)$ is a known regression function of the data x_i and a vector β of p unknown parameters. Given a sample (y_i, x_i), the NLTS estimate $\hat{\beta}_n^{(\mathrm{NLTS},h)}$ is defined by

$$\hat{\beta}_n^{(\mathrm{NLTS},h)} = \arg\min_{\beta \in B} \sum_{i=1}^{h} r_{[i]}^2(\beta), \tag{3}$$

where $B \subset \mathbb{R}^p$ is the parameter space, $r_{[i]}^2(\beta)$ represents the ordered sample of squared residuals $r_i^2(\beta) = (y_i - h(x_i, \beta))^2$, and h is the trimming constant (see Section 2.1). This estimator shares in most cases its robustness properties with the already reviewed LTS (see Stromberg (1992)) and the same is true for many of its asymptotic properties as will become gradually evident in what follows.

3. Consistency and asymptotic normality of NLTS

In this section I present the main asymptotic results concerning NLTS, namely, its asymptotic linearity, consistency, and asymptotic normality. Before proving these properties, an alternative definition of NLTS and some notational conventions used in the rest of the paper are mentioned as well as the assumptions necessary for the mentioned asymptotic results.

3.1. **Alternative definition of NLTS, notation.** Given a sample (y_i, x_i), the NLTS estimator of unknown parameter vector β is defined for model (1) by equation (3). The true underlying value of the vector β in (1) will be referred to by β^0. Further, the most important function of y_i, x_i, and β is the residual function $r^2(x_i, y_i; \beta) = (y_i - h(x_i, \beta))^2$. Given some fixed $\beta \in B$, the distribution function of residuals $r^2(x_i, y_i; \beta)$ is denoted G_β and its density g_β, if it exists. Specifically for $\beta = \beta^0$, the index will be omitted, that is $G_{\beta^0} \equiv G$ and $g_{\beta^0} \equiv g$, which represent the distribution function and probability density function of ε_i^2.

Next, an alternative definition of NLTS employed in the theoretical part of this paper instead of (3) is given by[1]

$$\hat{\beta}_n^{(\mathrm{NLTS},h)} \quad = \quad \arg\min_{\beta \in B} \sum_{i=1}^{n} r_i^2(\beta) \cdot I\Big(r_i^2(\beta) \le r_{[h]}^2(\beta)\Big). \tag{4}$$

To obtain formula (4), one has to realize that for a given value of $\beta \in B$, the minimization of the h smallest squared residuals means that we include in the objective function only those residuals that satisfy $r_i^2(\beta) \le r_{[h]}^2(\beta)$.[2]

One additional note concerns the trimming constant: whenever asymptotic properties of NLTS are studied, that is $n \to +\infty$, we have to work with a sequence of trimming constants h_n (for every sample size n, there has to be a corresponding choice of h). As this constant determines the robustness properties of NLTS, we want to prescribe asymptotically a fixed fraction λ of observations that are considered to be

[1]By I(property describing a set A) we denote the indicator of the set A.

[2]In general, this definition is not equivalent to the first one. They are exactly equivalent if and only if all the residuals are different from each other. Under the assumptions given in Section 3.2, this happens with zero probability and definitions (3) and (4) are equivalent almost surely as the cumulative distribution function of $r_i(\beta)$ is assumed to be absolutely continuous. Therefore, we further use definition (4) for convenience.

correct, $\frac{1}{2} < \lambda \leq 1$, or alternatively, a fraction $1-\lambda$ of observations that are excluded from the objective function of NLTS ($0 \leq 1-\lambda < \frac{1}{2}$). The trimming constant for a given $n \in \mathbb{N}$ can be then defined by $h_n = [\lambda n]$, where $[\cdot]$ represents the integer part, and hence $h_n/n \to \lambda$. From now on, we assume that there is such a number $\lambda \in (\frac{1}{2}, 1\rangle$ for a sequence h_n of trimming constants defining the NLTS estimator.

To close this section, we discuss some purely mathematical notation. An open neighborhood of a point $x \in \mathbb{R}^l$ is denoted by $U(x,\delta) = \left\{z \in \mathbb{R}^l : \|z-x\| < \delta\right\}$. Moreover, let us denote the convex span of $x_1, \ldots, x_m \in \mathbb{R}^l$ by $[x_1, \ldots, x_m]_\kappa$. Finally, let 1_n represent n-dimensional vector of ones and $\mathcal{I}_n$ be the $n \times n$ identity matrix.

3.2. **Assumptions.** Now, I specify all the assumptions necessary to prove the asymptotic linearity of NLTS. They form two groups–distributional assumptions D for random variables in (1) and assumptions H concerning properties of function $h(x,\beta)$.

First of all, let me discuss the distributional assumptions dealing with the random variables used in model (1). We argue in a number of remarks that the following assumptions do not restrict us in any way in real applications.

Assumption D.

D1: Let $(x_i, \varepsilon_i) \in \mathbb{R}^k \times \mathbb{R}$, $i = 1, \ldots, n$, be a sequence of independent identically distributed random vectors with finite fourth moments and let x_i and ε_i be mutually independent. Moreover,

$$n^{-1/4} \max_{i,j} |x_{ij}| = \mathcal{O}_p(1). \tag{5}$$

***Remark* 1.** *The necessity to include assumption (5) is caused by the discontinuity of the objective function of NLTS. A nonrandom version of this assumption was used, for example, by Víšek (1999). Using Proposition 1, I claim that equation (5) holds even for some distribution functions with polynomial tails, namely for those that have finite second moments. This becomes apparent once we realize that a distribution with tails behaving like one over a polynomial of the third or lower order does not have finite second moments. As the existence of finite second moments is one of the necessary conditions here, assumption (5) should not pose a considerable restriction on the explanatory variables. You can also notice that random variables with a finite support are not restrained by this assumption in any way.*

Proposition 1. *Let $x_1, x_2, \ldots$ be a sequence of independent identically distributed random variables with a distribution function $F(x)$. Let $b(x)$ be a lower bound for $F(x)$ in a neighborhood U_1 of $+\infty$. If $b(x)$ can be chosen as $1 - \frac{1}{P_4(x)}$, where $P_4(x)$ is a polynomial of the fourth order, then it holds that $n^{-\frac{1}{4}} \max_{i=1,\ldots,n} x_i = \mathcal{O}_p(1)$ as $n \to +\infty$. Analogously, let $c(x)$ be an upper bound for $F(x)$ in a neighborhood U_2 of $-\infty$. If $c(x)$ can be chosen as $\frac{1}{P_4(x)}$, where $P_4(x)$ is a polynomial of the fourth order, then it holds that $n^{-\frac{1}{4}} \min_{i=1,\ldots,n} x_i = \mathcal{O}_p(1)$ as $n \to +\infty$.*

Proof. See Čížek (2001, Proposition 1). □

D2: We assume $\mathsf{E}\, x_i x_i^T = Q$, where Q is a nonsingular matrix, and

$$\mathsf{E}\left(\varepsilon_i I\Big(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\Big)\Big|\, x_i\right) = 0, \qquad \mathsf{E}\left(\varepsilon_i^2 I\Big(r_i^2(\beta^0) \leq r_{[h_n]}^2(\beta^0)\Big)\Big|\, x_i\right) = \sigma^2,$$

where $\sigma^2 \in (0, +\infty)$.

***Remark* 2.** *This is a natural analogy of usual orthogonality condition $\mathsf{E}(\varepsilon|x) = 0$ and sphericality condition $\mathsf{E}(\varepsilon\varepsilon^T|x) = \sigma^2\mathcal{I}$ in the case of the linear regression model.*

D3: The distribution function F of ε_i is absolutely continuous. Let f denote the probability density of F, which is assumed to be positive, bounded by $M_f > 0$ and differentiable on the whole support of the distribution function F. Let f' denote the first derivative of f.

***Remark* 3.** *Note that Assumption D3 implies the following property of the distribution function $F(x)$ and its density $f(x)$: for any $0 < \alpha < 1$ we can find $\varepsilon > 0$ such that $\inf_{x \in (F^{-1}(\alpha)-\varepsilon, F^{-1}(\alpha)+\varepsilon)} \min\{F(x), f(x)\} > 0$. The same is true for $G(x)$and $g(x)$.*

***Remark* 4.** *In Assumption D3, the existence of the probability density function f and its derivative is required. If the explanatory variables are uniformly bounded, $\sup_{i \in \mathbb{N}, j=1,\ldots,k} |x_{ij}| = \mathcal{O}(1)$, then it is sufficient for these densities and their derivatives to exist only in a neighborhood of $-\sqrt{G^{-1}(\lambda)}$ and $\sqrt{G^{-1}(\lambda)}$. The same applies to the assumption that the probability densities are bounded.*

As we aim to apply NLTS to nonlinear models, several conditions on the regression function $h(x, \beta)$ have to be specified. Most of them are just regularity conditions that are employed in almost any work concerning nonlinear regression models. Because the assumptions stated below rely on the value of β and I do not have to require their validity over the whole parametric space, I restrict β to a neighborhood $U(\beta^0, \delta)$ in these cases and suppose that there exists a positive constant δ such that all the assumptions are valid.

Assumptions H. Let the following assumptions hold for some $\delta > 0$.

H1: Let $h(x_i, \beta)$ be a continuous in $\beta \in B$ (uniformly over any compact subset of the support of x) and twice differentiable function in β on $U(\beta^0, \delta)$ almost surely. The first derivative is continuous for $\beta \in U(\beta^0, \delta)$.

H2: Furthermore, let us assume that the second derivatives $h''_{\beta_j\beta_k}(x, \beta)$ satisfy locally the Lipschitz property in a neighborhood of β^0.

H3: Let

$$n^{-1/4} \max_{1 \le i \le n} \max_{1 \le j \le p} \left\| h'_{\beta_j}(x_i, \beta) \right\| = \mathcal{O}_p(1) \tag{6}$$

and

$$n^{-1/2} \max_{1 \le i \le n} \max_{1 \le j,k \le p} \left\| h''_{\beta_j\beta_k}(x_i, \beta) \right\| = \mathcal{O}_p(1) \tag{7}$$

as $n \to +\infty$ uniformly over $\beta \in B$.

***Remark* 5.** *Assumption H3 depicts another regularity condition that is going to be fulfilled in most cases. For example, for a function of the form $h(x_i^T \beta)$, where h is twice differentiable, we can immediately observe that $h'_{\beta_j}(x, \beta) = h'(x_i^T \beta) x_{ij}$, and analogously, $h''_{\beta_j\beta_k}(x, \beta) = h''(x_i^T \beta) x_{ij} x_{ik}$. Hence, assumptions (6) and (7) are a direct consequence of (5) as long as the first two derivatives of $h(x, \beta)$ are bounded on any compact subset of the support of random variable x.*

H4: Moreover, we assume that $\mathsf{E}\, h'_\beta(x_i, \beta^0) h'_\beta(x_i, \beta^0)^T = Q_h$, where Q_h is a nonsingular positive definite matrix.

3.3. Asymptotic linearity.

3.3.1. *Normal equations.* In order to analyze the behavior of the NLTS estimator, we use normal equations as the starting point, that is, instead of minimizing the objective function $\rho(\beta) = \sum_{i=1}^{n} r_i^2 \cdot I\Big(r_i^2(\beta) \leq r_{[h]}^2(\beta)\Big)$ over all $\beta \in B$, we consider a solution of $\frac{\partial \rho(\beta)}{\partial \beta} = 0$.

3.3.2. *Asymptotic linearity.* Analogously as for M-estimators or LTS (see Jurečková and Sen (1989), Víšek (1996), or Víšek (1999)), we shall first investigate the term $S_n(t) = \frac{\partial \rho(\beta^0 - n^{-\frac{1}{2}} t)}{\partial \beta} - \frac{\partial \rho(\beta^0)}{\partial \beta}$ for $t \in \mathcal{T}_M = \{t \in \mathbb{R}^p : \|t\| \leq M\}$, where $0 < M < \infty$ is an arbitrary, but fixed constant. More precisely, we show that $S_n(t)$ behaves asymptotically as a linear function of $n^{-\frac{1}{2}} t$ over the whole set $\mathcal{T}_M$.

Theorem 1. *Let Assumptions D and H hold. Then for a given $\lambda \in (\frac{1}{2}, 1\rangle$, it holds that*

$$n^{-\frac{1}{4}} \sup_{t \in \mathcal{T}_M} \left\| S_n(t) + n^{\frac{1}{2}} \cdot Q_h t \cdot [\lambda - q_\lambda \cdot \{f(-q_\lambda) + f(q_\lambda)\}] \right\| = \mathcal{O}_p(1)$$

as $n \to +\infty$, where $q_\lambda = \sqrt{G^{-1}(\lambda)}$.

Proof. See Čížek (2001, Theorem 1). □

3.4. $\sqrt{n}$ consistency.

In this section, I derive the consistency of NLTS. To provide as complete picture as possible, I specify two sets of assumptions. The first group, Assumption NC, is as general as possible and is sufficient just for proving consistency of NLTS; the second group, Assumption NN, allows us to derive $\sqrt{n}$-consistency and asymptotic normality of NLTS. In the presented form, Assumption NC correspond mostly to the assumptions required for the uniform law of large numbers in nonlinear models, which is in a very general form presented in Andrews (1987).

Assumption NC. Let the following assumptions are satisfied for function $q(x_i, \varepsilon_i; \beta) = r_i^2(\beta) \cdot I\Big(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\Big)$, where $r_i(\beta) = \varepsilon_i + h\big(x_i, \beta^0\big) - h(x_i, \beta)$.[3]

- **NC1:** Let the parameter space B be a compact metric space.
- **NC2:** Let $q(x_i, \varepsilon_i; \beta)$, $q^*(x_i, \varepsilon_i; \beta, \rho) = \sup\{q(x_i, \varepsilon_i; \beta) : \beta' \in U(\beta, \rho)\}$, and $q_*(x_i, \varepsilon_i; \beta, \rho) = \inf\{q(x_i, \varepsilon_i; \beta) : \beta' \in U(\beta, \rho)\}$ be measurable random variables for all $\beta \in B, i \in \mathbb{N}$, and for all $\rho > 0$ sufficiently small.
- **NC3:** Let $\mathsf{E}\left\{\sup_{\beta \in B} |q(x_i, \varepsilon_i; \beta)|\right\}^{1+\delta} < \infty$ for some $\delta > 0$.

***Remark* 6.** *Assumptions NC1–NC3 are necessary (together with the assumption concerning the differentiability of function $h(x, \beta)$ with respect to β) for the uniform law of large number. Assumption NC3 is actually a standard condition used in this context to ensure that functions $\{q^*(x_i, \varepsilon_i; \beta, \rho)\}$ and $\{q_*(x_i, \varepsilon_i; \beta, \rho)\}$ satisfy pointwise the strong law of large numbers for any $\beta \in B$ and all ρ sufficiently small; see Andrews (1987), for instance. Moreover, note that*

$$r_i(\beta) = \varepsilon_i + h\big(x_i, \beta^0\big) - h(x_i, \beta) = \varepsilon_i + h_\beta^{'}(x_i, \xi(\beta)) \cdot (\beta - \beta^0), \tag{8}$$

where $h_\beta^{'}(x_i, \xi)$ is bounded independently of β (under Assumption H3) and $\beta - \beta^0$ is bounded as well (B is a compact space). Hence, the existence of an upper bound for $r_i^2(\beta)$ follows from Assumptions H and NC1 and Assumption NC3 just requires existence of a certain expectation of this upper bound.

---

[3]For the case of nonlinear least squares, $\lambda = 1$ and $G_\beta^{-1}(\lambda) = \infty$. Therefore, this case corresponds to $q(x_i, \varepsilon_i; \beta) = r_i^2(\beta)$.

NC4: For any $\varepsilon > 0$ and $U(\beta^0, \varepsilon)$ such that $B - U(\beta^0, \varepsilon)$ is compact, there exists $\alpha(\varepsilon) > 0$ such that it holds

$$\min_{\beta \in B - U(\beta^0, \varepsilon)} \mathsf{E}\, q(x_i, \varepsilon_i; \beta) - \mathsf{E}\, q(x_i, \varepsilon_i; \beta^0) > \alpha(\varepsilon).$$

Remark 7. *This is nothing but an analogy of the identification condition for the nonlinear least squares, see for example White (1980).*

NC5: Assume that $I_G = \inf_{\beta \in B} G_\beta^{-1}(\lambda) > 0$ and $m_{gg} = \inf_{\beta \in B} \inf_{z \in (-\delta, \delta)} g_\beta\left(G_\beta^{-1}(\lambda) + z\right) > 0$, where G_β and g_β are the cumulative distribution function and the probability density function of $r_i^2(\beta)$.

NC6: Let $g_\beta(z)$ is bounded on $(I_G, +\infty)$ uniformly in $\beta \in B$, that is, there is $M_{gg} > 0$ such that $\sup_{\beta \in B} \sup_{z \in (I_G, +\infty)} g_\beta(z) \leq M_{gg}$.

Remark 8. *Although Assumptions NC5 and NC6 might look unfamiliar at the first sight, they just guarantee that the distribution functions of random variables $r_i^2(\beta)$ do not converge to some extreme cases for some $\beta \in B$. Namely, these conditions exclude cases when the expectation or variance of $r_i^2(\beta)$ converge to infinite values for some $\beta \in B$ or when the distribution function G_β converges to a discrete distribution function for some $\beta \in B$. This does not restrict us in commonly used regression models, because the parametric space B is compact.*

The following theorem confirms that Assumption NC is sufficient for the consistency of NLTS.

Theorem 2. *Let Assumptions D, H, and NC hold. Then the nonlinear least trimmed squares estimator defined for model (1) by*

$$\hat{\beta}_n^{(\mathrm{NLTS},h)} = \arg\min_{\beta \in B} \sum_{i=1}^{n} r_i^2(\beta) \cdot I\left(r_i^2(\beta) \leq r_{[h]}^2(\beta)\right)$$

is consistent, that is, $\hat{\beta} \to \beta^0$ in probability as $n \to +\infty$.

Proof. See Čížek (2001, Theorem 2). □

Next, let us recall that Assumption NC is the sufficient condition for the consistency of NLTS. However, if we enrich Assumption NC to below stated Assumption NN, we are able to prove even the $\sqrt{n}$-consistence and asymptotic normality of NLTS. Assumption NN is in principle equivalent to Assumption NC applied not only to the residual function $r_i^2(\beta)$, but also to its first two derivatives.

Assumption NN. Let Assumption NC hold, and additionally, Assumptions NC1–NC3 are satisfied for functions

- $q(x_i, \varepsilon_i; \beta) = r_i^2(\beta) \cdot I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\right)$,
- $q(x_i, \varepsilon_i; \beta) = r_i(\beta) \cdot h''_{\beta_j \beta_k}(x_i, \beta) \cdot I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\right)$,
- $q(x_i, \varepsilon_i; \beta) = h'_{\beta_j}(x_i, \beta) \cdot h'_{\beta_k}(x_i, \beta) \cdot I\left(r_i^2(\beta) \leq G_\beta^{-1}(\lambda)\right)$,

where $j, k = 1, \ldots, p$.

Finally, combining all the conditions stated so far, namely, D, H, and NN, we can prove the $\sqrt{n}$-consistence of NLTS.

Theorem 3. *Let Assumptions D, H, and NN hold. Then $\hat{\beta}_n^{(\mathrm{NLTS},h_n)}$ is $\sqrt{n}$-consistent, that is,*

$$\sqrt{n}\left(\hat{\beta}_n^{(\mathrm{NLTS},h_n)} - \beta^0\right) = \mathcal{O}_p(1)$$

as $n \to +\infty$.

Proof. See Čížek (2001, Theorem 3). □

3.5. **Asymptotic normality.** The $\sqrt{n}$-consistency of NLTS derived in Theorem 3 allows us to use the asymptotic linearity of the objective function (Theorem 1) for $t = \hat{\beta}_n^{(NLTS,h_n)}$ and to derive this way the main result concerning NLTS – its asymptotic normality.

Theorem 4. *Let Assumptions D, H, and NN are fulfilled and*

$$C_\lambda = \lambda - q_\lambda \cdot \left\{ f\left(G^{-1}(\lambda)\right) + f\left(G^{-1}(\lambda)\right) \right\} \neq 0.$$

Then $\hat{\beta}_n^{(\mathrm{NLTS},h_n)}$ is asymptotically normal with its expectation equal to β^0:

$$\sqrt{n}\left(\hat{\beta}_n^{(\mathrm{NLTS},h_n)} - \beta^0\right) \xrightarrow{F} N(0, V),$$

where $V = C_\lambda^{-2} \cdot Q_h^{-1} \,\mathsf{var}\left[\varepsilon_i \cdot h'_\beta\left(x_i, \beta^0\right) \cdot I\left(\varepsilon_i^2 \leq G^{-1}(\lambda)\right)\right] Q_h^{-1}$.

Proof. See Čížek (2001, Theorem 4). □

4. Conclusion

In this paper, I have introduced the nonlinear least trimmed squares estimator and derived its asymptotic properties. Thus, the applicability of LTS is extended to various intrinsically nonlinear models (asymptotic distribution is known). However, given the rather theoretical character of the paper, it remains to be seen whether the existing computational procedures designed for LTS in the linear regression model suit well various nonlinear settings.

Reference

[1] Andrews, D. W. K. (1987): Consistency in Nonlinear Econometic Models: A Generic Uniform Law of Large Numbers, *Econometrica*, Vol. 55 / 6, 1465–1471.

[2] Box, G. E. P. and Cox, D. (1964): An Analysis of Transformations, *Journal of the Royal Statistical Society, Series B*, 211–264.

[3] Chen, Y., Stromberg, A., and Zhou, M. (1997): The least trimmed squares estimate in nonlinear regression, *Technical report, University of Kentucky*, 365/2000.

[4] Čížek, P. (2001): Nonlinear Least Trimmed Squares, *SFB Discussion Paper, Humboldt University*, 25/2001.

[5] Jurečková, J., Sen, P. K. (1989): Uniform Second Order Asymptotic Linearity of M-statistics, *Statistics & Decisions*, Vol. 7, 263–276.

[6] Rousseeuw, P. J. (1985): Multivariate Estimation With High Breakdown Point, in W. Grossman, G. Pflug, I. Vincze, and W. Wertz eds., *Mathematical Statistics and Applications, Vol. B*, Reidel, Dordrecht, The Netherlands, 283–297.

[7] ———, Leroy, A. M. (1987): *Robust Regression and Outlier Detection*, Wiley, New York.

[8] ———, Van Driessen, K. (1999): Computing LTS Regression for Large Data Sets, *Technical Report, University of Antwerp.*

[9] Stromberg, A. J. (1992): Breakdown In Nonlinear Regression, *Journal of American Statistical Association*, Vol. 87, 991–997.

[10] Tong, H. (1990): *Nonlinear time series: A Dynamical Systems Approach*, Oxford University Press, Oxford.
[11] Víšek, J. Á (1996): Sensitivity Analysis of M-estimators, *Annals of the Institute of Statistical Mathematics*, Vol. 48, 469–495.
[12] ——— (1999): The least trimmed squares—random carriers, *Bulletin of the Czech Econometric Society*, Vol. 10/1999, 1–30.
[13] White, H. (1980): Nonlinear Regression on Cross-Section Data, *Econometrica*, Volume 48, No. 3, 721–746.

Humboldt-Universität zu Berlin Wirtschaftswissenschaftliche, Fakultät Institut für Statistik und Ökonometrie, Alexanderplatz 6 10178 Berlin