Least Trimmed Squares - Sensitivity Study

Jan Ámos Víšek

Abstract: Asymptotic representation of the difference of estimates by the least trimmed squares for the whole population of data and for the data without the $\ell$-th observation is derived. It shows that in the case when the supremum of absolute value of any explanatory variable is not (uniformly in number of observations) bounded (e.g. as (usual) in the framework with random carriers), this difference cannot be bounded, too.

MSC 2000: 62J05, 62F35

Key words: The least trimmed squares, sensitivity analysis.

Introduction and notation

Let $N$ denote the set of all positive integers, $R$ the real line and $R^p$ the $p$-dimensional Euclidean space. For any set $A$ let $A^o$ denote the interior of the set (in the topology implied by Euclidean metric). We shall consider for any $n \in N$ the linear regression model

$$Y_i = x_i'\beta^0 + e_i = \sum_{j=1}^{p} x_{ij}\beta_j + e_i, \quad i = 1, 2, \ldots, n$$  (1)

where $Y_i$ and $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ are values of response and of explanatory variables for the $i$-th case, respectively. $\beta^0$ is the vector of regression coefficients and $e_i$ represents random fluctuation (disturbance) of $Y_i$ from the mean value $EY_i$.

For any $\beta \in R^p$ let us denote by $r_i(\beta) = Y_i - x_i' \beta$ and $r^2_i(\beta), i = 1, 2, \ldots, n$ the $i$-th residual and the $h$-th order statistics of squared residuals, respectively.* In what follows the definition of the least trimmed squares (LTS) will be considered in the form:

**Definition 1.** For a compact set $K$ such that the vector of the true regression coefficients $\beta^0 \in K^o$ the estimator given as

$$\hat{\beta}^{\text{LTS},n,h} = \arg\min_{\beta \in K} h \sum_{i=1}^{n} r^2_i(\beta)$$  (2)

will be called the least trimmed squares (LTS).† It is clear that (2) is equivalent to‡

$$\hat{\beta}^{\text{LTS},n,h} = \arg\min_{\beta \in K} \sum_{i=1}^{n} r^2_i(\beta) \cdot I \{ r^2_i(\beta) \leq r^2_{(h)}(\beta) \}$$  (3)

Research was supported by grant of GA ČR 402/06/0408.

*Hereafter we shall assume that the random variables are defined on a basic probability space $(\Omega, \mathcal{A}, P)$.

†The reasons for restricting on a compact $K$, instead of considering $R^p$, were given in [13].

‡The paper is a continuation of the papers The least trimmed squares: Consistency (Part I), √n-consistency (Part II) and Asymptotic representation (Part III) [13], and many details (which were here, due to the limited space, omitted, can be found there).
(where \( I \{ A \} \) denote the indicator of the set \( A \)). As \( \frac{\partial}{\partial \beta} I \left\{ r^2_h(\beta) \leq r^2_{(h)}(\beta) \right\} = 0 \) (see [13], Part 1), \( \hat{\beta}^{(LTS,n,h)} \) is one of solution of the normal equations

\[
\sum_{i=1}^{n} \left[ Y_i - x'_i(\beta) x_i \right] I \left\{ r^2_i(\beta) \leq r^2_{(h)}(\beta) \right\} = 0. \tag{4}
\]

Now, denote \( G(z) \) the distribution function of \( c^2_1 \) and for any \( \alpha \in (0,1) \), \( u^2_{n\alpha} \) will be the upper \( \alpha \)-quantile of \( G(z) \). Further, denote by \( [a]_{\text{int}} \) the integer part of \( a \) and for any \( n \in N \) and \( \alpha \in [0, \frac{1}{2}] \) put§ \( h_n = [(1-\alpha)n]_{\text{int}}. \tag{5} \)

Hereafter a fix \( \alpha_0 \in [0, \frac{1}{2}] \) and the following assumptions will be considered:

**Assumptions** \( \mathcal{A} \) The sequences \( \{ x_i \}_{i=1}^{\infty} (x_i \in R^p) \) is a fix sequence of nonrandom vectors from \( R^p \). Moreover,

\[
\lim_{n \to \infty} Q_n = Q \quad \text{and} \quad \max_{1 \leq i \leq n, 1 \leq j \leq p} |x_{ij}| = O(1) \quad \text{as} \quad n \to \infty \quad (6)
\]

where \( Q \) is a regular matrix (and convergence is of course assumed coordinate-wise). The sequence \( \{ e_i \}_{i=1}^{\infty} (e_i \in R) \) is a sequence of independent and identically distributed random variables with absolutely continuous symmetric distribution function \( F(z) \). There is a neighbourhood of \( u_{\alpha_0} \) in which the distribution \( F(z) \) has a bounded density \( f(z) \) which is positive and has bounded in absolute value the first and the second derivative. The second derivative is further Lipschitz of the first order. Moreover, the density \( f(z) \) is strictly decreasing on \( R^+ \) and \( Ee_i^2 = \kappa_4 \in (0,\infty) \). Finally, there are distribution functions \( H_j(\beta), t \in R, \beta \in R^p \) such that for any compact set \( W \subset R^p \)

\[
\sup_{\beta \in W} \sup_{t \in R} \left| \frac{1}{n} \sum_{i=1}^{n} I \left\{ x'_i(\beta - \beta^0) \leq t \right\} - H_j(\beta)(t) \right| = O(n^{-\frac{1}{2}}). \quad (7)
\]

**Remark** 1. In [13] except of these assumptions, also an alternative collection was used. They did not assume the second part of (6) but on the other hand they had to ask for more restrictive assumptions on disturbances \( e_i \)'s.

**Sensitivity analysis of the least trimmed squares**

For any \( t, v \in R^p \) denote \( r_i(t,v) = r_i(\beta^0 - n^{-\frac{1}{2}} t - n^{-1}v) \) and \( r^2_{(h)}(t,v) \) the \( h \)-th order statistic among \( r^2_i(t,v), i = 1, 2, ..., n \) (write \( r_i(t) \) and \( r^2_{(h)}(t) \) if \( v = 0 \)). Further, denote \( (a,b)_{\text{ord}} = (\min\{a,b\}, \max\{a,b\}) \) and the same for the closed intervals.

**Assertion 1.** Let Assumptions \( \mathcal{A} \) be fulfilled. Then for any \( \varepsilon > 0 \) there is \( K(\varepsilon) \in (0,\infty) \) and \( n_\varepsilon \in N \) such that for all \( n > n_\varepsilon \)

\[
P \left( \sup_{\beta \in K} \left| r^2_{(h)}(\beta) - u^2_{\alpha_0}(\beta) \right| < n^{-\frac{1}{2}} K(\varepsilon) \right) > 1 - \varepsilon \tag{8}
\]

and for any \( \ell, 1 \leq \ell \leq n, \) any \( \tau \in (\frac{1}{2}, \frac{3}{4}) \) and any \( K \in (0,\infty) \) there is \( n_{\varepsilon,K,\tau} \in N \) such that for any \( n > n_{\varepsilon,K,\tau} \) there is a set \( B_n \subset \Omega \) such that \( P(B_n) > 1 - \varepsilon \) and

\[
P \left( \left( c^2_1(\ell), r^2_{(h)}(\ell) \right)_{\text{ord}} \cap B_n \right) \leq n^{-\tau} K. \quad (9)
\]

§Since it seems reasonable to have \( \frac{2}{3} < h_n \leq n. \)
For the proof see Lemma 1 of [13] in Part I and Corrolary 1 in Part II. In the same way we can prove

**Assertion 2.** Let Assumptions \( A \) hold. Then for any \( \ell, 1 \leq \ell \leq n \), any \( \varepsilon > 0 \), any \( K \in (0, \infty) \) and any \( \tau \in (1, \frac{5}{4}) \) there is an \( n_{\varepsilon,K,\tau} \in N \) so that for any \( n > n_{\varepsilon,K,\tau} \) there is a set \( B_n \subset \Omega \) such that \( P(B_n) > 1 - \varepsilon \) and

\[
P\left(e_1^2 \in \left( r_\varepsilon^2(t), r_\varepsilon^2(t,v) \right)_{\text{ord}} \cap B_n \right) \leq n^{-\tau} K. \tag{10}
\]

We shall study the sum

\[
S_n(t,v,h) = \sum_{i=1}^{n} \left[ r_i(t,v)x_iI \left\{ r_i^2(t,v) \leq r_{\varepsilon}^2(t,v) \right\} - r_i(t)x_iI \left\{ r_i^2(t) \leq r_{\varepsilon}^2(t) \right\} \right].
\]

Further let us denote

\[
\xi_i(t,v) = I \left\{ r_i^2(t,v) \leq r_{\varepsilon}^2(t,v) \right\} - I \left\{ r_i^2(t) \leq r_{\varepsilon}^2(t) \right\} \tag{11}
\]

and

\[
T_M = \{ t,v \in R^p, ||t|| < M, ||v|| < M \} . \tag{12}
\]

**Theorem 1.** Denote \( R_n(t,v) = u_{ao} \sum_{i=1}^{n} \text{sign}(e_i)x_i(\xi_i(t,v) - E\xi_i(t,v)) \). Under Assumptions \( A \) we have

\[
\sup_{t,v \in T_M} ||S_n(t,v,h_n) + Q_nv[(1-\alpha_o) - u_{ao}(f(u_{ao}) + f(-u_{ao}))] + R_n(t,v)|| = o_p(1)
\]

as \( n \to \infty \) and where \( \sup_{t,v \in T_M} R_n(t,v) = O_p(1) \) and \( h_n = [(1-\alpha_o)n]_{\text{int}} \).

**Proof.** Write \( h \) and \( \xi_i \) instead of \( h_n \) and \( \xi_i(t,v) \), respectively, and consider

\[
S_n(t,v,h) = \sum_{i=1}^{n} \left\{ [r_i(t,v) - r_i(t)]x_i \cdot I \left\{ r_i^2(t,v) \leq r_{\varepsilon}^2(t,v) \right\} - r_i(t)x_i\xi_i \right\}
\]

\[
= \sum_{i=1}^{n} \frac{1}{n} x_i x'_i \xi_i + \frac{1}{n} x_i x'_i I \left\{ r_i^2(t) \leq r_{\varepsilon}^2(t) \right\} - [r_i(t) - e_i]x_i\xi_i - e_i x_i\xi_i. \tag{13}
\]

According to Assumptions \( A \) for any \( n > n_{\varepsilon,K,\tau} \) we have

\[
P(\{ |\xi_i| \neq 0 \} \cap B_n) = P \left( \left\{ I \left\{ r_i^2(t,v) \leq r_{\varepsilon}^2(t,v) \right\} \neq I \left\{ r_i^2(t) \leq r_{\varepsilon}^2(t) \right\} \right\} \cap B_n \right)
\]

\[
\leq P \left( e_1^2 \in (r_{\varepsilon}^2(t), r_{\varepsilon}^2(t,v))_{\text{ord}} \right) + \sup_{a \in (u_{ao} - \delta, u_{ao} + \delta)} P \left( e_1^2 \in (a, a + n^{-1}x'_iu)_{\text{ord}} \right)
\]

(for some \( \delta > 0 \) and Assertion 2 allows to find for any \( \varepsilon > 0 \) and any \( K_1 \in (0, \infty) \) such \( n_{\varepsilon,K_1} \in N \) and \( \delta > 0 \) so that for any \( n > n_{\varepsilon,K_1} \), we have for \( i = 1, 2, ..., n \)

\[
|\pi_i(t,v)| = |P(\{ |\xi_i| \neq 0 \} \cap B_n)| \leq n^{-1} \cdot K_1 \tag{14}
\]
(where the last equality defines \( \pi_t(t,v) \)) and hence, for any \( L > 0 \),

\[
P \left( \left\{ \left\| \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \xi_i \right\| > L \right\} \cap B_n \right) \leq \frac{1}{Ln} \sum_{i=1}^{n} \left\| x_i \right\|^2 \cdot |v| \cdot \mathbb{E} \left| \xi_i \cdot I \{ B_n \} \right|
\]

\[
\leq \frac{1}{Ln} K_1 \cdot M \sum_{i=1}^{n} \left\| x_i \right\|^2 = O(n^{-1}) = o(1)
\]

(15)

(for \( M \) see (12)), i.e. the first term in (13) is \( o_p(1) \). Similarly, a straightforward computation gives that for the second term of (13) we have

\[
\left\| n^{-1} \sum_{i=1}^{n} x_i x_i' I \left\{ r_i^2(t) \leq r_{(h)}^2(t) \right\} - Qv(1 - \alpha_0) \right\| = o_p(1) \text{ as } n \to \infty.
\]

(16)

Let us recall that \( r_i(t) - e_i = r_i(\beta_0 - n^{-\frac{1}{2}} t) - e_i = Y_i - x'_i \beta_0 + n^{-\frac{1}{2}} x'_i t - e_i = n^{-\frac{1}{2}} x'_i t \)

and consider the third term in (13). Employing (14) once again we arrive for any \( \delta > 0 \) at

\[
P \left( \left\{ \left\| \sum_{i=1}^{n} \left[ r_i(t) - e_i \right] x_i \xi_i \right\| > \delta \right\} \cap B_n \right) = P \left( \left\{ \left\| n^{-\frac{1}{2}} \sum_{i=1}^{n} x_i x_i' \xi_i \right\| > \delta \right\} \cap B_n \right)
\]

\[
\leq \delta^{-1} n^{-\frac{1}{2}} K_1 M \sum_{i=1}^{n} \left\| x_i \right\|^2 = o(1).
\]

(17)

Finally, denote \( C_i = \{ \omega \in \Omega : |\xi_i| = 1 \} \cap B_n \) and consider the last term in (13). First of all, let us fix \( \varepsilon > 0 \) and employing Assertion 1 we find \( K_4 \) and \( n_{\varepsilon,K_4} \in N \) so that for all \( n > n_{\varepsilon,K_4} \) for the set

\[
D_n = \left\{ \omega \in \Omega : \left| e_{(h)}^2 - u_{\alpha_0}^2 \right| < n^{-\frac{1}{2}} K_4 \right\}
\]

(18)

has probability \( P(D_n) > 1 - \varepsilon \). As \( e_{(h)}^2 - u_{\alpha_0}^2 = \left( \sqrt{e_{(h)}^2} - u_{\alpha_0} \right) \cdot \left( \sqrt{e_{(h)}^2} + u_{\alpha_0} \right) \)

and for \( \alpha_0 \in (0, \frac{1}{2}) \) we have \( u_{\alpha_0} > u_{\frac{1}{2}}^2 \), (18) implies that there is also \( K_5 \) so that for all \( n > n_{\varepsilon,K_4} \) and any \( \omega \in D_n \)

\[
\left| \sqrt{e_{(h)}^2} - u_{\alpha_0} \right| < n^{-\frac{1}{2}} K_5.
\]

(19)

Now, due to (6) and (12) there is \( K_6 < \infty \) so that for any \( t \in T_M \)

\[
\left| e_{(h)}^2 - r_{(h)}^2(t) \right| < n^{-\frac{1}{2}} K_6 \quad \text{and} \quad \left| r_{(h)}^2(t) - r_{(h)}^2(t,v) \right| < n^{-\frac{1}{2}} K_6.
\]

Then the definition of \( C_i \) implies that there is a positive \( K_7 < \infty \) so that for all \( \omega \in C_i \) we have

\[
\left| e_i - \sqrt{e_{(h)}^2} \right| < n^{-\frac{1}{2}} K_7.
\]

(20)

Now, (19) and (20) imply that there is a \( K_8 < \infty \) so that for any \( n > n_{\varepsilon,K_4} \) and \( \omega \in C_i \cap D_n \)

\[
\left| e_i - u_{\alpha_0} \right| < n^{-\frac{1}{2}} K_8 \quad \Rightarrow \quad e_i = \text{sign}(e_i) \cdot u_{\alpha_0} + \eta_i
\]

(21)
with \(|\eta_i| < n^{-\frac{1}{2}}K_8\).

Then (21) for any \(n > n_{\varepsilon, K_4}\) and \(\omega \in B_n \cap D_n\) gives

\[
\left| \sum_{i=1}^{n} e_i x_i \cdot \xi_i \right| \leq \left| \text{sign}(e_i) \cdot u_{\alpha} \sum_{i=1}^{n} x_i \xi_i \right| + \sum_{i=1}^{n} \|x_i\| \cdot |\eta_i| \cdot |\xi_i|.
\]

(23)

Let us denote \(\sup_{n \in N} \|x_i\| = U\) (which is due to (6) finite). Now recalling how we have defined in (11) \(\xi_i\)'s and taking into account (14) and (22), we arrive at

\[
\mathbb{E} \left\{ I\{B_n \cap D_n\} \sum_{i=1}^{n} \|x_i\| \cdot |\eta_i| \cdot |\xi_i| \right\} \leq n^{-\frac{1}{2}} U \cdot K_8 \sum_{i=1}^{\infty} \pi_t(t, v) < n^{-\frac{1}{2}} \cdot K_9
\]

for some constant \(K_9 < \infty\) which in turn means that \(\sum_{i=1}^{n} \|x_i\| \cdot |\eta_i| \cdot |\xi_i| = O_p(1)\).

But it allows to write for any \(n > n_{\varepsilon, K_4}\) and \(\omega \in B_n \cap D_n\) the last term of (13) as

\[
u_{\alpha} \sum_{i=1}^{n} \text{sign}(e_i) \cdot x_i \xi_i + o_p(1) = u_{\alpha} \sum_{i=1}^{n} x_i \left( \text{sign}(e_i) \cdot \xi_i - \mathbb{E} \left[ \text{sign}(e_i) \cdot \xi_i \right] \right)
\]

\[
+ \nu_{\alpha} \sum_{i=1}^{n} x_i \mathbb{E} \left[ \text{sign}(e_i) \cdot \xi_i \right] + o_p(1).
\]

(24)

Now, it follows from definition of \(\pi_i^{(j+)}\) and Assertion 2 that there is \(K_{10} \sim \infty\) so that (for all \(i, j, k\))

\[
\pi_i^{(j+)} \leq n^{-1} \cdot K_{10}.
\]

(25)

Then, employing Lemma A.1, we arrive at (see again [3])

\[
u_{\alpha} \sum_{i=1}^{n} x_i \left( I\{e_i > 0\} \cdot \xi_i - \mathbb{E} I\{e_i > 0\} \cdot \xi_i \right) = \mathbb{D} \left( \sum_{i=1}^{n} \left( \pi_{0, i, k}^{(j+)}(t, v) + \pi_{1, i, k}^{(j+)}(t, v) \right) \right)
\]

(26)

and

\[
\mathbb{E} \left[ \pi_{i, k}^{(j+)}(t, v) \right] = \nu_{\alpha}^{2} \left( x_{ik} \right)^{2} \left[ 1 - \pi_i^{(j+)}(t, v) \right] \pi_i^{(j+)}(t, v).
\]

(27)

Then due to (25)

\[
\sum_{i=1}^{n} \left( \pi_{0, i, k}^{(j+)}(t, v) + \pi_{1, i, k}^{(j+)}(t, v) \right) = O_p(1)
\]

which in turn means that for any \(\varepsilon > 0\) there is \(K_{11} \sim \infty\) and \(n_{K_{11}} \in N\) such that for all \(n > n_{K_{11}}\),

\[
|\eta_i| < n^{-\frac{1}{2}}K_8.
\]
\[ P(D_n) = P \left( \omega \in \Omega : \sum_{i=1}^{n} \left( \tau_{i,k}^{(0+)} + \tau_{i,k}^{(1+)} \right) \leq K_{11} \right) > 1 - \varepsilon. \] (28)

Notice please, that the equality (26) means equality of all finite dimensional distribution of two random processes (the “time” parameter of the processes is of course the vector \((t, v)\)). Now, employing [4] we find that the process \(u_n \sum_{i=1}^{n} x_i (I\{e_i > 0\} \cdot \xi_i - \mathbb{E}[I\{e_i > 0\} \cdot \xi_i])\) is tight and hence also (due to (28))

\[ \sup_{t,v \in T_M} u_n \sum_{i=1}^{n} x_{ik}(I\{e_i > 0\} \cdot \xi_i - \mathbb{E}[I\{e_i > 0\} \cdot \xi_i]) \cdot I(D_n) \leq \sup_{z \in (0,K_{11})} |W(z)| = \mathcal{O}_p(1). \] (29)

The same we obtain for \(u_n \sum_{i=1}^{n} x_{ik}(I\{e_i < 0\} \cdot \xi_i - \mathbb{E}[I\{e_i < 0\} \cdot \xi_i]) \cdot I(D_n)\). As

\[ \sup_{t,v \in T_M} \left\| R_n(t,v) \right\| = \sup_{t,v \in T_M} u_n \alpha \sum_{i=1}^{n} \text{sign}(e_i) \cdot x_i (\xi_i - \mathbb{E}[\text{sign}(e_i) \cdot \xi_i]) = \mathcal{O}_p(1). \] (30)

Finally, we need to analyze \(\mathbb{E}[\text{sign}(e_i) \cdot \xi_i]\). Employing Assertion 2, and a routine analysis we arrive at

\[ \sup_{(t,v) \in T_M} \left\| u_n \alpha \sum_{i=1}^{n} x_i \left\{ \mathbb{E}[\text{sign}(e_i) \cdot \xi_i] - n^{-1} x_i \cdot v \left( f(u_n) + f(-u_n) \right) \right\} \right\| = o(1) \] (31)

as \(n \to \infty\). Taking into account (15), (16), (17), (29), (30) and (31) we conclude the proof of Theorem.

**Theorem 2.** Let Assumptions \( \mathcal{A} \) be fulfilled. Then, denoting \( \hat{\beta}^{(LTS,n-1,t,h)} \) estimate for the data from which \(t\)-th observation has been deleted, we have

\[ n \left( \hat{\beta}^{(LTS,n,h)} - \hat{\beta}^{(LTS,n-1,t,h)} \right) = \mathcal{O}_p(1) \] as \(n \to \infty\).

**Proof.** Let us write \(e_{(h,n)}^2\) and \(e_{(h,n-1,t)}^2\) for the \(h\)-th order statistic among squared random fluctuations \(e_{i,k}^2\)'s in the “whole” data set and for data set without the \(t\)-th observation, respectively. Recalling that \(e_{(h)}^2 = \tau_{(h)}^2(\beta^0)\), we immediately obtain from (8) that for any \(\varepsilon > 0\) there are \(n_{\varepsilon} \in \mathbb{N}\) and \(K_1 < \infty\) so that for any \(n > n_{\varepsilon}\) we have

\[ B_n = \left\{ \omega : |e_{(h,n)}^2 - e_{(h,n-1,t)}^2| \leq n^{-\frac{1}{4}} K_1 \right\} \text{ with } P(B_n) > 1 - \varepsilon. \] (32)

Now, let us define a mapping \(a : \Omega \to \{1,2,...,n\}^h\) so that \(a(\omega) = (i_1,i_2,...,i_h)\) is the point of Cartesian product \(\{1,2,...,n\}^h\) for which

\[ \min_{\beta \in \mathbb{R}^p} \sum_{i \in a(\omega)} (Y_i(\omega) - x_i'\beta)^2 = \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{h} \tau_{i,j}^2(\beta,\omega) \] (33)
(for a “prove” of existence of such point, see [13], Part I). Further, let us denote by
\( Y(a) = Y(a(ω)) \), \( e(a) = e(a(ω)) \) and \( X(a) = X(a(ω)) \) corresponding \( h \)-dimensional
subvector of the vector \( Y \), of random fluctuations \( e \) and submatrix of the matrix
\( X \) containing all lines indices of which fall into the set \( a(ω) \), respectively. Finally,
denote by \( β^{(LS,h)}(Y(a), X(a)) = β^{(LS,h)}(Y(a(ω)), X(a(ω))) \) the least squares estimator
evaluated just for subpopulation \( [Y(a(ω)), X(a(ω))] \). Then it is clear that
\[
β^{(LTS,n,h)}(ω) = β^{(LS,h)}(Y(a(ω)), X(a(ω)))
\]
where \( a^{(n)} = a^{(n)}(ω) \) denotes corresponding mapping \( a^{(n)} : Ω → \{1, 2, ..., n\}^h \)
and \( β^{(LTS,n,h)}(ω) \) the value of \( β^{(LTS,n,h)}(ω) \) at the point \( ω \) (and a similar expression we
have for \( β^{(LTS,n−1,ℓ,h)} \)). Then, writing \( X(n) \) for \( X(a^{(n)}) \), \( X(n−1,ℓ) \) for \( X(a^{(n−1,ℓ)}) \)
and similarly for \( e(a^{(n)}) \) and \( e(a^{(n−1,ℓ)}) \), we arrive at
\[
n(β^{(LTS,n,h)} − β^{(LTS,n−1,ℓ,h)} = n(β^{(LTS,n,h)} − β^0) − n(β^{(LTS,n−1,ℓ,h)} − β^0)
\]
\[
= (\frac{1}{n}X′(n)X(n))^{-1}X′(n)e(n) − \frac{1}{n}X′(n−1,ℓ)X(n−1,ℓ)^{-1}X′(n−1,ℓ)e(n−1,ℓ)
\]
\[
= (\frac{1}{n}X′(n)X(n))^{-1}[X′(n)e(n) − X′(n−1,ℓ)e(n−1,ℓ)]
\]
(34)

Now, we may write
\[
\frac{1}{n}X′(n)X(n)
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} x_i x_i′ \left[ I(e_i^2 ≤ e_{(h,n)}^2) − ΕI(e_i^2 ≤ e_{(h,n)}^2) \right] + \frac{1}{n} \sum_{i=1}^{n} x_i x_i′ ΕI(e_i^2 ≤ e_{(h,n)}^2)
\]
and hence
\[
(\frac{1}{n}X′(n)X(n)) → Q(1 − α_0)
\]
(35)
in probability. Similarly
\[
X′(n)e(n) − X′(n−1,ℓ)e(n−1,ℓ)
\]
\[
= \sum_{i=1}^{n} x_i e_i I(e_i^2 ≤ e_{(h,n)}^2) − \sum_{i=1}^{n} x_i e_i I(e_i^2 ≤ e_{(h,n−1,ℓ)}^2)
\]
\[
= x_ℓ e_ℓ I(e_ℓ^2 ≤ e_{(h,n)}^2) + \sum_{i=1,i≠ℓ}^{n} x_i e_i u_{iα} \left[I(e_i^2 ≤ e_{(h,n)}^2) − I(e_i^2 ≤ e_{(h,n−1,ℓ)}^2)\right]
\]
(36)

First of all, let us notice that either \( e_i^2 ≤ e_{(h,n)}^2 \) or \( e_i^2 > e_{(h,n)}^2 \). In the former case,
we have \( e_{(h,n−1,ℓ)}^2 = e_{(h,n−1,ℓ)}^2 \) and taking into account that
\( P(e_{(h,n)}^2 ≠ e_{(h,n−1,ℓ)}^2) = 1 \), we find that only for such index \( i \) for which \( e_i^2 = e_{(h,n−1,ℓ)}^2 = e_{(h,n−1,ℓ)}^2 = e_{(h,n−1,ℓ)}^2 \)
(where the last equality defines \( k_{h+1}^0(β^0) \))

\[
I(e_i^2 ≤ e_{(h,n)}^2) − I(e_i^2 ≤ e_{(h,n−1,ℓ)}^2) = −1.
\]
For all other indices the expression is zero. In the latter case \( e_i^2 \leq e_{(h,n)}^2 \) and 
\[ I\{e_i^2 \leq e_{(h,n)}^2\} - I\{e_i^2 \leq e_{(h,n-1)\ell}^2\} = 0. \]
It means that 
\[ \sum_{i=1, i \neq \ell}^n x_i e_i u_n \left[ I\{e_i^2 \leq e_{(h,n)}^2\} - I\{e_i^2 \leq e_{(h,n-1)\ell}^2\} \right] = -x_{h+1}(\beta) e_{h+1}(\beta) = O_p(1) \]
(remember that \( e_{h+1}(\beta) = e_{(h+1,n)}^2 \) which is, with probability at least \( 1 - \varepsilon \), near to \( u_{(h)}^2 \)). Since \( x_i e_i I\{e_i^2 \leq e_{(h,n)}^2\} = O_p(1) \) as \( n \to \infty \) (again due to the fact that \( e_{(h,n)}^2 \) is with high probability near to \( u_{(h)}^2 \)), the first term of (34) is bounded in probability. Along very similar lines we may treat the second term in (34) and that concludes the proof.

Now, let us put \( R_n^* = R_n \left( \sqrt{n} \left( \hat{\beta}(LTS,n,h) - \beta^0 \right), n \left( \hat{\beta}(LTS,n,h) - \hat{\beta}(LTS,n-1,\ell,h) \right) \right) \)
(for \( R_n(t,v) \) see Theorem 1).

**Theorem 3.** Let \( [(1 - \alpha_0) - u_{(h)} \cdot (f(u_{(a)}) + f(-u_{(a)})) + R_n^*]^{-1} = O_p(1). \) Then under Assumptions \( \mathcal{A} \) we have
\[ n \left( \hat{\beta}(LTS,n,h) - \hat{\beta}(LTS,n-1,\ell,h) \right) = Q_n^{-1} \left[ (1 - \alpha_0) - u_{(h)} \cdot (f(u_{(a)}) + f(-u_{(a)})) + R_n^* \right]^{-1} \times \]
\[ \left( Y_t - x_t' \hat{\beta}(LTS,n,h) \right) x_t I \left\{ r_t^2(\hat{\beta}(LTS,n,h)) \leq r_{(h)}^2(\hat{\beta}(LTS,n,h)) \right\} + o_p(1) \] (37)
as \( n \to \infty \).

**Proof.** First of all, let us recall that 
\[ \sum_{i=1}^n r_i(\hat{\beta}(LTS,n,h)) \cdot x_i \cdot I \left\{ r_t^2(\hat{\beta}(LTS,n,h)) \leq r_{(h)}^2(\hat{\beta}(LTS,n,h)) \right\} = 0 \]
(since the equations are the corresponding normal equations) as well as
\[ \sum_{i=1, i \neq \ell}^n r_i(\hat{\beta}(LTS,n-1,\ell,h)) \cdot x_i \cdot I \left\{ r_t^2(\hat{\beta}(LTS,n-1,\ell,h)) \leq r_{(h)}^2(\hat{\beta}(LTS,n-1,\ell,h)) \right\} = 0. \]
So, denoting 
\[ t^{(n)} = \sqrt{n} \left( \hat{\beta}(LTS,n,h) - \beta^0 \right) \quad \text{and} \quad v^{(n)} = n \left( \hat{\beta}(LTS,n,h) - \hat{\beta}(LTS,n-1,\ell,h) \right), \]
we have
\[ S_n(t^{(n)}, v^{(n)}, h_n) = r_t(\hat{\beta}(LTS,n-1,\ell,h)) \cdot x_t \cdot I \left\{ r_t^2(\hat{\beta}(LTS,n-1,\ell,h)) \leq r_{(h)}^2(\hat{\beta}(LTS,n-1,\ell,h)) \right\} \]
\[ = r_t(\hat{\beta}(LTS,n,h)) \cdot x_t \cdot I \left\{ r_t^2(\hat{\beta}(LTS,n,h)) \leq r_{(h)}^2(\hat{\beta}(LTS,n,h)) \right\} + o_p(1). \]
Now the proof of theorem follows from Theorems 1 and 2 when we substitute 
\( t = t^{(n)}, v = v^{(n)} \).
Conclusions
Just concluded study of sensitivity of $\hat{\beta}^{(LTS,n,h)}$ confirmed conjecture that the estimator may behave similarly as $M$-estimators which have discontinuous $\psi$-function, namely that it can be sometimes very sensitive even to deletion of one observation. This is indicated by presence of the term $R_n^*$ in the asymptotic representation (37). Let us recall that $L_1$-estimator, i.e. the estimator which minimizes the sum of absolute values of residuals, also contains in its asymptotic representation a term of type $R_n^*$ (see [6]). The sensitivity study of all these estimators showed that the change of their values, when deleting even one observation, is sometimes much larger than the corresponding change of the $M$-estimators which have continuous $\psi$-function, see again [6], [7] and [11] (see [2] or [14] for an analogical result for the least squares).

From the application point of view this high sensitivity even to (a shift of) one point means that the estimator relies too much on selected subsample while the complementary part of data is assumed, with uncritical belief, to be a contamination. It may lead to “underrating” the role of random fluctuations - i.e. an underestimation of the scale of random fluctuations possibly together with a strange shape of their empirical distribution function. So it seems that it would be preferable to use $M$-estimators with continuous (and preferably redescending) $\psi$ functions. However, $M$-estimators are not (generally) scale- and regression-equivariant and so to achieve this equivariance we have to studentize the residuals by scale-equivariant and regression-invariant scale estimator. Although it is principally possible to evaluate such estimate (see [7]), it is not simple and quick. So, it would be better to avoid it and to use directly such estimator of regression coefficients which is scale- and regression-equivariant.

So to achieve all possible desirable properties of estimator as controllable breakdown point, low gross error sensitivity, low local shift sensitivity, low subsample sensitivity and the scale- and regression-equivariance, we would prefer an estimator with smooth rejection of influential points but with similar behaviour as $\hat{\beta}^{(LTS,n,h)}$ in the case of high contamination (of finite set of data). One can conjecture that the least weighted squares, of course with smoothly decreasing weights (assigned to the order statistics of squared residuals), could meet with these expectations, see [10], [12]. Moreover, unlike $\hat{\beta}^{(LTS,n,h)}$ the least weighted squares can be employed also for panel data.

1 Appendix

Lemma A.1. ([5], page 420, VII.2.8 or Proposition 13.7 (page 277) of [1]) Let $a$ and $b$ be positive numbers. Further let $\xi$ be a random variable such that $P(\xi = -a) = \pi$ and $P(\xi = b) = 1 - \pi$ (for $a$ $\pi \in (0,1)$) and $\mathbb{E}\xi = 0$. Moreover let $\tau$ be the time for the Wiener process $W(s)$ to exit the interval $(-a,b)$. Then $\xi \equiv_{\mathcal{D}} W(\tau)$ where “$\equiv_{\mathcal{D}}$” denotes the equality of distributions of the corresponding random variables. Moreover, $\mathbb{E}\tau = a \cdot b = \text{var}\xi$. 

References


Jan Ámos Víšek: Dept. of Macroeconomics and Econometrics, Inst. of Economic Studies, Fac. of Social Sciences, Charles University and Dept. of Stochastic Informatics, Inst. of Information Theory and Automation, Academy of Sciences of the Czech Republic, Smetanovo nabreží 6, Prague, 110 01, the Czech Republic, visek@mbox.fsv.cuni.cz