A model where the Least Trimmed Squares estimator is maximum likelihood

Supplement

Vanessa Berenguer-Rico[*], Søren Johansen[†] & Bent Nielsen[‡]

19 September 2021

This supplement has two parts. In appendix C, we analyze the Least Median of Squares (LMS) estimators along the lines of the analysis in the main paper of the Least Trimmed Squares (LTS) estimator. In appendix D, we provide the detailed derivation of some identities used in the main paper when proving the asymptotic theory for the LTS estimator.

C Least Median of Squares

The Least Median of Squares (LMS) estimator was suggested along with LTS by Rousseeuw (1984). Originally, LMS was seen as more important than LTS for computational reasons. However, LMS is only $n^{1/3}$-consistent in i.i.d. models and it has a rather complicated asymptotic distribution theory (Kim and Pollard, 1990). Here, we give a likelihood theory for LMS by replacing the normal distribution in the LTS model with a uniform distribution. We show that, in the location-scale case, the LMS estimator has the same asymptotic theory as the infeasible Chebychev estimator applied to the set of 'good' observations. That is, it is h-consistent with a standard Laplace limit.

C.1 The LMS estimator

As before, let $r^2_{(1)}(\beta) \leq \cdots \leq r^2_{(n)}(\beta)$ be the order statistics of $r^2_i(\beta) = (y_i - \beta' x_i)^2$. The LMS estimator is

$$\hat{\beta}_{LMS} = \arg\min_{\beta} r^2_{(h)}(\beta). \tag{C.1}$$

Rousseeuw (1984) was concerned with the case where h is one plus the integer part of $n/2$, but other quantiles are routinely used. The LMS estimator is related to the Chebychev estimator, also referred to as the L_∞ estimator or the minimax estimator. LMS divides the observations into 'good' observations and 'outliers'. The estimated set of 'good' observations is

$$\hat{\zeta}_{LMS} = \arg\min_{\zeta} \max_{i \in \zeta} |y_i - x_i' \hat{\beta}^{\zeta}_{Cheb}| \quad \text{where} \quad \hat{\beta}^{\zeta}_{Cheb} = \arg\min_{\beta} \max_{i \in \zeta} |y_i - \beta' x_i|. \tag{C.2}$$

[*]Department of Economics, University of Oxford. E-mail: vanessa.berenguer-rico@economics.ox.ac.uk.

[†]Department of Economics, University of Copenhagen and CREATES, Department of Economics and Business, Aarhus University, DK-8000 Aarhus C. E-mail: soren.johansen@econ.ku.dk.

[‡]Department of Economics, University of Oxford. Address for correspondence: Nuffield College, Oxford OX1 1NF, UK. E-mail: bent.nielsen@nuffield.ox.ac.uk.

Some properties of Chebychev estimators carry over to LMS. Wagner (1959) pointed out that the Chebychev estimator can be found as a regular linear programming problem with $p+1$ relations, where p is the dimension of x_i. Harter (1953) and Schechtman and Schechtman (1986) found that the Chebychev estimator is a non-unique maximum likelihood estimator in a model with uniform errors with known range. We find, below, that it is maximum likelihood estimator in a uniform model with unknown range. Knight (2017) studied the asymptotic theory of the Chebychev regression estimator in the context of i.i.d. errors.

C.2 The LMS regression model and maximum likelihood

Model C.1 (LMS regression model)**.** *Consider the regression model $y_i = \beta' x_i + \sigma\varepsilon_i$ for data y_i, x_i for $i = 1, \ldots, n$. Let $h \le n$ be given. Follow the setup of the LTS regression Model 3.1 apart from assuming that the 'good' errors ε_i for $i \in \zeta$ are uniformly distributed on $[-1, 1]$.*

The LMS likelihood mimicks the LTS likelihood in (4.3). The only difference is that the normal ϵ-probability $\Delta^\epsilon\Phi(y)$ is replace by a uniform ϵ-probability $\Delta^\epsilon\mathsf{U}(y) = \mathsf{U}(y) - \mathsf{U}(y-\epsilon)$, where $\mathsf{U}(y) = (y+1)/2$ for $|y| \le 1$. In the maximization, the difference is that the profile likelihood (4.4) has the essential part

$$\lim_{\epsilon\to 0}\prod_{i\in\zeta}\epsilon^{-1}\Delta^\epsilon\mathsf{U}(y_i^{\beta\sigma}) = \prod_{i\in\zeta}(2\sigma)^{-1}1_{(|y_i-\beta' x_i|\le\sigma)} = (2\sigma)^{-h}1_{(\max_{i\in\zeta}|y_i-\beta' x_i|\le\sigma)},$$

which is a uniform likelihood. For given β, ζ, this is maximized by choosing σ as small as possible, subject to the constraint $\sigma \ge \max_{i\in\zeta}|y_i - \beta' x_i|$. Minimizing the lower bound for σ over β for fixed ζ gives the Chebychev estimator. Thus, for given ζ the maximizers are

$$\hat\beta_{\zeta,LMS} = \arg\min_\beta\max_{i\in\zeta}|y_i - \beta' x_i| \qquad \text{and} \qquad \hat\sigma_{\zeta,LMS} = \max_{i\in\zeta}|y_i - \hat\beta'_{\zeta,LMS}x_i|. \tag{C.3}$$

The remainder of the LTS maximum likelihood argument stands. We summarize.

Theorem C.1. *Consider the LMS regression Model C.1. The ϵ-likelihood is maximized for $\epsilon \to 0$ as follows. For any h-subsample ζ, define the Chebychev estimators $\hat\beta_{\zeta,LMS}$ and $\hat\sigma_{\zeta,LMS}$ as in (C.3). Let $\hat\zeta_{LMS} = \arg\min_\zeta\hat\sigma_{\zeta,LMS}$, subject to the constraint that $\hat\epsilon_i \ne \hat\epsilon_\ell$ for $i \in \zeta$, $1 \le \ell \le n$ and $\ell \ne i$ and $\ell \ne i$ and where $\hat\epsilon_i = y_i - \hat\beta'_{\zeta,LMS}x_i$. Then, $\hat\beta_{LMS} = \hat\beta_{\hat\zeta_{LMS},LMS}$ and $\hat\sigma_{LMS} = \hat\sigma_{\hat\zeta_{LMS},LMS}$.*

C.3 Asymptotic theory for the location-scale model

We consider a sequence of LMS location-scale models $y_i = \mu + \sigma\varepsilon_i$ defined along the lines of §5.1. The LMS estimator simplifies. Consider sets ζ of the form $y_{(\delta+1)}, \ldots, y_{(\delta+h)}$ for $\delta = 0, \ldots, n-h$, where $y_{(1)} \le \cdots \le y_{(n)}$ are the order statistics. Following Rousseeuw (1984), the LMS estimators reduce to $\hat\mu_{LMS} = \hat\mu_{\hat\delta_{LMS},LMS}$ and $\hat\sigma_{LMS} = \hat\sigma_{\hat\delta_{LMS},LMS}$, and where

$$\hat\mu_{\delta,LMS} = \frac{1}{2}\{y_{(\delta+h)} + y_{(\delta+1)}\}, \quad \hat\sigma_{\delta,LMS} = \frac{1}{2}\{y_{(\delta+h)} - y_{(\delta+1)}\}, \quad \hat\delta_{LMS} = \arg\min_{0\le\delta\le n-h}\hat\sigma_{\delta,LMS}. \tag{C.4}$$

We show that with LMS, the set of 'good' observations is estimated consistently. Regularity conditions are needed, regardless of the proportion λ of 'good' observations, to ensure that the uniform 'good' observations are sufficiently separated from the 'outliers'.

Assumption C.1. *Let $G(x)$ for $x \geq 0$ represent $\overline{\mathsf{G}}(x)$ or $\underline{\mathsf{G}}(x)$. Suppose*
(i) $\exists \epsilon > 0$ so that $\forall 0 < \psi < 1$ then $G^{-1}(\psi) \geq 2\psi\varrho$ where $\varrho = (1-\rho+\epsilon)(1-\lambda)/\lambda$;
(ii) $\exists \psi_0 > 0, \tau < 1$ so that $\forall 0 < \psi < \psi_0$ then $G^{-1}(\psi) \geq \psi^{\tau}$.

In the proof, the first condition is needed to establish that $\hat{\delta}_{LMS} = \delta_n + \mathsf{o_P}(h_n)$. It gives a global bound on the entire distribution function, while allowing bursts of very concentrated 'outliers' as long as they are not in the vicinity of the 'good' observations. The second condition is binding for 'outliers' in the vicinity of the 'good' observations. It is needed to improve the rate of the remainder term for $\hat{\delta}_{LMS}$.

Theorem C.2. *Consider a sequence of LMS location-scale models. Let $1/2 < \lambda < 1$ and suppose Assumption C.1. Let $\overline{e}, \underline{e}$ be independent, standard exponential variables. Then*

$$\mathsf{P}(\hat{\delta}_{LMS} = \delta_n) \to 1, \qquad h_n(\hat{\mu}_{LMS} - \mu)/\sigma \overset{\mathsf{D}}{\to} \underline{e} - \overline{e}, \qquad h_n(\hat{\sigma}_{LMS} - \sigma)/\sigma \overset{\mathsf{D}}{\to} -(\underline{e} + \overline{e}),$$

where $\underline{e} - \overline{e}$ and $\underline{e} + \overline{e}$ are dependent $\mathsf{Laplace}(0,1)$ *and* $-\mathsf{Gamma}(2,1)$ *variables.*

The result generalizes to the case $0 < \lambda < 1$ allowing more 'outliers' than 'good' observations (Berenguer-Rico et al., 2019). It contrasts with the previous, complicated $n^{1/3}$-consistent, complicated asymptotic theory for i.i.d. models (Kim and Pollard, 1990). We expect that the above result would be more complicated for regression models in light of the recent theory of Knight (2017) for Chebychev regression estimators with i.i.d. errors.

C.4 The OLS estimator in the LMS location-scale model

We show that the OLS estimator is inconsistent but with bounded bias in the LMS model. This indicates that the least squares estimator is robust in the sense of Hampel (1971) within a wider class of contamination in the LMS model than in the LTS model.

Theorem C.3. *Consider a sequence of LMS location-scale models. Let $0 < \lambda < 1$ and $\rho = 0$. Suppose $\overline{\mathsf{G}}$ has finite expectation $\mu_{\mathsf{G}} = \int_0^\infty \{1 - \overline{\mathsf{G}}(x)\}dx$. The sample average satisfies*

$$(\bar{\mu} - \mu)/\sigma \overset{\mathsf{P}}{\to} (1-\lambda)(1+\mu_{\mathsf{G}}) > 0.$$

C.5 Proofs of asymptotic theory for the LMS model

We consider the LMS estimator in Model C.1. We suppress the 'LMS' index on estimators.

C.5.1 Uniform spacings & sums of exponential variables

We consider uniform spacings and derive some properties of sums of exponential variables.

Lemma C.1. *(Pyke, 1965, §4.1) Let $u_1, \ldots, u_n$ be i.i.d. standard uniform with order statistics $u_{(1)} < \cdots < u_{(n)}$. The spacings $s_1, \ldots, s_n$ are defined as $s_i = u_{(i)} - u_{(i-1)}$ for $i = 2, \ldots, n$ while $s_1 = u_{(1)}$ and $s_{n+1} = 1 - u_{(n)}$. Further, let $e_1, \ldots, e_{n+1}$ be standard exponential variables $e_1, \ldots, e_{n+1}$. Then $(s_1, \ldots, s_{n+1})$ have the same distribution as $(e_1, \ldots, e_{n+1})/(e_1 + \cdots + e_{n+1})$.*

Lemma C.2. *Let $e_1, e_2, \ldots$ be independent standard exponentially distributed. Define $g_{jn} = \sum_{i=1}^n e_{j+i}$ for $n, j+1 \in \mathbb{N}$. Then*
(a) g_{jn} is $\Gamma(n,1)$ distributed and $\mathsf{E}|g_{jn} - n|^4 = 3n(n+2) \leq 9n^2$;
(b) $\mathsf{P}(|n^{-1}(g_{jn} - n)| \geq x) \leq 9x^{-4}n^{-2}$;
(c) $\mathsf{P}(\max_{0<j<n_1} |n_0^{-1}(g_{jn_0} - n_0)| > x) \leq 9x^{-4}n_1 n_0^{-2}$.

Proof. (a) see (Johnson et al., 1994, §17.6, equation 17.10).

(b) By the Markov inequality, $\mathsf{P}(|n^{-1}(g_{jn}-n)| \geq x) \leq (nx)^{-4}\mathsf{E}|g_{jn}-n|^4$. Apply (a).

(c) Let $z_j = n_0^{-1}(g_{jn_0}-n_0)$ and $\mathcal{P}_n = \mathsf{P}(\max_{0<j<n_1}|z_j| > x)$. Boole's inequality gives $\mathcal{P}_n \leq \sum_{0<j<n_1}\mathsf{P}(|z_j|>x)$. Here, $\mathsf{P}(|z_j|>x) \leq 9x^{-4}n_0^{-2}$ by (b), so that $\mathcal{P}_n \leq 9x^{-4}n_1n_0^{-2}$. □

C.5.2 Fewer 'outliers' than 'good' observations

We show that the minimizer $\hat{\delta}_{LMS}$ is close to $\delta_n = \sum_{j\in\zeta_n} 1_{(\varepsilon_j<\min_{i\in\zeta_n}\varepsilon_i)}$. Due to the argument (B.1), it suffices to analyze the asymptotic behaviour for deterministic sequences δ_n.

Lemma C.3. *Suppose Assumption C.1 holds. Let $\lambda, \rho < 1$. Then, conditional on δ_n, an $\epsilon > 0$ exists, so that* $\min_{1\leq s<h_n} h_n(\hat{\sigma}_{\delta_n+s}-\hat{\sigma}_{\delta_n}) \geq \epsilon + \mathrm{o}_{\mathsf{P}}(1)$.

Proof. Let $S_s = (\hat{\sigma}_{\delta_n+s}-\hat{\sigma}_{\delta_n})/\sigma = \{\varepsilon_{(\delta_n+s+h_n)}-\varepsilon_{(\delta_n+s+1)}\}/2 - \{\varepsilon_{(\delta_n+h_n)}-\varepsilon_{(\delta_n+1)}\}/2$. Reorganize as $S_s = \{\varepsilon_{(\delta_n+h_n+s)}-\varepsilon_{(\delta_n+h_n)}\}/2 - \{\varepsilon_{(\delta_n+s+1)}-\varepsilon_{(\delta_n+1)}\}/2$.

The 'good' errors $\varepsilon_{(\delta_n+s+1)}$ and $\varepsilon_{(\delta_n+1)}$ are order statistics of uniform errors on $[-1,1]$. Thus, $\{\varepsilon_{(\delta_n+1+s)}-\varepsilon_{(\delta_n+1)}\}/2 = u_{(1+s)}-u_{(1)}$ is a standard uniform spacing. The uniform spacings Lemma C.1 shows that there exists independent standard exponential variables e_k where $1\leq k\leq h_n+1$ so that $u_{(1+s)}-u_{(1)} = \sum_{k=2}^{1+s}e_k/\sum_{k=1}^{h_n+1}e_k$.

The 'outliers' satisfy $\varepsilon_{(\delta_n+h_n+s)}-\varepsilon_{(\delta_n+h_n)} = \overline{\varepsilon}_{(s)}$ where $\overline{\varepsilon}_{(s)}$ is positive and an order statistic of the distribution function $\overline{\mathsf{G}}$. By the inverse probability transformation, there exist independent standard uniform variables $\overline{u}_s$, so that $\overline{\varepsilon}_{(s)} = \overline{\mathsf{G}}^{-1}\{\overline{u}_{(s)}\}$. Thus, $\varepsilon_{(\delta_n+h_n+s)}-\varepsilon_{(\delta_n+h_n)} = \overline{\mathsf{G}}^{-1}\{\overline{u}_{(s)}\}$.

We consider the cases $1\leq s<s_n$ and $s_n\leq s<h_n$ separately for some sequence $s_n\to\infty$, but $s_n/n\to 0$. We choose $s_n = n^{(1-\tau)/2}$ for $\tau<1$ defined in Assumption C.1(ii).

The case $s_n\leq s\leq h_n$. For the *'good' observations* bound

$$\{\varepsilon_{(\delta_n+1+s)}-\varepsilon_{(\delta_n+1)}\}/2 = \frac{\sum_{k=2}^{1+s}e_k}{\sum_{k=1}^{h_n+1}e_k} \leq \frac{\sum_{k=2}^{s+1}e_k}{\sum_{k=2}^{h_n+1}e_k} = (\frac{s}{h_n})\frac{1+s^{-1}\sum_{k=2}^{s+1}(e_k-1)}{1+h_n^{-1}\sum_{k=2}^{h_n+1}(e_k-1)}. \tag{C.5}$$

Let $m_n^{\text{large}} = \max_{t\geq s_n}|t^{-1}\sum_{i=2}^{t+1}(e_i-1)|$. By the Law of Large Numbers $x_t = t^{-1}\sum_{i=1}^{t}(e_i-1)\to 0$ $a.s.$ for $t\to\infty$. This implies $m_n^{\text{large}}\to 0$ $a.s.$, since for each outcome we have deterministic sequence $x_t\to 0$, say. But, if $x_t\to 0$, then $\limsup_{t\to\infty}|x_t|\to 0$. In particular, for $s_n\to\infty$ we get $\max_{t\geq s_n}|x_t|\to 0$. In summary, we get

$$\{\varepsilon_{(\delta_n+1+s)}-\varepsilon_{(\delta_n+1)}\}/2 \overset{a.s.}{\leq} (s/h_n)\{1+\mathrm{o}(1)\}. \tag{C.6}$$

The 'outliers'. By Assumption C.1(i), we have $\overline{\mathsf{G}}^{-1}(\psi)\geq 2\psi\varrho$, where $\varrho = (1-\rho+\epsilon)(1-\lambda)/\lambda$. Thus, we get $\{\varepsilon_{(\delta_n+h_n+s)}-\varepsilon_{(\delta_n+h_n)}\}/2 = \overline{\mathsf{G}}^{-1}\{\overline{u}_{(s)}\}/2 \geq \varrho\overline{u}_{(s)}$. The uniform spacings Lemma C.1 shows that there exists independent standard exponential variables $\overline{e}_k$ for $1\leq k\leq\overline{n}+1$ so that $\overline{u}_{(s)} = \sum_{k=1}^{s}\overline{e}_k/\sum_{k=1}^{\overline{n}+1}\overline{e}_k$. Thus, we can bound

$$\{\varepsilon_{(\delta_n+h_n+s)}-\varepsilon_{(\delta_n+h_n)}\}/2 \geq \varrho\frac{\sum_{k=1}^{s}\overline{e}_k}{\sum_{k=1}^{\overline{n}+1}\overline{e}_k} = \varrho(\frac{s}{\overline{n}+1})\frac{1+s^{-1}\sum_{k=1}^{s}(\overline{e}_k-1)}{1+(\overline{n}+1)^{-1}\sum_{k=1}^{\overline{n}+1}(\overline{e}_k-1)}. \tag{C.7}$$

Let $\overline{m}_n^{\text{large}} = \max_{s\geq s_n}|s^{-1}\sum_{j=1}^{s}(\overline{e}_j-1)|$. As before, $\overline{m}_n^{\text{large}} = \mathrm{o}(1)$ $a.s.$ Thus, we get

$$\{\varepsilon_{(\delta_n+h_n+s)}-\varepsilon_{(\delta_n+h_n)}\}/2 \overset{a.s}{\geq} \varrho(\frac{s}{\overline{n}+1})\{1+\mathrm{o}(1)\}.$$

Combine with the bound (C.6) to see that $\min_{s_n\leq s\leq h_n}(\overline{n}+1)S_s \geq s\{\varrho-(\overline{n}+1)/h_n\}\{1+\mathrm{o}(1)\}$ $a.s.$ Since $(\overline{n}+1)/h_n \to \tilde{\rho} = (1-\rho)(1-\lambda)/\lambda$, so that $\varrho - \tilde{\rho} = \epsilon(1-\lambda)\lambda^{-1} > 0$, while $s > s_n$ we get $\min_{s_n\leq s\leq h_n}(\overline{n}+1)S_s \geq s_n\epsilon(1-\lambda)\lambda^{-1}\{1+\mathrm{o}(1)\}$ $a.s.$ which goes to infinity with s_n, while $\overline{n}/h_n \to \tilde{\rho} > 0$.

The case $1 \leq s < s_n = h_n^{(1-\tau)/2}$. For the 'good' observations bound

$$\{\varepsilon_{(\delta_n+1+s)} - \varepsilon_{(\delta_n+1)}\}/2 = \frac{\sum_{k=2}^{1+s} e_k}{\sum_{k=1}^{h_n+1} e_k} \leq \frac{\sum_{k=2}^{1+s_n} e_k}{\sum_{k=2}^{h_n+1} e_k} = (\frac{s_n}{h_n})\frac{1+s_n^{-1}\sum_{k=2}^{s_n+1}(e_k-1)}{1+h_n^{-1}\sum_{k=2}^{h_n+1}(e_k-1)}.$$

By the strong Law of Large Numbers we get that the averages in the numerator and denominator vanish, so that $\{\varepsilon_{(\delta_n+1+s)} - \varepsilon_{(\delta_n+1)}\}/2 \leq (s_n/h_n)\{1+\mathrm{o}(1)\}$ $a.s.$

For the 'outliers', Assumption C.1(ii) is $\overline{\mathsf{G}}^{-1}(\psi) \geq \psi^\tau$ for some $\tau < 1$ and $\psi < \psi_0$. Thus, we get $\varepsilon_{(\delta_n+h_n+s)} - \varepsilon_{(\delta_n+h_n)} = \overline{\mathsf{G}}^{-1}\{\overline{u}_{(s)}\} \geq \{\overline{u}_{(s)}\}^\tau$. Further, $\overline{u}_{(s)} \geq \overline{u}_{(1)}$. As before, $\overline{u}_{(1)} = \overline{e}_1/\sum_{k=1}^{\overline{n}+1}\overline{e}_k$. Thus, we can replace (C.7) with

$$\varepsilon_{(\delta_n+h_n+s)} - \varepsilon_{(\delta_n+h_n)} = \overline{\mathsf{G}}^{-1}\{\overline{u}_{(s)}\} \geq \{\overline{u}_{(1)}\}^\tau = (\frac{\overline{e}_1}{\sum_{k=1}^{\overline{n}+1}\overline{e}_k})^\tau \overset{a.s.}{=} (\frac{\overline{e}_1}{\overline{n}+1})^\tau\{1+\mathrm{o}(1)\},$$

by the Strong Law of Large Numbers. Since $\overline{e}_1$ is exponential, then for all $\epsilon > 0$ exists a $\eta > 0$ so that $\mathsf{P}(\overline{e}_1 > 2\eta) \geq 1-\epsilon$. As before, $\overline{n}/h_n \to \tilde{\rho} > 0$. In combination, we get

$$\min_{1\leq s\leq s_n} h_n S_s \geq [(h_n/2)\{\eta/(\tilde{\rho}h_n)\}^\tau - h_n(s_n/h_n)]\{1+\mathrm{o}_{\mathsf{P}}(1)\}.$$

Thus, for some constant $C > 0$, we get $\min_{1\leq s\leq s_n} h_n S_s \geq (Ch_n^{1-\tau} - s_n)\{1+\mathrm{o}_{\mathsf{P}}(1)\}$. This diverges since $s_n = h_n^{(1-\tau)/2} = \mathrm{o}(h_n^{1-\tau})$ when $1-\tau > 0$. □

Proof of Theorem C.2. We proceed as in the proof of Theorem 5.2, conditioning on sequences δ_n satisfying $\delta_n/(n-h_n) \to \rho$, and only considering $\hat{s} = \hat{\delta}_{LMS} - \delta_n > 0$. First, it suffices to show that $n(\hat{\sigma}^2_{\delta_n+s} - \hat{\sigma}^2_{\delta_n}) > \epsilon + \mathrm{o}_{\mathsf{P}}(1)$ for some $\epsilon > 0$, uniformly in $1 \leq s < h_n$. This was proved in Lemma C.3 using Assumption C.1.

Second, since $\mathsf{P}(\hat{\delta} = \delta_n) \to 1$, we get $\hat{\mu}_{LMS} = \hat{\mu}_{\delta_n}, \hat{\sigma}_{LMS} = \hat{\sigma}_{\delta_n}$ with large probability.

Third, we analyze $\hat{\mu}_{\delta_n}$ and $\hat{\sigma}_{\delta_n}$. Since ε_i is uniform on $[-1,1]$ then $u_i = (\varepsilon_i+1)/2$ is uniform on $[0,1]$. The spacings Lemma C.1 shows that independent standard exponential variables e_i exist, so that $u_{(\delta_n+1)} = e_1/\sum_{k=1}^{h_n+1} e_k$ and $1 - u_{(\delta_n+h_n)} = e_{h_n+1}/\sum_{k=1}^{h_n+1} e_k$. In particular,

$$(\hat{\sigma}_{\delta_n} - \sigma)/\sigma = \{\varepsilon_{(\delta_n+h_n)} - \varepsilon_{(\delta_n+1)} - 2\}/2 = u_{(\delta_n+h_n)} - u_{(\delta_n+1)} - 1 = -(e_1 + e_{h_n+1})/\textstyle\sum_{k=1}^{h_n+1}e_k,$$
$$(\hat{\mu}_{\delta_n} - \mu)/\sigma = \{\varepsilon_{(\delta_n+h_n)} + \varepsilon_{(\delta_n+1)}\}/2 = u_{(\delta_n+h_n)} + u_{(\delta_n+1)} - 1 = (e_1 - e_{h_n+1})/\textstyle\sum_{k=1}^{h_n+1}e_k.$$

By the Law of Large Numbers, $(h_n+1)^{-1}\sum_{i=1}^{h_n+1} e_i = 1+\mathrm{o}_{\mathsf{P}}(1)$. Let $\underline{e} = e_1$ and $\overline{e} = e_{h_n+1}$. □

C.5.3 The OLS estimator in the LMS model

Proof of Theorem C.3. Proceed as in the proof of Theorem 5.4, apart from a different analysis of the order statistic $\varepsilon_{(h_n)}$. Specifically, for $1 \leq i \leq h_n$ then ε_i is uniform on $[-1,1]$, so that $\varepsilon_{(h_n)} = 1 + \mathrm{o}_{\mathsf{P}}(1)$. Thus, the two last terms in (B.8) have the same order here. □

D Identities in LTS proofs

The formula (B.3).

We expand $S_s = (\hat{\sigma}^2_{\delta_n+s} - \hat{\sigma}^2_{\delta_n})/\sigma^2$ when $0 < s < h_n$. By definition

$$S_s = h_n^{-1} \sum_{i=1}^{h_n} \varepsilon^2_{(\delta_n+s+i)} - \{h_n^{-1} \sum_{i=1}^{h_n} \varepsilon_{(\delta_n+s+i)}\}^2 - h_n^{-1} \sum_{i=1}^{h_n} \varepsilon^2_{(\delta_n+i)} + \{h_n^{-1} \sum_{i=1}^{h_n} \varepsilon_{(\delta_n+i)}\}^2.$$

We have that $s < \delta_n < s + h_n$. We can then divide the h_n errors $\{\varepsilon_{(\delta_n+s+1)}, \ldots, \varepsilon_{(\delta_n+s+h_n)}\}$ into

$$\{\varepsilon_{(\delta_n+s+1)}, \ldots, \varepsilon_{(\delta_n+h_n)}\} \qquad \text{and} \qquad \{\varepsilon_{(\delta_n+h_n+1)}, \ldots, \varepsilon_{(\delta_n+s+h_n)}\}.$$

The first group are order statistics of 'good' errors. The second group consists of 'outliers' for which $\varepsilon_{(\delta_n+h_n+j)} = \varepsilon_{(\delta_n+h_n)} + \overline{\varepsilon}_{(j)}$ for $1 \le j \le s$. Thus, for the second moment we have

$$\begin{aligned} \sum_{i=1}^{h_n} \varepsilon^2_{(\delta_n+s+i)} &= \sum_{i=s+1}^{h_n} \varepsilon^2_{(\delta_n+i)} + \sum_{j=1}^{s} \{\varepsilon_{(\delta_n+h_n)} + \overline{\varepsilon}_{(j)}\}^2 \\ &= \sum_{i=s+1}^{h_n} \varepsilon^2_{(\delta_n+i)} + s\varepsilon^2_{(\delta_n+h_n)} + 2\varepsilon_{(\delta_n+h_n)} \sum_{j=1}^{s} \overline{\varepsilon}_{(j)} + \sum_{j=1}^{s} \overline{\varepsilon}^2_{(j)}. \end{aligned}$$

For the squared first moment we have

$$\begin{aligned} \{\sum_{i=1}^{h_n} \varepsilon_{(\delta_n+s+i)}\}^2 &= [\sum_{i=s+1}^{h_n} \varepsilon_{(\delta_n+i)} + \sum_{j=1}^{s} \{\varepsilon_{(\delta_n+h_n)} + \overline{\varepsilon}_{(j)}\}]^2 \\ &= \{\sum_{i=s+1}^{h_n} \varepsilon_{(\delta_n+i)}\}^2 + s^2 \varepsilon^2_{(\delta_n+h_n)} + \{\sum_{j=1}^{s} \overline{\varepsilon}_{(j)}\}^2 \\ &\quad + 2s\varepsilon_{(\delta_n+h_n)} \sum_{i=s+1}^{h_n} \varepsilon_{(\delta_n+i)} + 2s\varepsilon_{(\delta_n+h_n)} \sum_{j=1}^{s} \overline{\varepsilon}_{(j)} + 2\{\sum_{i=s+1}^{h_n} \varepsilon_{(\delta_n+i)}\}\{\sum_{j=1}^{s} \overline{\varepsilon}_{(j)}\}. \end{aligned}$$

Further, we can expand

$$\begin{aligned} \sum_{i=1}^{h_n} \varepsilon^2_{(\delta_n+i)} &= \sum_{i=1}^{s} \varepsilon^2_{(\delta_n+i)} + \sum_{i=s+1}^{h_n} \varepsilon^2_{(\delta_n+i)}, \\ \{\sum_{i=1}^{h_n} \varepsilon_{(\delta_n+i)}\}^2 &= \{\sum_{i=1}^{s} \varepsilon_{(\delta_n+i)}\}^2 + \{\sum_{i=s+1}^{h_n} \varepsilon_{(\delta_n+i)}\}^2 + 2\sum_{i=1}^{s} \varepsilon_{(\delta_n+i)} \sum_{i=s+1}^{h_n} \varepsilon_{(\delta_n+i)}. \end{aligned}$$

Inserting the expansions of the moments in the expression for S_s gives

$$\begin{aligned}
S_s = & \frac{1}{h_n}\{\sum_{i=s+1}^{h_n} \varepsilon^2_{(\delta_n+i)} + s\varepsilon^2_{(\delta_n+h_n)} + 2\varepsilon_{(\delta_n+h_n)}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)} + \sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)}\} \\
& - \frac{1}{h_n^2}[\{\sum_{i=s+1}^{h_n} \varepsilon_{(\delta_n+i)}\}^2 + s^2\varepsilon^2_{(\delta_n+h_n)} + \{\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}^2 \\
& \qquad + 2s\varepsilon_{(\delta_n+h_n)}\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)} + 2s\varepsilon_{(\delta_n+h_n)}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)} + 2\{\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}\}\{\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}] \\
& - \frac{1}{h_n}\{\sum_{i=1}^{s}\varepsilon^2_{(\delta_n+i)} + \sum_{i=s+1}^{h_n}\varepsilon^2_{(\delta_n+i)}\} \\
& + \frac{1}{h_n^2}[\{\sum_{i=1}^{s}\varepsilon_{(\delta_n+i)}\}^2 + \{\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}\}^2 + 2\sum_{i=1}^{s}\varepsilon_{(\delta_n+i)}\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}]
\end{aligned}$$

This reduces as

$$S_s = \frac{s}{h_n}(1-\frac{s}{h_n})\varepsilon^2_{(\delta_n+h_n)} + A_n$$

where

$$\begin{aligned}
A_n = & \frac{1}{h_n}\{\sum_{i=s+1}^{h_n} \varepsilon^2_{(\delta_n+i)} + 2\varepsilon_{(\delta_n+h_n)}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)} + \sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)}\} \\
& - \frac{1}{h_n^2}[\{\sum_{i=s+1}^{h_n} \varepsilon_{(\delta_n+i)}\}^2 + \{\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}^2 \\
& \qquad + 2s\varepsilon_{(\delta_n+h_n)}\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)} + 2s\varepsilon_{(\delta_n+h_n)}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)} + 2\{\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}\}\{\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}] \\
& - \frac{1}{h_n}\{\sum_{i=1}^{s}\varepsilon^2_{(\delta_n+i)} + \sum_{i=s+1}^{h_n}\varepsilon^2_{(\delta_n+i)}\} \\
& + \frac{1}{h_n^2}[\{\sum_{i=1}^{s}\varepsilon_{(\delta_n+i)}\}^2 + \{\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}\}^2 + 2\sum_{i=1}^{s}\varepsilon_{(\delta_n+i)}\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}]
\end{aligned}$$

There are two cancellations: term 1 in line 1 with term 2 in line 4 and term 1 in line 2 with term 2 in line 5. Thus, A_n reduces to

$$\begin{aligned}
A_n = & 2\varepsilon_{(\delta_n+h_n)}\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)} + \frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)} - \{\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}^2 \\
& - 2\frac{s}{h_n}\varepsilon_{(\delta_n+h_n)}\frac{1}{h_n}\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)} - 2\frac{s}{h_n}\varepsilon_{(\delta_n+h_n)}\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)} \\
& - 2\{\frac{1}{h_n}\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}\}\{\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\} - \frac{1}{h_n}\sum_{i=1}^{s}\varepsilon^2_{(\delta_n+i)} \\
& + \{\frac{1}{h_n}\sum_{i=1}^{s}\varepsilon_{(\delta_n+i)}\}^2 + 2\frac{1}{h_n}\sum_{i=1}^{s}\varepsilon_{(\delta_n+i)}\frac{1}{h_n}\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}.
\end{aligned}$$

Rearrange as

$$A_n = [\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)} - \{\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}^2] - [\frac{1}{h_n}\sum_{i=1}^{s}\varepsilon^2_{(\delta_n+i)} - \{\frac{1}{h_n}\sum_{i=1}^{s}\varepsilon_{(\delta_n+i)}\}^2]$$
$$+ 2\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\{(1-\frac{s}{h_n})\varepsilon_{(\delta_n+h_n)} - \frac{1}{h_n}\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}\}$$
$$- 2\frac{1}{h_n}\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}\{\frac{s}{h_n}\varepsilon_{(\delta_n+h_n)} - \frac{1}{h_n}\sum_{i=1}^{s}\varepsilon_{(\delta_n+i)}\}.$$

The terms in second and in third line, respectively, can be simplified to give

$$A_n = [\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)} - \{\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}^2] - [\frac{1}{h_n}\sum_{i=1}^{s}\varepsilon^2_{(\delta_n+i)} - \{\frac{1}{h_n}\sum_{i=1}^{s}\varepsilon_{(\delta_n+i)}\}^2]$$
$$+ 2\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\frac{1}{h_n}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+h_n)} - \varepsilon_{(\delta_n+i)}\}$$
$$- 2\frac{1}{h_n}\sum_{i=s+1}^{h_n}\varepsilon_{(\delta_n+i)}\frac{1}{h_n}\sum_{i=1}^{s}\{\varepsilon_{(\delta_n+h_n)} - \varepsilon_{(\delta_n+i)}\}.$$

which has the desired form $A_n = A_{n1} - A_{n2} + 2A_{n3} - 2A_{n4}$ □

The formula (B.7).

We have $h_n - \underline{s}_n \leq s < h_n$ where $\overline{s}_n = (2\log h_n)^{-1/4}h_n$. By definition

$$\hat{\sigma}^2_{\delta_n+s}/\sigma^2 = \frac{1}{h_n}\sum_{i=1}^{h_n}\varepsilon^2_{(\delta_n+s+i)} - \{\frac{1}{h_n}\sum_{i=1}^{h_n}\varepsilon_{(\delta_n+s+i)}\}^2.$$

A residual sums of squares is invariant to subtracting a constant from each observation. Thus, subtracting $\varepsilon_{(\delta_n+h_n)}$ from each $\varepsilon_{(\delta_n+s+i)}$ gives

$$\hat{\sigma}^2_{\delta_n+s}/\sigma^2 = \frac{1}{h_n}\sum_{i=1}^{h_n}\{\varepsilon_{(\delta_n+s+i)} - \varepsilon_{(\delta_n+h_n)}\}^2 - [\frac{1}{h_n}\sum_{i=1}^{h_n}\{\varepsilon_{(\delta_n+s+i)} - \varepsilon_{(\delta_n+h_n)}\}]^2.$$

Split into 'good' and 'outlier' errors to get

$$\hat{\sigma}^2_{\delta_n+s}/\sigma^2 = \frac{1}{h_n}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+i)} - \varepsilon_{(\delta_n+h_n)}\}^2 + \frac{1}{h_n}\sum_{j=1}^{s}\{\varepsilon_{(\delta_n+h_n+j)} - \varepsilon_{(\delta_n+h_n)}\}^2$$
$$- [\frac{1}{h_n}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+i)} - \varepsilon_{(\delta_n+h_n)}\} + \frac{1}{h_n}\sum_{j=1}^{s}\{\varepsilon_{(\delta_n+h_n+j)} - \varepsilon_{(\delta_n+h_n)}\}]^2.$$

Note that $\varepsilon_{(\delta_n+h_n+j)} - \varepsilon_{(\delta_n+h_n)} = \overline{\varepsilon}_{(j)}$ while $\varepsilon_{(\delta_n+i)} < \varepsilon_{(\delta_n+h_n)}$ so that

$$\hat{\sigma}^2_{\delta_n+s}/\sigma^2 = \frac{1}{h_n}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+h_n)} - \varepsilon_{(\delta_n+i)}\}^2 + \frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)}$$
$$- [-\frac{1}{h_n}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+h_n)} - \varepsilon_{(\delta_n+i)}\} + \frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}]^2.$$

Rearrange as

$$\begin{aligned}\hat{\sigma}^2_{\delta_n+s}/\sigma^2 &= \frac{1}{h_n}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+h_n)} - \varepsilon_{(\delta_n+i)}\}^2 - [\frac{1}{h_n}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+h_n)} - \varepsilon_{(\delta_n+i)}\}]^2 \\ &+ \frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)} - \{\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}^2 \\ &+ 2[\frac{1}{h_n}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+h_n)} - \varepsilon_{(\delta_n+i)}\}]\{\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}.\end{aligned}$$

The term in the second line satisfies

$$\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)} - \{\frac{1}{h_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}^2 = \frac{s}{h_n}(1-\frac{s}{h_n})\frac{1}{s}\sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)} + (\frac{s}{h_n})^2[\frac{1}{s}\sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)} - \{\frac{1}{s_n}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}^2].$$

Thus, we get

$$\hat{\sigma}^2_{\delta_n+s}/\sigma^2 = A_n = A_{n1} + A_{n2} + A_{n3} + 2A_{n4},$$

which is (B.7), where

$$\begin{aligned}A_{n1} &= h_n^{-1}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+i)} - \varepsilon_{(\delta_n+h_n)}\}^2 - [h_n^{-1}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+i)} - \varepsilon_{(\delta_n+h_n)}\}]^2, \\ A_{n2} &= (\frac{s}{h_n})^2[\frac{1}{s}\sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)} - \{\frac{1}{s}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}^2], \\ A_{n3} &= \frac{s}{h_n}(1-\frac{s}{h_n})\frac{1}{s}\sum_{j=1}^{s}\overline{\varepsilon}^2_{(j)} \\ A_{n4} &= [h_n^{-1}\sum_{i=s+1}^{h_n}\{\varepsilon_{(\delta_n+h_n)} - \varepsilon_{(\delta_n+i)}\}]\{h_n^{-1}\sum_{j=1}^{s}\overline{\varepsilon}_{(j)}\}.\end{aligned}$$

This completes the proof of (B.7). □

References

Berenguer-Rico, V., Johansen, S., and Nielsen, B. (2019). Models where the Least Trimmed Squares and Least Median of Squares estimators are maximum likelihood. Discussion Paper 879, Dept. Econ., Oxford.

Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.*, 42:1887–1896.

Harter, H. L. (1953). Maximum likelihood regression equations (abstract). *Ann. Math. Statist.*, 24:687.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous univariate distributions*, volume 1. John Wiley & Sons, New York, 2nd edition.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.*, 18:191–219.

Knight, K. (2017). On the asymptotic distribution of the l_∞ estimator in linear regression. Mimeo.

Pyke, R. (1965). Spacings (with discussion). *J. Roy. Statist. Soc. Ser. B*, 27:395–449.

Rousseeuw, P. J. (1984). Least median of squares regressions. *J. Amer. Statist. Assoc.*, 79:871–880.

Schechtman, E. and Schechtman, G. (1986). Estimating the parameters in regression with uniformly distributed errors. *J. Statist. Comput. Simul.*, 26:269–281.

Wagner, H. M. (1959). Linear programming techniques for regression analysis. *J. Amer. Statist. Assoc.*, 54:206–212.