International Journal of Statistics and Management System
Vol. 5 No. 1 (January-June, 2020)

Submitted: 12th January 2020 *Revised: 07th April 2020* *Accepted: 19th May 2020*

On an *L*-estimator with data-dependent coefficients *

Yijun Zuo †and Du Juan ‡

Abstract

A classical L-estimator is a linear combination of order statistics with *constant* coefficients. This paper studies an L-estimator which has *data-dependent* coefficients. The paper focuses on the efficiency behavior of the estimator and addresses the robustness and the asymptotics issues as well. It turns out that the random-coefficient estimator enjoys a remarkably high *absolute efficiency* relative to the most efficient estimators at a variety of light and heavy tailed models while sharing the best breakdown point robustness of the univariate median. Findings in the paper suggest that the random-coefficient L-estimator can serve very well as a location estimator and an alternative to both the median and the mean.

1. Introduction Classical L-estimators of location parameters are defined to be linear combinations of order statistics with *constant* coefficients; see, e.g., Serfling (1980). They compete well against M- and R- estimators from both the efficiency and the robustness view points. Indeed, the mean, an L-estimator, is most efficient at normal models whereas the median, another L-estimator, is most robust. Both the mean and the median, however, bear some shortcomings. For example, the median is just 64% efficient relative to the

*Research partially supported by NSF Grant DMS–0134628.

Key words and phrases: L-estimator, random coefficient, efficiency, robustness, order statistics.
AMS 2000 subject classifications. Primary 62G05; secondary 62H12, 62G35.

†*Mailing Address:* Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824. *E–mail:* zuo@msu.edu.

‡*Mailing Address:* Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824. *E–mail:* dujuan@msu.edu.

mean at normal models while the mean has the worst breakdown-point robustness (see Section 3). The mean, albeit most efficient at normal models, is just 50% efficient relative to the most efficient estimator, the median, at double exponential models. All these raise a natural question as to whether there are estimators that can share the best robustness of the median while possessing a universally high efficiency relative to the most efficient estimators at different models.

The *random*-coefficient L-estimator studied in this paper provides a positive answer to the question. The estimator has data-dependent coefficients. The idea of data-dependent weighting dates back to Tukey (1970) or earlier. The estimator here has strong connections to but different from, M-, Tukey W- and other existing estimators (see Section 6). Furthermore, it has never been singled out and subjected to a careful scrutiny in the literature.

After defining the estimator in Section 2, the paper shows in Section 3 that it shares the best breakdown point of the median and is locally robust with a bounded influence function. Asymptotic representation and limiting distribution of the estimator are presented in Section 4. The large and finite sample efficiencies of the estimator relative to the median, the mean, and the best estimators at a variety of light and heavy tailed symmetric and asymmetric models with different weighting schemes are carefully examined in Section 5. With appropriately selected weight functions, the estimator turns out to be overwhelmingly more efficient than the median and the mean, except at the double exponential and the normal models where the latter two are respectively most efficient. On the other hand, the estimator enjoys remarkably high (about 90% or higher) efficiency relative to the most efficient estimators universally across a variety of distributions range from very light to very heavy tailed ones. The main body of the paper ends in Section 6 with discussions and concluding remarks. Proofs of main results are reserved for the Appendix.

2. The random-coefficient L-estimator Consider location and scale models and let μ and σ be some (initial) estimators of location and scale, respectively. Choices of (μ, σ) include (the mean, the standard deviation) and (the median, the median absolute deviation). Write μ_n and σ_n for $\mu(X)$ and $\sigma(X)$ at a random sample $X = \{X_1, \cdots, X_n\}$. Denote $|x - \mu_n|/\sigma_n$, the scaled deviation of x to the center $\mu(X)$ of the data set, by $d(x, X)$. When $x - \mu_n = \sigma_n = 0$, we defines $d(x, X) = 0$. The *random-coefficient L-estimator*, L_n,

a scaled deviation weighted mean, is defined as

$$L_n := \sum_{i=1}^{n} a_{ni} X_{ni} := \sum_{i=1}^{n} w(d(X_i, X)) X_i \Big/ \sum_{i=1}^{n} w(d(X_i, X)) \tag{2.1}$$

where $X_{n1} \leq \cdots \leq X_{nn}$ are ordered values of $X_1, \cdots, X_n$, the random coefficient $a_{ni} = w(d(X_{ni}, X))/\sum_{i=1}^{n} w(d(X_{ni}, X))$, and weight $w(r) > 0$ is bounded and even on $[-\infty, \infty]$.

Throughout the paper let μ and σ be *affine equivariant*: $\mu(aX + b) = a\mu(X) + b$ and $\sigma(aX + b) = |a|\sigma(X)$ for any $a, b \in R^1$, where $aX + b = \{aX_1 + b, \cdots, aX_n + b\}$. Roughly speaking, this means that μ and σ do not depend on the underlying coordinate system and measurement scale. Common choices of μ and σ are affine equivariant. This affine equivariance implies that $d(x, X)$ is *affine invariant*, that is, $d(ax + b, aX + b) = d(x, X)$. Consequently L_n is also affine equivariant. If $X_i \sim F$ is *symmetric* about a point $\theta \in \mathbb{R}^1$ (i.e. $\pm(X_i - \theta)$ have the same distribution), then so is L_n. Furthermore, if $E(X_i)$ exists, then L_n is *unbiased* for θ. Let F_n be the empirical version of F based on X and write X and F_n interchangeably, then $L_n = L(F_n) = L(X)$ with the functional $L(\cdot)$ defined as follows:

$$L(F) = \int x w((x - \mu(F))/\sigma(F)) dF(x) \Big/ \int w((x - \mu(F))/\sigma(F)) dF(x). \tag{2.2}$$

Assume in the following the functionals $\mu(\cdot)$ and $\sigma(\cdot)$ are *affine equivariant*, then so is $L(\cdot)$. Further, $L(\cdot)$ is *Fisher consistent* in the sense that $L(F) = \theta$ if F is symmetric about θ.

3. Robustness Robustness is a fundamental issue in statistics. Any statistical procedures are desired to be robust. This section studies the robustness aspects of $L(\cdot)$. At the sample level, we investigate its *global* robustness in term of the finite sample breakdown point. At the population level, we examine its *local* robustness in term of the influence function.

3.1. Finite sample breakdown point

This notion, introduced in Donoho and Huber (1983), has become the most popular quantitative measure of the global robustness of an estimator. Roughly speaking, the *finite sample breakdown point* of a location estimator is the minimal fraction of "bad points" (or

contaminated points) in a data set that can render the estimator unbounded. More precisely, the replacement breakdown point (RBP) of an estimator T_n at $X = \{X_1, \cdots, X_n\}$ is defined as

$$\text{RBP}(T_n, X) = \min\{\frac{m}{n} : \sup_{X^m} \|T(X^m) - T(X)\| = \infty\}, \tag{3.1}$$

where X^m denotes a contaminated sample resulting from replacing m points of X with arbitrary values. Clearly the sample mean has the lowest breakdown point $1/n$ whereas the median (Med) can be seen to possess the highest one, $\lfloor (n+1)/2 \rfloor / n$, among all affine equivariant estimators. Here $\lfloor x \rfloor$ denotes the largest number no larger than x.

For a scale estimator $S(\cdot)$, we can define its breakdown point with the same definition but with $T(\cdot)$ on the right side replaced by $\log(S(\cdot))$. That is, we say that S_n breaks down if it vanishes or gets unbounded. If all the data points coincide, then any reasonable S_n will be 0 (hence breaks down). To avoid trivial cases like this, we will consider the random sample X that is *in general position*, namely, data points of X are distinct with each other. For X in general position, the standard deviation has the lowest breakdown point $1/n$ whereas the median absolute deviation (MAD): MAD(X)=Med$\{|X_i - \text{Med}\{X_i\}|\}$ is readily seen to have the highest one, $\lfloor n/2 \rfloor / n$, among all affine equivariant estimators. It turns out that L_n can also have the best breakdown point with Med and MAD for μ and σ, respectively.

Theorem 3.1 *Let* $\mu = \text{Med}$ *and* $\sigma = \text{MAD}$. *Let* $w(r)$ *and* $rw(r)$ *be bounded on* $[0, \infty]$ *and* $\inf_{0 \le r \le 1} w(r) > 0$. *Then for any* X *in general position,* $\text{RBP}\,(L_n, X) = \lfloor (n+1)/2 \rfloor / n$.

From the proof we see that the theorem remains valid if Med and MAD is replaced with any μ and σ that have the same breakdown points as those of Med and MAD respectively. The restrictions on w are quite mild and examples of such w are given in later sections.

The random-coefficient L-estimator shares with the median the best breakdown point. As a *global* robustness measure, the breakdown point alone, however, does not depict the *entire* picture of the robustness of the estimator. The influence function of the estimator provided in the next section can complement the picture and fill the *local* robustness gap.

3.2. Influence function and gross error sensitivity

Denote by δ_x the point mass probability distribution at a fixed point $x \in \mathbb{R}^1$. For a given distribution F and an $\epsilon > 0$, the distribution resulting from contaminating F with an ϵ

amount of δ_x is denoted by $F(\epsilon, \delta_x) = (1-\epsilon)F + \epsilon\,\delta_x$. The *influence function* of a statistical functional T at a given point $x \in \mathbb{R}^1$ for a given F is defined as [see Hampel et al. (1986)]

$$IF(x; T(F)) = \lim_{\epsilon \to 0^+} \frac{T(F(\epsilon, \delta_x)) - T(F)}{\epsilon}. \tag{3.2}$$

This function describes the relative effect (influence) on T of an infinitesimal point-mass contamination at x, capturing the local robustness of T. A functional with a bounded influence function is therefore robust and desirable. The supremum of $|IF(x; T(F))|$ is called the *gross error sensitivity* (GES) of T at F [see Hampel et al. (1986)]. That is

$$GES(T(F)) = \sup_{x \in \mathbb{R}^d} |IF(x; T(F))|, \tag{3.3}$$

which is the maximum relative effect on T of an infinitesimal point-mass contamination and measures the local (and the global in some sense as well) robustness of T.

The functional $L(\cdot)$ turns out to possess a bounded influence function for appropriate μ and σ as shown in the following result. Hence $L(\cdot)$ is locally robust.

Theorem 3.2 *Let $IF(x; \mu(F))$ and $IF(x; \sigma(F))$ exist for $x \in \mathbb{R}^1$, $w(r)$ be continuously differentiable on $[0, \infty]$, and $\int |w'(d(y,F))|(d(y,F))^i dF(y) < \infty$ $(i = 1, 2)$. Then*

$$IF(x; L(F)) = \frac{\int w'(d(y,F))(y - L(F))IF(x; d(y,F))dF(y) + (x - L(F))w(d(x,F))}{\int w(d(y,F))dF(y)}, \tag{3.4}$$

where

$$IF(x; d(y,F)) = \frac{I_{y=\mu(F)}|IF(x;\mu(F))| - \text{sign}(y - \mu(F))IF(x;\mu(F)) - d(y,F)IF(x;\sigma(F))}{\sigma(F)}, \tag{3.5}$$

and the gross error sensitivity of $L(F)$, $\sup_x |IF(x; L(F))|$, is finite as long as those of $\mu(F)$ and $\sigma(F)$ are finite and $rw(r)$ is bounded on $[0, \infty]$.

Location and scale measures that possess bounded influence functions include a large class of M-functionals; see Huber (1981). The Med and MAD are the special cases in the class. The restrictions on w are mild and examples of such w are given in later sections.

When F is symmetric about a point θ, the influence function in the theorem simplifies. Since $L(F)$ is affine equivariant, we can assume without loss of generality that $\theta = 0$.

Corollary 3.3 *If F is symmetric about 0, then under the conditions of Theorem 3.2,*

$$IF(x; L(F)) = \Big(\int -w'(|y|)|y| IF(x; \mu(F)) dF_0(y) + xw(|x|/\sigma(F))\Big) \Big/ \int w(|y|) dF_0(y), \tag{3.6}$$

where $F_0(y) = F(\sigma(F)y)$. The gross error sensitivity of $L(F)$ is finite as long as that of $\mu(F)$ is and $rw(r)$ is bounded on $[0, \infty]$. When $\mu = \text{Med}$, $f(0) = F'(0) > 0$, then

$$GES(L(F)) = \Big(\int -w'(|y|)|y| dF_0(y)/(2f(0)) + \sigma(F) \sup_{r>0} rw(r)\Big) \Big/ \int w(|y|) dF_0(y), \tag{3.7}$$

if the numerator is positive, otherwise

$$GES(L(F)) = \int w'(|y|)|y| dF_0(y)/(2f(0)) \Big/ \int w(|y|) dF_0(y), \tag{3.8}$$

The corollary indicates that the influence function of $L(F)$ does not depend on that of the scale functional as long as F is symmetric about a point $\theta \in \mathbb{R}^1$. A sufficient condition for the display (3.7) to hold is that $w(r)$ is non-increasing on $[0, \infty)$. This condition, of course, will fail the display (3.8). That is, for (3.8) $w(r)$ must increase for r in some part of $[0, \infty)$.

Both the median and the L functionals (with appropriate choices of μ and σ) share the best global robustness in terms of their breakdown point. Both are also locally robust in terms of their bounded gross error sensitivities. In light of these, they are *equally robust*. A possible more detailed comparison between the two, of course, is to see how large their gross error sensitivities are. The median has a very moderate gross error sensitivity $1/(2f(0))$ for F symmetric about 0 with $f(0) = F'(0) > 0$. An immediate question is can the $GES(L(F))$ in Corollary 3.3 be smaller than $GES(\text{Med}(F))$? From the $GES(L(F))$ results in Corollary 3.3, this seems possible for suitable w and F. For example, when the display (3.8) is true, this holds if $\int w'(|r|)|r| dF_0(r) < \int w(|r|) dF_0(r)$ for appropriate w and F. Providing some examples of such w and F seems non-trivial though. On the other hand, when w is non-increasing and f is unimodal, the answer to the above question is negative.

Theorem 3.4 *Assume that the density $f(x)$ of F is even and non-increasing on $[0, \infty)$ and that $w(r)$ is non-increasing on $[0, \infty]$. Let $\mu = \text{Med}$. Then $GES(L(F)) \geq GES(\text{Med}(F))$.*

The theorem indicates that the gross error sensitivity of the median can not be improved if (i) the median itself is employed to define the L functional, (ii) the weight function in L is non-increasing, and (iii) the distribution F is symmetric with a non-increasing density.

Now a natural question raised is: how larger can the gross sensitivity of L be, compared to that of the median under the setting of Theorem 3.4? To partly answer the question, we consider, for simplicity, the likelihood weighting scheme case. That is, $w(r)$ is the Lebsegue density $f_0(r)$ of $F_0(r)$. Under the setting of Theorem 3.4, we observe that

$$\int -w'(|y|)|y|dF_0(y) \;=\; 2\int_0^\infty -f_0'(y)yf_0(y)dy \;=\; \int_0^\infty f_0^2(y)dy \;=\; \frac{1}{2}\int w(|y|)dF_0(y).$$

Let $N(f) = \sup_{r>0} rf(r)$ and $D(f) = \int_0^\infty f^2(r)dr$. Then it is readily seen that

$$GES(L(F)) = (1/(2f(0)) + N(f)/D(f))/2. \tag{3.9}$$

It is interesting to note that the scale measure σ plays no role in $GES(L(F))$ in this setting.

The distribution of greatest interest is, of course, the normal. We have $D(f) = 1/(4\sqrt{\pi})$ and $N(f) = 1/\sqrt{2e\pi}$. Thus $GES(\text{Med}(F)) = \sqrt{\pi/2} < \sqrt{\pi/8} + \sqrt{2/e}) = GES(L(F)$.

For a slightly heavier tailed one, the logistic LG$(0,1)$ with p.d.f. $1/(e^{x/2} + e^{-x/2})^2$, $D(f) = 1/12$ and $N(f) = 0.22387$. Thus $GES(\text{Med (LG(F)))} = 2 < 2.34323 = GES(L(F))$.

For the double exponential DE$(0,1)$ with p.d.f. $e^{-|x|}/2$, a yet more heavier tailed one, $D(f) = 1/8$ and $N(f) = 1/(2e)$. Thus $GES(\text{Med (DE(F)))} = 1 < 1/2+2/e = GES(L(F))$.

Finally for the most heavy tailed one, the Cauchy CAU$(0,1)$ with p.d.f. $1/(\pi(1+x^2))$, $D(f) = 1/(4\pi)$ and $N(f) = 1/(2\pi)$. $GES(\text{Med (DE(F)))} = \pi/2 < \pi/4 + 1 = GES(L(F))$.

We now list the above results in the following table.

Table 1. The gross error sensitivity of the median and L functionals

distribution F	$N(0,1)$	LG$(0,1)$	DE$(0,1)$	CAU$(0,1)$
$GES(\text{Med}(F))$	1.2533	2.0000	1.0000	1.5708
$GES(L(F))$	1.4844	2.3432	1.2358	1.7854

Remark 3.5 (i) The table indicates that $\sup_x |IF(x; \text{Med}(F))|$ is (slightly) smaller than that of $L(F)$ for all the four distributions. The median has a slight advantage over the

L functional in term of the *worst case* relative influence of an infinitesimal point mass contamination on the underlying estimator. (ii) On the other hand, $|IF(x; \mathrm{Med}(F))|$ is actually greater than that of $L(F)$ for most of points $x \in \mathbb{R}^1$ for these distributions, as exemplified in Figure 1. (iii) This implies that the relationship between $E(IF^2(X; \mathrm{Med}(F))$ and $E(IF^2(X; \mathrm{L}(F))$, the asymptotic variances of the median and the L estimator, can be the very opposite, namely, $E(IF^2(X; \mathrm{Med}(F)) > E(IF^2(X; \mathrm{L}(F))$. Indeed, this is the case for three of the four distributions, as shown in Section 5.

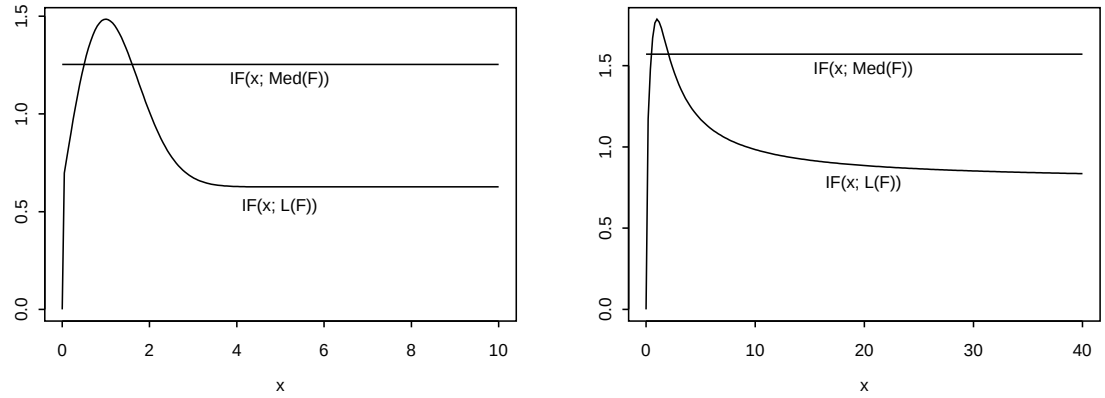

Figure 3.1: The influence functions of $\mathrm{Med}(F)$ and $L(F)$ at normal (left) and Cauchy (right) F.

4. Asymptotics Asymptotic results offer guidance for finite sample practice. They are often employed to develop practical inference procedures. Indeed, the essence of model statistical practice is asymptotic approximation. Here we focus on the asymptotic normality of $L(F_n)$. The strong consistency of $L(F_n)$ can be established in a similar but less involved manner with much less (and weaker) assumptions. Let F_n be the empirical version of F based on a random sample from F. To establish desired results, we need the following condition:

$$\mathbf{C}\text{: } \mu(F_n) - \mu(F) = \frac{1}{n}\sum_{i=1}^{n} f_\mu(X_i) + o_p(\frac{1}{\sqrt{n}}), \quad \sigma(F_n) - \sigma(F) = \frac{1}{n}\sum_{i=1}^{n} f_\sigma(X_i) + o_p(\frac{1}{\sqrt{n}}),$$

where $X_i \sim F$, $Ef_\mu(X_i) = Ef_\sigma(X_i) = 0$ and $Ef_\mu^2(X_i)$ and $Ef_\sigma^2(X_i)$ exist, $i = 1, \cdots, n$.

Condition C holds true for smooth M-estimators of location and scale; see p.133 (Theorem 3.1) of Huber (1981), for example. It also holds for less smooth estimators such as Med and MAD at smooth F; see Theorem 2.5.1 of Serfling (1980), for example.

Theorem 4.1 *Let $w'(r)$ be continuous and $w(\infty) = 0$, $\mu(F_n)$ and $\sigma(F_n)$ be strongly consistent for $\mu(F)$ and $\sigma(F)$, respectively, condition C hold, and $P(X = \mu(F)) = 0$. Then*

$$L(F_n) - L(F) = \frac{1}{n}\sum_{i=1}^{n} g(X_i) + o_p(\frac{1}{\sqrt{n}}),$$

where

$$g(x) = \frac{\int (L(F) - y)w'(d(y,F))\frac{d(y,F)f_\sigma(x) + \text{sign}(y-\mu(F))f_\mu(x)}{\sigma(F)}dF(y) + (x - L(F))w(d(x,F))}{\int w(d(y,F)dF(y)}$$

and $Eg(X) = 0$ and $v := Eg^2(X) < \infty$. Hence $\sqrt{n}\,(L(F_n) - L(F)) \xrightarrow{d} N(0, v)$.

The consistency of $L(F_n)$ follows. The proof of the theorem indicates that the strong consistency of $L(F_n)$ can be established less challengingly under weaker assumptions. Note that $g(x)$ has exactly the same form as $IF(x; L(F))$ except with $IF(x; \mu(F))$ and $IF(x; \sigma(F))$ replaced by $f_\mu(x)$ and $f_\sigma(x)$ respectively. In the location setting, it is very common to assume that F is symmetric, say, about 0, by virtue of the affine equivarience of the underlying location estimator. For symmetric F about the origin, $g(x)$ takes a much simpler form which also does not depend on f_σ, as manifested in the following result.

Corollary 4.2 *Let F be symmetric about 0. Then under the conditions of Theorem 4.1, the results in the theorem hold with*

$$g(x) = \Big(- \int |r|w'(|r|)dF_0(r)\ f_\mu(x) + xw(|x|/\sigma(F))\Big) \Big/ \int w(|r|)dF_0(r),$$

where $F_0(r) = F(\sigma(F)r)$. When $\mu = Med$ and $f(0) = F'(0) > 0$, $f_\mu(x) = \text{sign}(x)/(2f(0))$,

$$v = \Big(a^2 + 2a \int |x|w(|x|/\sigma(F))dF(x) + \int x^2 w^2(|x|/\sigma(F))dF(x)\Big) \Big/ b^2,$$

where $a = -\int |r|w'(|r|)dF_0(r)/(2f(0))$ and $b = \int w(|r|)dF_0(r)$.

The asymptotic result can be very useful for inferring the underlying location parameter in practice. Although, quantities such as a and b in v are unknown, we can use the corresponding sample versions as strong consistent estimators. The density $f(0)$ can be dealt with in the same manner as done in the inference procedures of the median and quantiles. Furthermore, bootstrap technique can be employed. Details will not be pursued here.

5. Efficiency Efficiency is another main concern in statistics in addition to robustness. Statistical procedures are desired to be efficient. This section investigates the large as well as finite sample relative efficiency of L_n. The asymptotic results in the last section become very useful here.

5.1. Asymptotic relative efficiency

Throughout the subsection, we assume that F is symmetric about 0, unless stated otherwise.

LIKELIHOOD WEIGHTING First we revisit the convenient case that $w(x)$ is the same as $f_0(x)$, the density of $F_0(x)$. This weighting scheme actually is a very reasonable one: points are weighted based on their likelihoods and large likelihoods correspond to large weights. With this likelihood weighting and $\mu = \text{Med}$, $g(x)$ in Corollary 4.2 takes the following form

$$g(x) = \Big(\int_0^\infty f^2(r)dr/(2f(0)) + |x|f(|x|)\Big)\text{sign}(x)\Big/\Big(2\int_0^\infty f^2(r)dr\Big), \tag{5.1}$$

which has nothing to do with $\sigma(F)$ any more. Now a straightforward calculation gives

$$v = \frac{1}{4}\Big(\frac{1}{2f(0)}\Big)^2 + \Big(\frac{1}{2f(0)}\Big)\frac{\int_0^\infty rf^2(r)dr}{\int_0^\infty f^2(r)dr} + \frac{\int_0^\infty r^2f^3(r)dr}{2\left(\int_0^\infty f^2(r)dr\right)^2} := \sigma^2_{L_n}, \tag{5.2}$$

where $\sigma^2_{T_n}$ denotes the asymptotic variance of $\sqrt{n}\,(T_n - T(F))$ for an estimator T_n. Note that $\sigma^2_{Med} = (1/2f(0))^2$ and $\sigma^2_{\bar{X}_n} = 1,\ \pi^2/3,\ 2,\ \infty$, respectively, for the four distributions in Table 1. We now list the asymptotic relative efficiencies of L_n and Med as follows.

Table 2 Asymptotic relative efficiency of $\boldsymbol{L_n}$ with $\boldsymbol{w(r) = f_0(r)}$

distribution F	$N(0,1)$	LG$(0,1)$	DE$(0,1)$	CAU$(0,1)$
$\sigma^2_{Med}/\sigma^2_{L_n}$	1.0580	1.0751	0.9558	1.1656
$\sigma^2_{\bar{X}_n}/\sigma^2_{L_n}$	0.6735	0.8843	1.9115	∞
$\sigma^2_{\bar{X}_n}/\sigma^2_{Med}$	0.6366	0.8225	2.0000	∞

The random-coefficient estimator L_n outperforms the median in but the double exponential distribution case where it is about 96% efficient relative to the median. The latter

case is no surprising since in this case the median is the uniformly minimum variance unbiased estimator (UMVUE). It even attains the Cramér-Rao lower bound.

There are two immediate concerns about the likelihood weighting scheme. First, one has to estimate the likelihood function $f(x)$, which can be even more challenging than estimating the location parameter in practice, and there is no fixed weight function. Second, the random-coefficient estimator, albeit more efficient than the median in most cases, has a low efficiency relative to the mean at light tailed (normal and logistic) distribution cases.

OUTLYINGNESS WEIGHTING Here we propose a single weight function so that L_n can have all the advantages in the likelihood weighting case but be very efficient relative to the sample mean at light tailed distributions. The weight function $w(x)$, which is continuously differentiable on $[0, \infty]$, takes the following form (cf Zuo, Cui and He (2004))

$$w(x, c, k) = I(x \leq c) + (e^{-k(1-((1+c)/(1+x))^2)^2} - e^{-k})/(1 - e^{-k})I(x > c), \qquad (5.3)$$

where $0 \leq c < \infty$ and $k > 0$ are two parameters. The function is plotted in Figure 2.

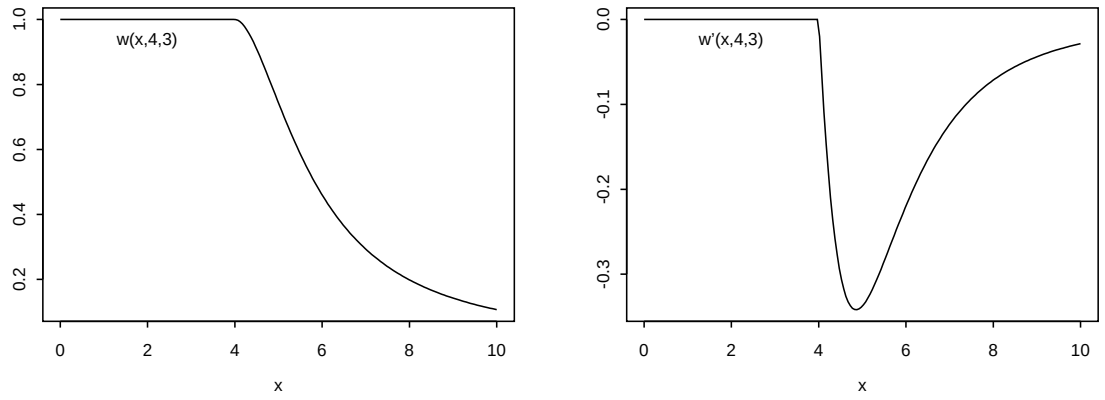

Figure 5.2: The weight function $w(x, c, k)$ and its derivative functions $w'(x, c, k)$.

The basic idea of this weight function is as follows. Points close to the center of symmetry are treated equally (and simply get averaged) and those away from the center are exponentially down-weighted based on their relative distance to the center. The parameter c controls the proportion of central points to be averaged and k determines the exponential down-weighting rate. Generally speaking, a large c is favorable at light tailed distributions whereas a small c at heavy tailed ones. The same is true in general for the parameter k.

A natural question raised here is how to determine the values of c and k for a given distribution or data set. We discuss a fully and a partially adaptive approaches here.

First, we can, of course, select the most favorable ones at each given model by appropriately tuning the parameters. For example, at normal models, if we select an extremely large c, then L_n becomes the most efficient sample mean. On the other hand, at double exponential models, if we let c be 0 and k be extremely large, then L_n becomes the most efficient median. The true distribution, of course, is unknown in practice. So for a given data set, we may select c and k such that the resulting L_n has the smallest sample variance. Here c and k become data-dependent. We can denote them by c_n and k_n. Suppose that for a given population distribution F there are optimal values $c(F)$ and $k(F)$ (the ones that can minimize the asymptotic variance of L_n). Then one can show that c_n and k_n are consistent estimators of $c(F)$ and $k(F)$ under some regularity conditions. Employing empirical process theory, one can further show that the asymptotic results in the last section still hold. Proofs and further details of this fully adaptive approach will not be pursued here.

Instead, we now focus on a partially adaptive approach and give a general rule here for selecting the values of k and c. If the distribution is light tailed, we select $c = 4\sigma(F)$ (σ = MAD); otherwise, we take $c = \sigma(F)/4$ (a more involved way is to replace 4 with $\sigma^2_{Med}/\sigma^2(F)$.) For simplicity, we fix $k = 3$ here. To determine if a distribution is light or heavy tailed, we can employ various tail indices (or tail probabilities) in the literature. Another way we propose here is to simple check the ratio of $\sigma^2_{Med}/\sigma^2_{\bar{X}_n}$. If it is greater than 0.9, we treat F as light tailed, or heavy tailed otherwise. Note that at the sample level, all the above quantities can be approximated by the corresponding (strongly consistent) sample versions. (With the help of empirical process theory, one can prove that the asymptotic results in the last section remain valid for a weight function with random parameters. Details will not be pursued here though. Details will not be pursued here.) Table 2 lists the efficiency results of L_n relative to Med and $\bar{X}_n$ with $w(x, c, k)$ as the weight function.

Table 3 Asymptotic relative efficiency of L_n with $w(r) = w(r, c, 3)$

distribution F	$N(0,1)$	LG$(0,1)$	DE$(0,1)$	CAU$(0,1)$
	$c = 4\sigma(F)$	$c = 4\sigma(F)$	$c = \sigma(F)/4$	$c = \sigma(F)/4$
$\sigma^2_{Med}/\sigma^2_{L_n}$	1.5071	1.2572	0.9782	1.1054
$\sigma^2_{\bar{X}_n}/\sigma^2_{L_n}$	0.9595	1.0340	1.9563	∞
$\sigma^2_{\bar{X}_n}/\sigma^2_{Med}$	0.6366	0.8225	2.0000	∞

The table contains two striking results. First, the efficiency of L_n with $w(x, c, k)$ is universally improved relative to that given in Table 2 with the likelihood weighting, expect in the cauchy case where it is very slightly reduced (but still higher than that of the median). The median is only about 66%, 80%, and 90% efficient relative to L_n at normal, logistic, and cauchy models, respectively. At double exponential model where the median performs the best, L_n still possesses a 98% relative efficiency. Second, L_n with $w(x, c, k)$ becomes extremely efficient at light tailed distributions. At normal model where the mean performs the best, it is 96% efficient. At logistic model, it is more efficient than the mean.

The efficiency of L_n relative to the best estimators, which is called *absolute efficiency*; see, e.g., p. 363 of Lehmann (1983), is very high at normal and double exponential models. A very relevant question is how is the performance of L_n at other two models relative to the best possible estimators? We answer this question by calculating the ratio of Cramér-Rao lower bound (crlb) to $\sigma^2_{L_n}$ at these models. The results are given in Table 4 below.

Table 4 Absolute efficiency of L_n with $w(r) = w(r, c, 3)$

distribution F	$N(0,1)$	LG$(0,1)$	DE$(0,1)$	CAU$(0,1)$
	$c = 4\sigma(F)$	$c = 4\sigma(F)$	$c = \sigma(F)/4$	$c = \sigma(F)/4$
crlb$/\sigma^2_{L_n}$	0.9595	0.9409	0.9782	0.8960

The estimator L_n is about 90% or high efficient relative the most efficient estimator at each model. This is remarkable. Leading competitors in the literature include Hodges-Lehmann estimator and the 12.5% trimmed mean. Their absolute efficiency [see p. 364 and p. 386 of Lehmann (1983)] is comparable to that of L_n, though lower in most cases. These two estimators, however, pay a price of very low breakdown points for the high efficiencies.

TUKEY BI-WEIGHTING The surprisingly high efficiency values of L_n resulting from the use of $w(r, c, k)$ raise the question as to whether other *popular* weight functions could do a even better job. A very popular weight function in the literature is Tukey bi-weight function

$$w_T(x, c) = x(1 - (x/c)^2)^2 I(x <= c), \quad x \in [0, \infty), \tag{5.4}$$

where $c > 0$ is a parameter. The efficiency of L_n with this function, however, is lower than that with $w(x, c, k)$ at all the four models. Indeed, even with the *best choice* of c (resulting in the maximum efficiency) at each of the four different models, the asymptotic efficiency of L_n relative to the median is respectively 123.31%, 106.33%, 95.77%, and 108.30% instead of 150.71%, 125.72%, 97.82%, and 110.54% in the $w(x, c, k,)$ case given in Table 3.

CONTAMINATED MODELS In practice, data may not follow exactly the models discussed above. For example, we may have the contaminated normal distribution suggested by Tukey (1960) as a model for observations, which usually follow a normal distribution but where occasionally something goes wrong with the experiment or its data processing step so that the resulting observation is a gross error. A legitimate concern about the location estimators above is how do they perform when the underlying models are slightly contaminated. That is, are the efficiency results robust against the model assumptions? We select the (Tukey) contaminated normal model, as an example, to investigate the asymptotic efficiency behavior of the estimators. Under this model, the distribution takes the form

$$F(x; \epsilon, \tau) = (1 - \epsilon)\Phi(x) + \epsilon\Phi(x/\tau), \quad 0 < \epsilon < 1/2, \quad 1 < \tau. \tag{5.5}$$

It has the same mean θ (which is assumed to be 0 w.l.o.g.) but a larger variance. For this mixture F, the density is $f(x; \epsilon, \tau) = (1 - \epsilon)\phi(x) + \frac{\epsilon}{\tau}\phi(\epsilon/\tau)$ and $\sigma^2_{\bar{X}_n} = (1 - \epsilon) + \epsilon\tau^2$, where $\phi(x)$ is the p.d.f. of the standard normal distribution. We now list the efficiency results for a number of combinations of ϵ and τ (as before MAD is the scale measure σ).

Table 5 Asymptotic relative efficiency of L_n at $F(x; \epsilon, \tau)$ with $w(r) = w(r, 4\sigma(F), 3)$

distribution F	$F(x; 0.01, 5)$	$F(x; 0.05, 4)$	$F(x; 0.10, 3)$	F(x; 0.15, 2)
$\sigma^2_{Med}/\sigma^2_{L_n}$	1.5112	1.4868	1.4211	1.4085
$\sigma^2_{\bar{X}_n}/\sigma^2_{L_n}$	1.1739	1.5345	1.4186	1.1124
$\sigma^2_{\bar{X}_n}/\sigma^2_{Med}$	0.7768	1.0321	0.9982	0.7898

The table conveys some remarkable messages. Firstly, L_n is overwhelmingly more efficient than the median in all cases. Indeed the median is only about 66%, 67%, 70% and 71% efficient relative to L_n in the four cases, respectively. Secondly, L_n outperforms the sample mean $\bar{X}_n$ in all cases. Note that the mean is the best location estimator (the UMVUE) without the slight contamination. Thirdly, with some special combinations ϵ and τ $(4((1-\epsilon)+\epsilon\tau^2)f^2(0;\epsilon,\tau) > 1)$, it is seen that the median can be more efficient than the mean when the distribution slightly deviates from a pure normal model. Note that the combinations in Table 4 are in favor of the mean. A large τ accompanied with a large ϵ would lead to a relatively worse performance for the mean and a relatively better one for L_n and the median. Finally, our calculation indicates that the variances of L_n (and Med) are very robust against slight departure from the assumed distribution.

5.2. Finite sample relative efficiency

All the results in the last section are in the asymptotic nature. This naturally raises concerns about their validity and relevance in finite sample applications. Indeed, asymptotic results are often vulnerable to criticism for their merits in finite sample practice. We address this concern in this section with finite sample simulations.

Here we first generate 1000 samples of size 50 from each of the four models in Table 3. Then we calculate L_n, the median, and the mean for each of the 1000 samples. Finally we calculate the sample variance (denoted by $s^2_{T_n}$ for an estimator T_n) of the 1000 estimates at each of the three cases. The weight function $w(x, c, 3)$ is used for L_n with $c = \sigma(F_n) * 4$ or $\sigma(F_n)/4$ for light or heavy tailed models, respectively. Here σ = MAD and μ = Med.

The efficiency results, listed in Table 6 below, are strikingly close to the asymptotic ones in Table 3 except in the double exponential case for the efficiencies of L_n and the median relative to the mean (1.6403 and 1.6469 in Table 6 versus 1.9563 and 2.0000 in Table 4). The slow convergence in this cases presumably is due to the lack of smoothness of f at zero.

Again L_n is seen to be much more efficient than the median except at the double exponential case. In the latter case it is almost as efficient as the median. But in the other three cases, the median is only about 68%, 79%, and 89% efficient relative to L_n. On the other hand, unlike the median, L_n is very efficient (with efficiency about 92% and 106%) relative to the mean at normal and logistic (light tailed) distributions and overwhelmingly more efficient than the mean at heavy tailed ones (164% and 576×10^6 %).

Table 6 Finite sample relative efficiency of L_n with $n = 50$, $w(r) = w(r, c, 3)$

distribution F	$N(0,1)$	LG$(0,1)$	DE$(0,1)$	CAU$(0,1)$
	$c = 4\sigma(F_n)$	$c = 4\sigma(F_n)$	$c = \sigma(F_n)/4$	$c = \sigma(F_n)/4$
$s^2_{Med}/s^2_{L_n}$	1.4642	1.2641	0.9960	1.1287
$s^2_{\bar{X}_n}/s^2_{L_n}$	0.9179	1.0567	1.6403	5.7596×10^6
$s^2_{\bar{X}_n}/s^2_{Med}$	0.6269	0.8359	1.6469	5.1031×10^6

As pointed out hereinbefore, data in practical applications do not follows exactly the above four distributions, it is therefore important to examine the relative efficiency of L_n at slightly contaminated models. We select the Tukey contaminated normal model again to exemplify the finite sample behavior of L_n. In practice, the contaminating distribution is most likely to have a small location change as well in addition to the scale change. We therefore consider the contaminated model which takes the following form

$$F(x; \epsilon, \tau, \eta) = (1-\epsilon)\Phi(x) + \epsilon\Phi((x-\eta)/\tau), \quad 0 < \epsilon < 1/2, \quad -\infty < \eta < \infty, \ 1 < \tau. \quad (5.6)$$

We generate 1000 samples of size 50 from this model and calculate L_n, the median, and the mean for each samples as before. The weight function $w(x, c, 3)$ with $c = 4\sigma(F_n)$ is used for L_n. Again $\mu =$ Med and $\sigma =$ MAD. Since the target location parameter θ is still 0, we will calculate the empirical mean squared error: $\text{mse}_{T_n} = \frac{1}{m}\sum_{i=1}^{m}(T_n^i - 0)^2$ for an estimator T_n. Here $m = 1000$ and $n = 50$. The ratio of the empirical mean squared errors will be used for the relative efficiency. The results are listed in Table 7 below.

Table 7 Finite sample relative efficiency of L_n with $n = 50$, $w(r) = w(r, 4\sigma(F_n), 3)$

distribution F	$F(x; 0.02, 5, 0.5)$	$F(x; 0.04, 4, 1)$	$F(x; 0.10, 3, 1.5)$	F(x; 0.14, 2, 2)
$\text{mse}_{Med}/\text{mse}_{L_n}$	1.4764	1.3238	1.2430	0.9520
$\text{mse}_{\bar{X}_n}/\text{mse}_{L_n}$	1.4669	1.3574	1.9143	1.9098
$\text{mse}_{\bar{X}_n}/\text{mse}_{Med}$	0.9936	1.0254	1.5401	2.0061

The table entries indicates that under these contaminated normal models L_n is more efficient than the median except at one case where it is 95.2% efficient relative to the median. On the other hand, L_n is overwhelmingly more efficient than the mean which is the best estimator if the underlying model is not slightly contaminated. The histograms of the 1000 medians, means, and L_n's based on the samples from $F(x; 0.02, 5, 0.5)$ are given in Figures 2 and 3. The superiority of L_n over the median and the mean is very clear visually.

6. Concluding remarks The random-coefficient estimator L_n shares the *best breakdown point* robustness of the median while possessing a overwhelmingly high efficiency relative to the latter and the mean at a variety of light- and heavy- tailed distributions with appropriate (outlyingness) weighting schemes. In fact, it can be much more efficient than both the median and the mean for the most models considered in the paper. Further, it enjoys an extremely high (about 90% or higher) *absolute efficiency* at the four distributions considered. In the more practical contaminated model settings, L_n can outperform both the mean and the median at most light- and heavy- tailed distributions we considered. Findings in the paper indicate that L_n can serve very well as an alternative to both the mean and the median in practice.

The success of L_n is mainly due to the random weighting idea. The latter appeared in the literature and consequently L_n has close connections with existing estimators. First, L_n is of the exact form of Tukey W-estimator: $T_n = \sum_{i=1}^n w_i X_i / \sum_{i=1}^n w_i$. It, however, is not a W-estimator since the latter is defined implicitly with $w_i = w((X_i - T_n)/(c\sigma_n))$ for the bi-weight w and σ_n =MAD$_n$; see, e.g., page 205 of Mosteller and Tukey (1977). It can be viewed as a *one-step* W-estimator with μ_n as the initial estimate of T_n, nevertheless.

Second, L_n clearly is the solution of $\sum_{i=1}^n w((X_i - \mu_n)/\sigma_n)((X_i - L_n)/\sigma_n) = 0$. On the other hand, an M-estimator T_n is the solution of $\sum_{i=1}^n \psi(X_i, T_n) = 0$; see, e.g., Serfling (1980). Hence L_n is not exactly an M-estimator. Neither is it an M-estimator with an initial scale estimate, the solution of $\sum_{i=1}^n \psi((X_i - T_n)/\sigma_n) = 0$; see Huber (1981). It can be viewed as a *one-step* M-estimator with initial scale and location estimates σ_n and μ_n.

Third, note that a weighted least squares estimator with an initial scale estimate minimizes: $\sum_{i=1}^n w((X_i - \theta)/\sigma_n)((X_i - \theta)/\sigma_n)^2$, among all $\theta \in \mathbb{R}^1$. Thus, L_n can be viewed as a weighted least squares estimator with initial *scale and location* estimates σ_n and μ_n since it minimizes: $\sum_{i=1}^n w((X_i - \mu_n)/\sigma_n)((X_i - \theta)/\sigma_n)^2$, among all $\theta \in \mathbb{R}^1$.

Finally, L_n can be viewed as a special one-dimensional version of Stahel-Donoho (Stahel (1981), Donoho (1982)) or more generally the multi-dimensional data depth weighted estimators discussed in Zuo, Cui and He (2004) and Zuo, Cui and Young (2004). The latter two papers focus on the general multi-dimensional versions and establish general results under very general (and hence strong) assumptions. This paper, on the other hand, focuses on the performance evaluation (relative efficiency) of L_n with respect to leading competitors. It also scrutinizes the robustness and asymptotic properties of this special one-dimensional case in a more precise manner and obtains specific and stronger results under weaker and more precise conditions. Furthermore, it covers topics such as gross error sensitivity and asymptotic efficiency that are not treated in the aforementioned two papers.

7. Appendix: proofs of main results **Proof of Theorem 3.1**. Write μ and σ for Med and MAD for convenience. Recall that we adopt the convention that $(y-\mu(Z))/\sigma(Z) = 0$ if $y-\mu(Z) = \sigma(Z) = 0$ for any y and data $Z = \{Z_1, \ldots, Z_n\}$ in $\mathbb{R}^1$ throughout the paper.

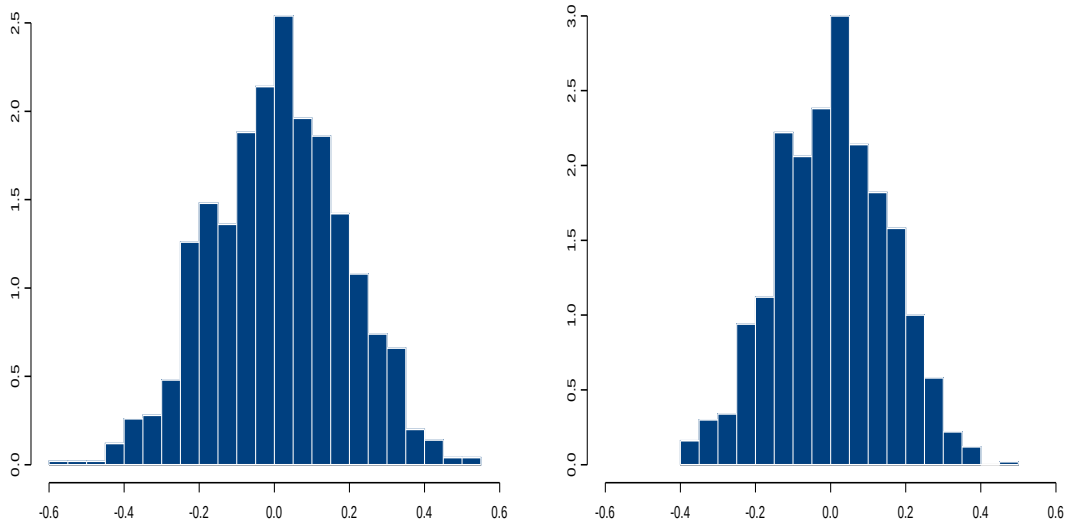

Figure 7.3: The histograms of 1000 medians (left) and L_n's (right) based on $F(x; 0.02, 5, 0.5)$.

First, we show that $m = \lfloor (n+1)/2 \rfloor$ points are sufficient to breakdown L_n. Move m original points in X to the same site y and let $y \to \infty$. Denote by $Z = X^m = \{Z_1, \ldots, Z_n\}$

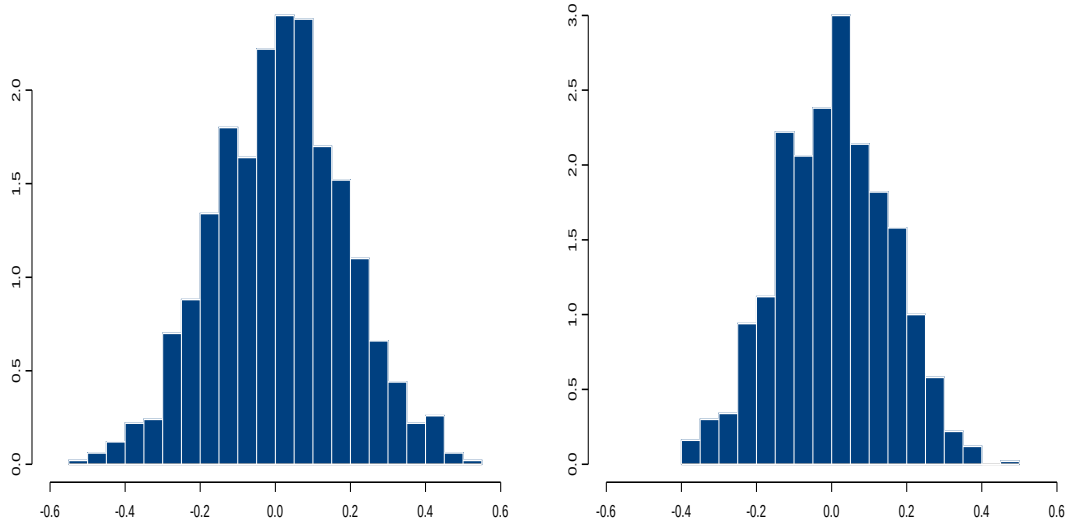

Figure 7.4: The histograms of 1000 means (left) and L_n's (right) based on $F(x; 0.02, 5, 0.5)$.

the contaminated data set. Since $n - m < (n + 1)/2$, thus $\sigma(Z) \to \infty$. It is readily seen that $d(y, Z) = 0$ for odd n and $= 1$ for even n. Hence $w_i = w(d(Z_i, Z)) \geq \delta > 0$ for some δ and for all m points Z_i at y. It is readily seen that the numerator of L_n approaches infinity as $y \to \infty$ while the denominator is no less than $m\delta$. Hence m points can break down L_n.

Second, we show that $m - 1$ points are not sufficient to breakdown L_n. We first show that the denominator of L_n is uniformly bounded away from zero for any Z resulting form contaminating X with $m-1$ points. Let $Z = \{Z_1, \cdots, Z_n\}$ and $Z_{(1)}, \ldots, Z_{(n)}$ be the ordered values of $Z_1, \ldots, Z_n$ such that $Z_{(1)} \leq Z_{(2)} \leq \ldots \leq Z_{(n)}$. Then it is not difficult to see that $d(Z_{(\lfloor (n+1)/2 \rfloor)}, Z) \leq 1$. Since $w(r) > 0$ and $\inf_{0 \leq r \leq 1} w(r) > 0$, thus the denominator of L_n is uniformly bounded away from zero for any Z mentioned above.

Now we show that the numerator of L_n is bounded above uniformly for any Z resulting form contaminating X with $m-1$ points. Since $n-(m-1) > m-1$ and $n-(m-1) > n/2$, thus μ and $\sigma(Z)$ are also bounded above uniformly for any contaminated data set Z. Assume without loss of generality that $w(r) \leq 1$ and $rw(r) \leq 1$ for $r \geq 0$. By the definition of $d(Z_i, Z)$ we have $|Z_i| \leq \sigma(Z)d(Z_i, Z) + |\mu(Z)|$. Thus

$$w(d(Z_i, Z))|Z_i| \leq \sup_Z \sigma(Z) + \sup_Z |\mu(Z)|,$$

which is bounded and hence implies that the numerator of L_n is bounded above uniformly for any aforementioned Z. Thus, $m - 1$ contaminating points can not break down L_n. □

Proof of Theorem 3.2. Denote $F(\epsilon, \delta_x)$ by $F_{\epsilon x}$ for convenience. First we can write

$$L(F_{\epsilon x}) - L(F) = \frac{(1-\epsilon)\int (y - L(F))w(d(y, F_{\epsilon x}))dF(y)}{\int w(d(y, F_{\epsilon x}))dF_{\epsilon x}(y)} + \frac{\epsilon w(d(x, F_{\epsilon x}))(x - L(F))}{\int w(d(y, F_{\epsilon x}))dF_{\epsilon x}(y)}.$$

By the existence of $IF(x; \mu(F))$ and $IF(x; \sigma(F))$ for any given x, we conclude that $\mu(F_{\epsilon x}) \to \mu(F)$ and $\sigma(F_{\epsilon x}) \to \sigma(F) > 0$ as $\epsilon \to 0$ for any fixed x. By the continuities of $d(y, F)$ in μ and σ, and $w(r)$ in r, and by the Lebesgue's dominated convergence theorem, we see that the second part on the right hand side gives immediately the second part of the desired influence function result in the theorem. We therefore focus on the first part of the last display. Denote it by $L_1(F_{\epsilon x})$. Then we observe that

$$\begin{aligned} L_1(F_{\epsilon x}) &= \frac{(1-\epsilon)\int (y - L(F))(w(d(y, F_{\epsilon x})) - w(d(y, F)))dF(y)}{\int w(d(y, F_{\epsilon x}))dF_{\epsilon x}(y)} \\ &= \frac{(1-\epsilon)\int (y - L(F))w'(\theta(y, F, F_{\epsilon x}))(d(y, F_{\epsilon x}) - d(y, F))dF(y)}{\int w(d(y, F_{\epsilon x}))dF_{\epsilon x}(y)}, \end{aligned}$$

for some $\theta(y, F, F_{\epsilon x})$ in-between $d(y, F)$ and $d(y, F_{\epsilon x})$. Now we can write

$$d(y, F_{\epsilon x}) - d(y, F) = \frac{(|y - \mu(F_{\epsilon x})| - |y - \mu(F)|) - (\sigma(F_{\epsilon x}) - \sigma(F))d(y, F)}{\sigma(F_{\epsilon x})}.$$

Hence we see that $IF(x; d(y, F))$ is exactly the one given in the theorem. Now the given conditions, Lebesgue's dominated convergence theorem, yield the first part of the desired result. The boundedness of $GES(L(F)$ under the given conditions is straightforward. □

Proof of Theorem 4.1. Let $D(y, F) = 1/(1 + d(y, F))$. Then $0 \le D(y, F) \le 1$ for any y and F in $\mathbb{R}^1$ (since $0 \le d(y, F) \le \infty$). Note that $w(d(y, F)) = w(1/D(y, F) - 1) := w_*(D(y, F))$ and $w_*(r)$ is continuously differentiable on $[0, 1]$ since $w(r)$ is on $[0, \infty]$. We will use $w_*(D(y, F))$ for $w(d(y, F))$ in the proof below to take the great technical advantage. To prove the theorem we first establish the following results.

Lemma 7.1 *Under the conditions of Theorem 4.1, we have*

$$\sup_{y \in \mathbb{R}^1} (1 + |y|)|D(y, F_n) - D(y, F)| = o(1), \quad a.s. \tag{7.1}$$

$$\sup_{y \in \mathbb{R}^1} (1 + |y|)|D(y, F_n) - D(y, F)| = O_p(\frac{1}{\sqrt{n}}), \tag{7.2}$$

$$D(y, F_n) - D(y, F) = \frac{s(F_n)d(y, F) + l(F_n)\text{sign}(y - \mu(F)) - |l(F_n)|I(y = \mu(F)}{\sigma(F)(1 + d(y, F))^2} + o_p(\frac{1}{\sqrt{n}}), \tag{7.3}$$

where $s(F_n) = \sigma(F_n) - \sigma(F)$ and $l(F_n) = \mu(F_n) - \mu(F)$.

PROOF: First we can write

$$d(y, F_n) - d(y, F) = \frac{|l(F_n)|I(y = \mu(F)) - l(F_n)\text{sign}(y - \mu(F)) - s(F_n)d(y, F)}{\sigma(F_n)}. \quad (7.4)$$

Therefore

$$|d(y, F_n) - d(y, F)| \leq (|\mu(F_n) - \mu(F)| + d(y, F)|\sigma(F_n) - \sigma(F)|)/\sigma(F_n).$$

Hence we have

$$|D(y, F_n) - D(y, F)| \leq \frac{|\mu(F_n) - \mu(F)|/(1 + d(y, F)) + |\sigma(F_n) - \sigma(F)|/(1 + d(y, F_n))}{\sigma(F_n)},$$

which yields immediately displays (7.1) and (7.2) by the strong consistency of $\mu(F_n)$ and $\sigma(F_n)$ and the condition C since $\sup_{y\in\mathbb{R}^1}(1 + |y|)/(1 + d(y, G)) < \infty$ a.s. for $G = F$ or F_n with large n. Now (7.3) follows from (7.4) and the condition C. □

PROOF OF THEOREM 4.1 Write

$$L(F_n) - L(F) = \int (y - L(F))w_*(D(y, F_n))dF_n(y) \Big/ \int w_*(D(y, F_n))dF_n(y). \quad (7.5)$$

We first work on the numerator. Observe that

$$\begin{aligned} \int (y - L(F))w_*(D(y, F_n))dF_n(y) &= \int (y - L(F))(w_*(D(y, F_n)) - w_*(D(y, F)))dF_n(y) \\ &+ \int (y - L(F))w_*(D(y, F))d(F_n - F)(y). \quad (7.6) \end{aligned}$$

The strong law of large numbers takes care of the second term on the right hand side. We now focus on the first term on the right side. Denote it by N_{1n}. Then we have

$$\begin{aligned} N_{1n} &= \int (y - L(F))w'_*(\theta(y, F, F_n))(D(y, F_n) - D(y, F))dF_n(y) \\ &= \int (y - L(F))w'_*(D(y, F))(D(y, F_n) - D(y, F))dF_n(y) + o_p(\frac{1}{\sqrt{n}}), \end{aligned}$$

by the continuous differentiability of w_* on $[0, 1]$ and Lemma 7.1, where $\theta(y, F, F_n)$ is a point in-between $D(y, F_n)$ and $D(y, F)$. Now we show via empirical process theory that

$$N_{1n} = \int (y - L(F))w'_*(D(y, F))(D(y, F_n) - D(y, F))dF(y) + o_p(\frac{1}{\sqrt{n}}). \quad (7.7)$$

Define a set of Borel measurable functions by

$$\mathcal{F} := \{h(y,\alpha,\beta) := \frac{(y - L(F))\, w_*'(D(y,F))}{1 + |y-\alpha|/\beta} : \mu(F)/2 < \alpha < 2\mu(F),\ \ \sigma(F)/2 < \beta < 2\sigma(F)\}$$

Clearly, $h(y, \mu(G), \sigma(G))$ belongs to $\mathcal{F}$ for $G = F$ and for $G = F_n$ a.s. with large n. Let $\gamma_i = (\alpha_i, \beta_i)' \in \mathbb{R}^2$. Then it can be seen that there is an $0 < M < \infty$ such that

$$\begin{aligned} |h(y,\alpha_1,\beta_1) - h(y,\alpha_2,\beta_2)| &\leq \frac{|y - L(F)|\,|w_*'(D(y,F))|}{\beta_2 + |y-\alpha_2|}(|\alpha_1-\alpha_2| + |\beta_1-\beta_2|) \\ &\leq M\, \|2(\gamma_1 - \gamma_2)\|, \end{aligned}$$

by the continuity of $w_*'(r)$ on $[0,1]$ and $|y - L(F)|/(\beta_2 + |y-\alpha_2|)$ on $[0,\infty]$. Now by 19.5 and 19.7 of van der vaart (1998) (or section 2.7.4 of Wellner and van der Vaart (1996)), we conclude that $\mathcal{F}$ is a P-Donsker class. On the other hand, by the preceding display and the strong consistency of $\mu(F_n)$ and $\sigma(F_n)$, it is readily seen that

$$\int (h(x,\mu(F),\sigma(F)) - h(x,\mu(F_n),\sigma(F_n)))^2 dF(x) = o(1), \quad a.s.$$

Now invoking Lemma 19.24 of van der Vaart (1998), we have display (7.7). This, in conjunction with Lemma 7.1, Fubini's Theorem, and display (7.6), gives

$$\begin{aligned} &\int (y - L(F)) w_*(D(y,F_n)) dF_n(y) \\ &= \int \Big(\int (x - L(F)) w_*'(D(x,F)) \frac{d(x,F) f_\sigma(y) + \mathrm{sign}(x-\mu(F)) f_\mu(y)}{\sigma(F)(1+d(x,F))^2} dF(x) \\ &+ (y - L(F)) w_*(D(y,F)) \Big) d(F_n - F)(y) + o_p(1/\sqrt{n}). \end{aligned} \tag{7.8}$$

Likewise and less challengingly, we can show for the denominator of display (7.5) that

$$\int w_*(D(y,F_n) dF_n(y) = \int w_*(D(y,F)) dF(y) + o(1), \quad a.s.$$

Note that $w_*'(D(y,F)) = -w'(d(y,F))(1+d(y,F))^2$. The desired result follows. □

REFERENCES

[1] Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Dept. Statistics, Harvard University.

[2] Donoho, D. L., and Huber, P. J. (1983). The notion of breakdown point. In *A Festschrift foe Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, CA.

[3] Hampel, F. R., Ronchetti, E. Z., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The approach based on influence functions*. Wiley, New York.

[4] Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons.

[5] Lehmann, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

[6] Monsteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression*. Addison-Wesley, Reading, Mass.

[7] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.

[8] Stahel, W. A. (1981). Breakdown of covariance estimators. Research Report 31, Fachgruppe für Statistik, ETH, Zürich.

[9] Tukey, J. W. (1960). A survey of sampling from contaminated distributions, in *Contributions to PRobability and Statistics*, Olkin, Ed. Stanford University Press, Stanford, Clif.

[10] Tukey, J. W. (1970). *Exploratory Data Analysis*, mimeographed preliminary edition.

[11] Tukey, J. W. (1975). Mathematics and picturing data. *Proc. Intern. Congr. Math. Vancouver 1974* **2** 523–531.

[12] van der Vaart, A. W. (1998). *Asymptotic Sattistics*. Cambridge University Press.

[13] van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer.

[14] Zuo, Y. (2003). Projection based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490.

[15] Zuo, Y., Cui, H., and He, X. (2004). On the Stahel-Donoho estimators and depth-weighted means of multivariate data. *Ann. Statist.* **32** (1) 167–188.

[16] Zuo, Y., Cui, H., and Young, D. (2004). Influence function and maximum bias of projection depth based estimators. *Ann. Statist.* **32** (1) 189–218.

[17] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28**(2) 461–482.