

SOME PATHOLOGICAL REGRESSION ASYMPTOTICS UNDER STABLE CONDITIONS

ROGER KOENKER AND STEPHEN PORTNOY

ABSTRACT. We consider a simple through-the-origin linear regression example introduced by Rousseeuw, van Aelst and Hubert (1999). It is shown that the conventional least squares and least absolute error estimators converge in distribution *without normalization* and consequently are inconsistent. A class of weighted median regression estimators, including the maximum depth estimator of Rousseeuw and Hubert (1999), are shown to converge at rate n^{-1} . Finally, the maximum likelihood estimator is considered, and we observe that there exist estimators that converge at rate n^{-2} . The results illustrate some interesting, albeit somewhat pathological, aspects of stable-law convergence.

1. INTRODUCTION

Rousseeuw and Hubert (1999) formulate a notion of data depth in linear regression settings and introduce a highly robust maximal depth estimator. In discussion of this paper, Koenker (1999) suggested that such high-breakdown bound estimators generally sacrifice efficiency in comparison to the median regression estimator in cases where the design observations have heavy tails. Though this is true in homoscedastic cases, Rousseeuw, van Aelst and Hubert (1999, p. 422) present an intriguing example of a simple heteroscedastic case for which their Monte Carlo simulations suggested that highly robust estimators that ignore large x -values are better than traditional estimators making stronger use of these values.

The example involves the simple, scalar-parameter, through-the-origin linear regression model. Let (Z_1, Z_2, Z_3) be *iid* standard normal and define,

$$\begin{aligned} X &= (Z_1/Z_3)^2 \text{sgn}(Z_1/Z_3) \\ Y &= X\beta + (Z_2/Z_3)^2 \text{sgn}(Z_2/Z_3) \end{aligned}$$

It is easy to see that the conditional distribution of Y given $X = x$ is symmetric about $x\beta$. Thus, a wide variety of estimation methods may be applied, even though the Z_3 term common to both X and the regression error makes the variability of Y highly dependent on X (see (2.5) below). Note that the marginal distributions of both X and the regression error have a symmetrized $F_{1,1}$ distribution, which has no

Key words and phrases. Asymptotics, median regression, LAD regression, stable law, data depth.

Version: March 27, 2014. Corresponding Author: Stephen Portnoy, Department of Statistics, University of Illinois, Champaign, Illinois, 61820. This research was partially supported by NSF grants SBR-9617206 and DMS 97-03758.

integer moments, so that the least squares estimator (OLS) might be expected to perform poorly. As alternatives to the OLS estimator, we consider several median regression estimators all of the form

$$\hat{\beta}_n = \arg \min_{b \in \mathbb{R}} \sum w_i |y_i - x_i b|$$

where the weights w_i are chosen to be: (i) $w_i \equiv 1$, (ii) $w_i = |x_i|^{-1}$ and (iii) $w_i = (x_i(1 + |x_i|))^{-1/2}$. The unweighted median regression estimator, or least absolute deviation (LAD) estimator may be viewed as the weighted median of the candidate slopes y_i/x_i with weights $|x_i|$. Case (ii) is Rousseeuw and Hubert's maximal depth estimator (MDE), which is simply the unweighted median of the y_i/x_i . Case (iii) corresponds to the optimally weighted median regression estimator (WLAD).

In Table 1 we report median absolute error estimates for the four estimators based on a small Monte-Carlo experiment with 10,000 replications. Note that the results for the OLS and LAD estimators are reported without normalization, while results for the two weighted median regression estimators have been normalized by n . Results are reported for 8 different sample sizes.

Estimator	n=100	n=200	n=300	n=400	n=500	n=1000	n=5000	n=10000
$\hat{\beta}_{OLS}$	0.437	0.428	0.428	0.421	0.417	0.421	0.428	0.425
$\hat{\beta}_{LAD}$	0.212	0.207	0.206	0.207	0.213	0.200	0.204	0.209
$n\hat{\beta}_{WLAD}$	0.919	0.956	0.898	0.911	0.891	0.890	0.914	0.896
$n\hat{\beta}_{MDE}$	1.164	1.136	1.094	1.108	1.088	1.125	1.135	1.136

TABLE 1. Monte-Carlo Median Absolute Errors

We show the following formal results: (1) Both the least squares and the unweighted median regression estimators (case (i)) are inconsistent; and in fact they converge *without normalization* to non-degenerate limiting distributions, and (2) the maximal depth MDE and optimally weighted (WLAD) estimators converge in distribution at rate $1/n$, with the MDE about 80% efficient with respect to the WLAD estimator. Finally, we note that estimators that converge at rate n^{-2} are also possible, as follows from results of Ibragimov and Has'minskii (1981).

2. SOME BASIC DISTRIBUTION THEORY

Since all the estimators considered are affine invariant, take $\beta = 0$ (without loss of generality). Thus, consider the random variables:

$$(2.1) \quad X = (Z_1/Z_3)^2 \text{sgn}(Z_1/Z_3) \quad Y = (Z_2/Z_3)^2 \text{sgn}(Z_2/Z_3)$$

Transforming from (Z_1^2, Z_2^2, Z_3^2) to (X, Y, Z_3^2) , integrating out Z_3^2 , and invoking symmetry around the origin gives the joint distribution of (X, Y) as

$$(2.2) \quad f_{X,Y}(x, y) = \frac{1}{8\pi} |xy|^{-1/2} (|x| + |y| + 1)^{-3/2}.$$

As noted above, the marginal distribution of X is a symmetrized $F_{1,1}$ with density,

$$(2.3) \quad f_X(x) = \frac{1}{2\pi} \frac{1}{\sqrt{|x|}} \frac{1}{1 + |x|}$$

so the conditional density is

$$(2.4) \quad f_{Y|X}(y|x) = \frac{1}{4\sqrt{|y|}} (1 + |x|)(1 + |x| + |y|)^{-3/2}$$

Thus, Y is clearly not independent of X , but integrating we find that the conditional distribution function is,

$$F_{Y|X}(y|x) = \frac{1}{2} + \frac{1}{2} \frac{\text{sgn}(y) \sqrt{|y|}}{\sqrt{1 + |x| + |y|}}.$$

The conditional quantile functions are

$$(2.5) \quad Q_{Y|X}(\tau|x) = \frac{(2\tau - 1)^2}{1 - (2\tau - 1)^2} (1 + |x|)$$

so the conditional median function is $Q_{Y|X}(\frac{1}{2}|x) = 0$, though all the other conditional quantile functions are proportional to $1 + |x|$, and thus have a piecewise linear scale effects.

3. CONVERGENCE IN DISTRIBUTION FOR THE LAD ESTIMATOR.

Our first result establishes the inconsistency of the usual median (LAD) regression estimator.

Theorem 3.1. *Let $\hat{\beta}_n$ be the LAD estimator of β . Then without normalization,*

$$B_n \equiv (\hat{\beta}_n - \beta) \rightarrow^D B$$

where B is symmetric about zero, and $|B|$ has a density

$$(3.1) \quad f_{|B|}(b) = \frac{b^{-1/2}}{2(1+b)^{3/2}} g\left(\left(\frac{b}{1+b}\right)^{1/2}\right) \quad \text{for } b > 0.$$

Here, $g(u)$ is the function

$$(3.2) \quad g(u) \equiv \frac{2}{\pi} E(S(u))^{-1/2}$$

where $S(u)$ is a stable random variable that may be expressed as follows:

$$(3.3) \quad S(u) \sim \frac{2}{\pi} (R(.5, u) + u) \sim \lambda^2(u) \left(\frac{2u/(1+u^2) + \sin(P)}{2W \cos^2(P)} \right).$$

In the first representation $R(.5, u)$ denotes a stable random variable with index $\alpha = .5$ and skewness parameter, u , as computed by the *Splus* function **rstab** (see Chambers, et al., 1976). In the second representation, $\lambda^2(u) = (2/\pi)(1+u^2)$ arises from the scale parameter of the stable domain of attraction for the sum in (3.6); and $P \sim \text{Unif}(-\pi/2, \pi/2)$ while W has an independent negative exponential distribution with mean 1.

Equivalently, transforming the asymptotic distribution using $u = (b/(1+b))^{1/2}$, the density of $|U|$ is just $g(u)$ as given by (3.2). (This follows since the Jacobian of the transformation is just the first factor in (3.1).)

Proof. Begin by using the finite sample density for the LAD estimator as given in Koenker and Bassett (1978). This result assumes the design is fixed; so conditioning on $\mathbf{X} \equiv \{X_1, X_2, \dots, X_n\}$, and assuming $\beta = 0$, without loss of generality, we let $Z_i \equiv |Y_i/X_i|$ and

$$(3.4) \quad f_{|B_n|}(b | \mathbf{x}) = \sum_{i=1}^n f_{Z_i}(b | x_i) P_{x_i}\{|S_{i,n}(b)| \leq |x_i|\}$$

for $b > 0$, where (using (2.4)) the conditional density of Z_i given $X_i = x$ is

$$(3.5) \quad f_{Z_i}(b | x) = \frac{b^{-1/2} (1 + \frac{1}{x})}{2(1 + b + \frac{1}{x})^{3/2}}$$

and $S_{i,n}(b)$ is the sum in the gradient condition:

$$(3.6) \quad S_{i,n}(b) = \sum_{j \neq i} X_j \text{sgn}(Y_j - bX_j).$$

Using the i.i.d. sampling, and taking expectations with respect to the distribution of X ,

$$f_{|B_n|}(b) = nE f_z(b | X) P_x\{|\tilde{S}_n(b)| \leq |X|\},$$

where

$$\tilde{S}_n(b) = \sum_{i=1}^{n-1} X_i \text{sgn}(Y_i - bX_i)$$

as in (3.6). Now, break the expectation into the pieces $|X| < n^{1/2}$ and $|X| \geq n^{1/2}$. First consider the expectation when $|X| < n^{1/2}$. It is straightforward to show that the density of the summands $\tilde{S}_n(b)$ have the form: $c|x|^{-3/2}(1 + \mathcal{O}(1/x))$. Thus, $\{\tilde{S}_n(b)/n^2\}$ has a stable limit law. In fact, the density of the summands is can be expanded as $|x|$ times an expansion in powers of $1/|x|$ as $|x| \rightarrow \infty$. Therefore, the

densities of $\{\tilde{S}_n(b)/n^2\}$ converge uniformly to the (bounded) limit stable density (by Theorem 7.21 of Christoph and Wolf, 1992). As a consequence,

$$P\left\{|\tilde{S}_n(b)| < n^{1/2}\right\} = P\left\{|\tilde{S}_n(b)/n^2| < n^{-3/2}\right\} = \mathcal{O}(n^{-3/2}) .$$

Therefore the contribution for $|X| < n^{1/2}$ vanishes at rate $\mathcal{O}(n^{-1/2})$ (since the conditional density (3.5) in (3.4) is uniformly bounded in x). Now consider the expectation for $|X| \geq n^{1/2}$. Here, the conditional density satisfies

$$f_z(b|x) = \frac{b^{1/2}}{2(1+b)^{3/2}} (1 + \mathcal{O}(n^{-1/2})) \quad \text{for } b > 0 .$$

uniformly in b . Therefore, since the probability term is positive,

$$(3.7) \quad f_{|B_n|}(b) = \frac{b^{1/2}}{2(1+b)^{3/2}} (1 + \mathcal{O}(n^{-1/2})) n P\{|\tilde{S}_n(b)| \leq |X|\}$$

where the probability is now unconditional.

Now make the transformation $u = (|b|/(1+|b|))^{1/2}$. Since the first term in (3.7) is just du/db (apart from a factor of 2 introduced by the absolute value in the transformation), it remains only to consider the limit of n times the probability in (3.7). But, the c.d.f. for $|x|$ satisfies $1 - F(x) = (1/\pi) \tan^{-1}(x^{1/2})$, which has a leading term $1/(\pi x^{1/2})$. It follows with some calculation that we may write $S_n^*(u)$ for $\tilde{S}_n(b)$ and obtain for some constant c and for $u > 0$ (taking expectations over X),

$$n P\{|S_n^*(u)| \leq |X|\} = \frac{1}{\pi} E \left\{ |S_n^*(u)/n^2|^{-1/2} + c \max\{1, n^{-2}|S_n^*(u)/n^2|^{-3/2}\} \right\} .$$

But, as noted above, for fixed u , the densities for $\{S_n^*(u)/n^2\}$ converge uniformly to a stable density. Thus, using the dominated convergence theorem for $|S_n^*(u)/n^2| < 1$, the expectations converge and

$$(3.8) \quad f_U(u) \rightarrow^D \frac{1}{\pi} E |S^*(u)|^{-1/2} \quad \text{for } u > 0 ,$$

where $S^*(u)$ is a stable random variable with index $\alpha = .5$ and skewness parameter depending appropriately on u .

It remains to compute the appropriate skewness parameter and describe the expectation in (3.8). This requires the limiting values for $|v|^{1/2}F(v)$ and $|v|^{1/2}(1 - F(v))$ where $F(v)$ is the c.d.f. for the summands, $V \equiv X \operatorname{sgn}(Y - bX)$. A trite calculation using first the conditional distribution for Y given X yields (for $b > 0$):

$$\begin{aligned} F(v) &= \frac{1}{2} E \left\{ 1 + \left(\frac{b|X|}{(1+b)|X| + 1} \right)^{1/2} \right\} I(|X| > v) \\ &\approx \pi^{-1/2} \left\{ 1 + \left(\frac{b}{(1+b)} \right)^{1/2} \right\} \end{aligned}$$

for large b ; and similarly,

$$(1 - F(v)) \approx \pi^{-1/2} \left\{ 1 - \left(\frac{b}{(1+b)} \right)^{1/2} \right\}.$$

As above, letting $u = (b/(1+b))^{1/2}$ and using equation (1.19) of Christoph and Wolf, the stable skewness and scale parameters (β' and λ in their notation) satisfy

$$\begin{aligned} \lambda \sin((1 + \beta')\pi/4) &= (1 - u)/\pi^{1/2} \\ \lambda \sin((1 - \beta')\pi/4) &= (1 + u)/\pi^{1/2} \end{aligned}$$

Solving in terms of the more conventional skewness parameter, $\beta = 4 \tan^{-1}(\beta')/\pi$ with $\beta = u$ yields: $\lambda = (2(1 + u^2)/\pi)^{1/2}$. Tradition demands the use of β for stable skewness parameters despite the statistical use of β for regression parameters.

The first representation for the limiting stable distribution given in (3.3) follows from Chambers, *et al.* (1976) (see also, Nolan (1998)); while the second uses Zolotarev's representation (see Zolotarev, 1983, or Weron, 1996), where the trigonometric identity

$$\sin(x) + \sin(y) = 2 \sin\left(\frac{1}{2}(x + y)\right) \times \cos\left(\frac{1}{2}(x - y)\right)$$

is applied. \square

In Figure 1 we compare the finite sample behavior of the LAD and OLS estimators based on estimated densities for the transformed $\hat{v} = \sqrt{|\hat{\beta}|/(1 + |\hat{\beta}|)}$. The dotted line indicating the theoretical density of the LAD estimator gives a quite accurate account of the empirical estimate for each of the sample sizes considered. The transformed OLS density given by the dashed line in each plot is clearly less concentrated around zero than the LAD fit. Like the LAD estimator the OLS estimator converges in distribution without normalization as we will show in the next section.

4. CONVERGENCE IN DISTRIBUTION FOR THE LEAST SQUARES ESTIMATOR.

It is also rather easy to show that the least squares estimator converges in distribution without normalization. Unfortunately, this seems to require using convergence to a bivariate stable distribution, and the general bivariate stable has a characterization that involves a (nontrivial) integration over the circle and as far as we are aware does not admit a known method for simulation.

Theorem 4.1. *Let $\hat{\beta}_n^*$ be the least squares estimator of β . Then without normalization, $(\hat{\beta}_n^* - \beta)$ converges in distribution to a functional of a bivariate stable distribution.*

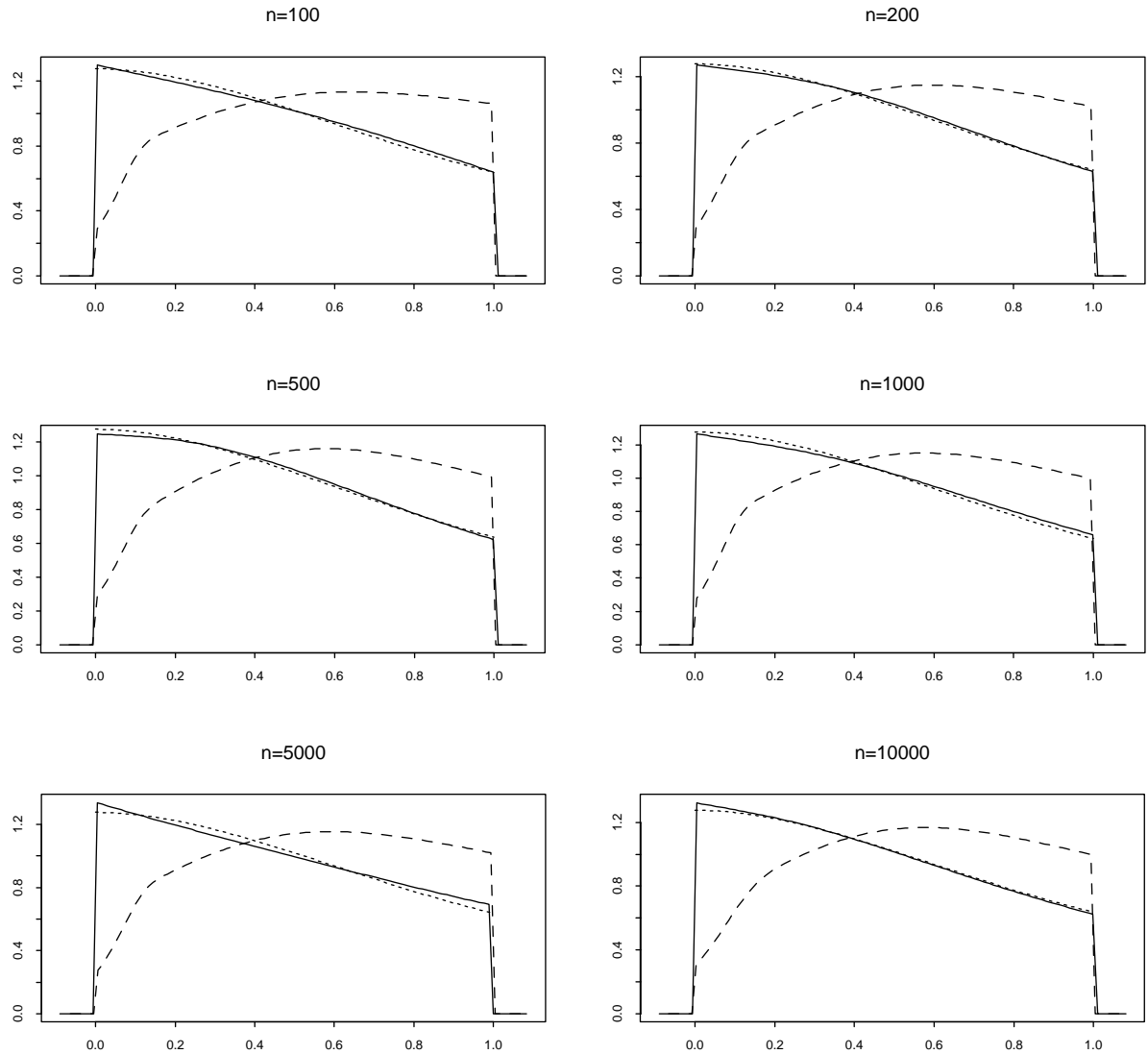


FIGURE 1. This figure illustrates the finite sample density of the LAD and OLS estimators after transformation. The dotted line indicates the theoretical density of the LAD estimator described in Theorem 1. The solid line is the fitted (transformed) LAD density based on 10,000 Monte-Carlo replications using the Stone and Kooperberg (1991) logspline method. The dashed line is the corresponding estimate of the OLS density.

Proof. The least squares estimator is

$$(4.1) \quad \hat{\beta}_n^* = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} .$$

Without loss of generality, take $\beta = 0$. Then, it suffices to consider the joint distribution of $U = X_i Y_i$ and $V = X_i^2$ and to show that the distribution of (U, V) lies in the domain of attraction of a bivariate stable distribution.

First, taking the joint distribution of (X_i, Y_i) in (2.2) and changing variables to (U, V) gives with some calculation,

$$(4.2) \quad f_{U,V}(u, v) = \frac{u^{-1/2} v^{-1/4}}{16\pi (v + u + v^{1/2})^{3/2}} \quad \text{for } u > 0, v > 0 .$$

Note that the density on the remainder of the domain is defined by symmetry. Now, changing to polar coordinates: $U = R \cos \theta$ and $V = R \sin \theta$,

$$(4.3) \quad f_{R,\theta}(r, t) = \frac{r^{-5/4} \cos(t)^{-1/2} \sin(t)^{-1/4}}{16\pi (\sin(t) + \cos(t) + \sin(t)^{1/2}/r^{1/2})^{3/2}} \quad \text{for } r > 0, 0 < t < \pi/2 .$$

As $r \rightarrow \infty$, this can be expanded in a series in terms of powers of $r^{-1/2}$. Thus, asymptotically, R and θ are independent; and the condition for a distribution to lie in a stable domain of attraction is trivial (see Rvačeva, 1962, p. 196). This means that the bivariate sums $(\sum_{i=1}^n X_i Y_i, \sum_{i=1}^n X_i^2)$ converge in distribution (in the plane) to a bivariate stable distribution. The convergence of the least squares estimator then follows immediately. \square

5. CONVERGENCE OF THE MAXIMAL DEPTH AND WEIGHTED LAD ESTIMATORS

Although the LAD estimator is inconsistent, it will now be shown that weighting may improve the situation considerably. Specifically, let $\check{\beta}_n$ be the weighted LAD estimator defined by

$$\check{\beta}_n = \arg \min_{b \in \mathbb{R}} \sum w(X_i) |Y_i - X_i b|$$

where w is a smooth weighting function.

Theorem 5.1. *Let X be a random variable with the distribuion of $|X_i|$ (2.3), and let $w(x)$ be a measurable weighting function for which*

$$V(w) \equiv E (X^2 w^2(X)) < +\infty .$$

Then, as $n \rightarrow \infty$,

$$n(\check{\beta}_n - \beta) \rightarrow^D vS$$

where S is a symmetric random variable with $|S| \sim \chi_1^2$, and v is a scale factor, $v \equiv \frac{M(w)}{\sqrt{V(w)}}$ with

$$(5.1) \quad M(w) \equiv E \left[\left(\frac{X}{1+X} \right)^{1/2} X w(X) \right].$$

Proof. The gradient for the weighted problem is

$$g_n(b) = - \sum \text{sgn}(Y_i - X_i b) x_i w(X_i)$$

and by the convexity of the objective function, hence the monotonicity of the gradient,

$$P(\hat{\beta}_n < \delta) = P(g_n(\delta) > 0).$$

For any sequence $\lambda_n \rightarrow \infty$,

$$P(\lambda_n \hat{\beta}_n < \delta) = P(g_n(\delta/\lambda_n) > 0) = P\left((g_n - M_n(w))/\sqrt{V_n(w)} > -M_n(w)/\sqrt{V_n(w)}\right),$$

where

$$M_n(w) = \sum (1 - 2F_i(X_i \delta/\lambda_n)) X_i w(X_i) \quad \text{and} \quad V_n(w) = \sum X_i^2 w^2(X_i).$$

It is easily checked that the weighted Bernoulli random variables, $\{(g_n - M_n(w))/\sqrt{V_n(w)}\}$ converge to a standard normal. Since $n^{-1}V_n g_n$ converges to $V(w)$, we have, letting $\lambda_n = n$,

$$\frac{M_n(w)}{\sqrt{V_n(w)}} = (V_n(w))^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\sqrt{X_i \delta/\lambda_n}}{\sqrt{1 + X_i + X_i \delta/\lambda_n}} X_i w(X_i) \rightarrow \frac{M(w)\sqrt{\delta}}{\sqrt{V(w)}}$$

Thus,

$$P\left(n|\hat{\beta}_n| < \delta\right) \rightarrow P\left(|Z| > -\sqrt{\delta} v\right) = P\left(Z^2 > \delta v^2\right)$$

where Z is a standard normal random variable. So $\hat{\beta}_n$ converges at rate $1/n$ to the symmetrically reflected, rescaled χ_1^2 distribution. \square

The improvement in the rate of convergence arises from the unboundedness of the conditional density at the median. To compare with the Monte-Carlo results of Rousseeuw, *et al.* (1999) and Table 1 above, we may compute the asymptotic approximation to the median absolute deviation of the estimator. That is, if $P(n|\hat{\beta}_n| < \delta_0) \rightarrow .5$, then $\delta_0 = \chi_{1,.5}^2/v^2$.

The optimal value for v follows from a straightforward application of Cauchy-Schwarz. Using the density for $|x_i|$ given by (2.3),

$$M(w) = \frac{1}{\pi} \int_0^\infty \frac{xw(x)}{(1+x)^{3/2}} dx \quad \text{and} \quad V(w) = \frac{1}{\pi} \int_0^\infty \frac{x^{3/2}w^2(x)}{(1+x)} dx.$$

Thus, by the Cauchy-Schwarz inequality,

$$(5.2) \quad M^2(w) \leq V(w) \times \frac{1}{\pi} \int_0^\infty \frac{\sqrt{x}}{(1+x)^2} dx = \frac{1}{2} V(w).$$

This inequality becomes an equality when the integrands for $M(w)$ and $V(w)$ are equal; that is, for $xw(x)/(1+x)^{3/2} = x^{3/2}w^2(x)/(1+x)$. Solving for $w^*(x)$ yields the following result.

Proposition 5.1. *Among all weight functions satisfying Theorem 5.1, the optimal weights and optimal scale factor are given by*

$$w^*(x) = (x(1+x))^{-1/2}; \quad v^* = 1/\sqrt{2}.$$

For this optimal weighting, the asymptotic mean absolute deviation is $\delta_0 = 2\chi_{1,5}^2 = .9099$.

We may contrast this optimally weighted median regression estimator with the even simpler weighting scheme $w_i = |x_i|^{-1}$, which yields the ordinary sample median, $\check{\beta}_n$, of the y_i/x_i 's as an estimator of β . This estimator is also the maximum depth estimator of Rousseeuw, *et al.* (1999). Here the $v_i = y_i/x_i$ are *iid* from the symmetrized $F_{1,1}$ distribution with distribution function,

$$F(v) = \frac{1}{2} + \frac{1}{\pi} \text{sgn}(v) \arctan(\sqrt{|v|/n}).$$

Since $\sqrt{n} \arctan \sqrt{|x|/n} \rightarrow \sqrt{|v|}$, for $\delta > 0$, $P(n\check{\beta}_n < \delta) \rightarrow P(Z > -\sqrt{\delta}2/\pi)$; we have

$$P(|n\check{\beta}_n| < \delta) \rightarrow P(Z^2 < 4\delta/\pi^2)$$

where Z is $\mathcal{N}(0,1)$. Again finding the asymptotic approximation for the median absolute deviation, we obtain,

$$\delta_0 = \chi_{1,5}^2 \pi^2 / 4 = 1.123.$$

The asymptotic relative efficiency of the two estimators is

$$\text{ARE} = .9099/1.123 = .81,$$

thus using the simpler weighting sacrifices almost 20% of the sample observations. Note that we have adopted the (Pitman) convention to measure ARE in this case as the ratio of scale parameters rather than variances since the convergence is at rate $1/n$ rather than the conventional $1/\sqrt{n}$. Thus, the ARE still reflects the limiting ratio of sample sizes required to achieve the same precision with the two procedures.

6. MAXIMUM LIKELIHOOD ESTIMATION

Having gone this far, it would be negligent to omit consideration of the maximum likelihood estimator. The log likelihood, obtained from (1.4), may be written as,

$$\ell(b) = K - \frac{1}{2} \sum_{i=1}^n [\log(|y_i - bx_i|/\sigma(x_i)) - 3 \log(1 + |y_i - bx_i|/\sigma(x_i))]$$

where $\sigma(x_i) = 1 + |x_i|$. But there is an immediate problem: $\ell(\beta)$ is unbounded at each of the sample points $b_i = y_i/x_i$. In the terminology of Ibragimov and Has'minskii(1981) the likelihood has a singularity of order $\alpha = -1/2$ at each of these

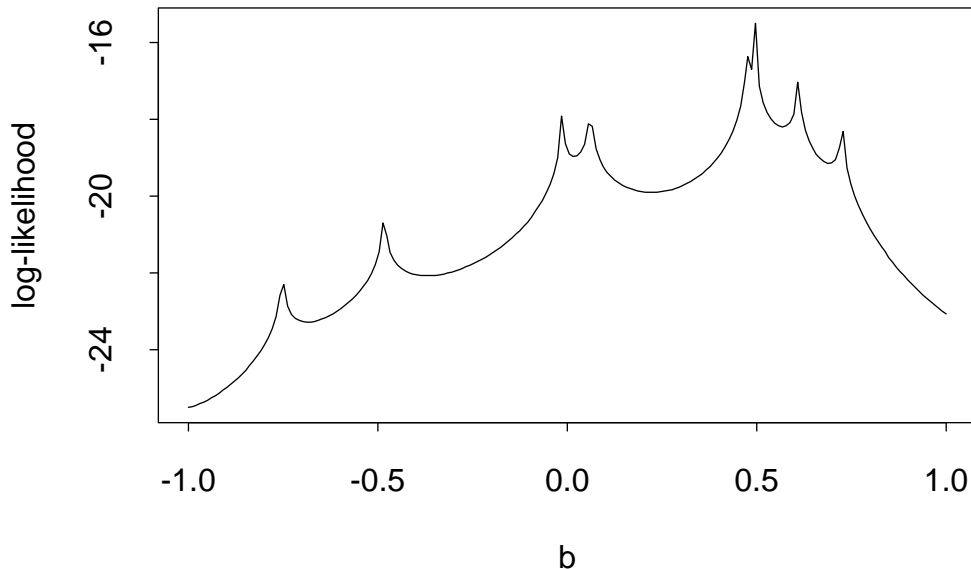


FIGURE 2. This figure illustrates the log likelihood function evaluated on a grid of 200 equally spaced points and based on a sample of size 21.

points. A typical log likelihood function is represented in Figure 2, where we have obviously avoided evaluation of the likelihood at any of the b_i 's, evaluating instead on a fairly fine grid. The MLE, which apparently expects an answer to the nonsensical question: “which infinity is largest?” can be sensibly defined as the b_i that maximizes $\ell(b_i + \varepsilon_n)$ over $i = 1, \dots, n$ for a suitably chosen sequence $\varepsilon_n \rightarrow 0$. Equivalently, we can simply define the leave-one-out likelihood at each b_i as

$$\tilde{\ell}(b_i) = K - \frac{1}{2} \sum_{j \neq i} [\log(|y_j - b_i x_j|/\sigma(x_j)) - 3 \log(1 + |y_j - b_i x_j|/\sigma(x_j))]$$

and maximize over $i = 1, \dots, n$. The value b^* at which $\tilde{\ell}(b_i)$ achieves its maximum will be called the jackknifed MLE (JMLE). For n large it is computationally helpful to observe that we can restrict attention to only the central order statistics of the b_i 's.

From Ibragimov and Has'minskii (1981, §VI.6) we know that there exist Bayes estimators that achieve convergence rate n^{-2} in this case. Thus, although the weighted median regression estimators seem to perform quite well, they all have asymptotic efficiency zero relative to the estimators of Ibragimov and Has'minskii (1981). However, it seems to be quite difficult to compute the Bayes estimators for the large sample

Estimator	n=20	n=50	n=100	n=200	n=500
$n\hat{\beta}_{MDE}$	1.31	1.19	1.17	1.18	1.11
$n^2\hat{\beta}_{JMLE}$	7.01	9.08	10.56	10.94	11.10

TABLE 2. Monte-Carlo Median Absolute Errors

sizes we have been using. In Table 2 we compare the performance of the JMLE with the MDE considered above based on 10,000 Monte-Carlo replications. We find that the performance of the JMLE is quite consistent with the convergence rate n^{-2} over the range of sample sizes employed in the experiment. Thus, the weighted median regression estimators all appear to have asymptotic efficiency zero relative to the JMLE.

References

- CHAMBERS, J, MALLOWS, C, AND STUCK, B. (1976). A method for simulating stable distributions, *J. Amer. Stat. Assoc.*, 71, 340-344.
- CHRISTOPH, G. AND WOLF, W. (1992). *Convergence Theorems with a Stable Law*, Akademie Verlag, Berlin.
- IBRAGIMOV, I.A. AND R.Z. HAS'MINSKII, (1981) *Statistical Estimation*, Springer: New York.
- KOENKER, R. AND BASSETT G. (1978). Regression Quantiles, *Econometrica*, 46, 33-50.
- KOENKER, R. (1999). Regression depth: Comment, *J. Amer. Stat. Assoc.*, 94, 417-419.
- KOOPERBERG, CHARLES AND STONE, CHARLES J., (1991). A Study of Logspline Density Estimation, *Computational Statistics and Data Analysis*, 12, 327-347,
- NOLAN, J. (1998). Parametrizations and modes of stable distributions, *Prob. Stat. Letters*, 38, 187-195.
- ROUSSEEUW, P. AND HUBERT, M. (1999). Regression depth, *J. Amer. Stat. Assoc.*, 94, 388-402.
- ROUSSEEUW, P., VAN AELST, S., AND HUBERT, M. (1999). Regression depth: Rejoinder, *J. Amer. Stat. Assoc.*, 94, 419-434.
- RVAČEVA, E. L. (1962). On domains of attraction of multi-dimensional distributions, *Selected Translations in Mathematical Statistics and Probability*, 2 (Amer. Math. Soc., Providence, RI.), 183-205.
- WERON, R. (1996). On the Chambers - Mallows - Stuck method for simulating skewed stable random variables, *Prob. Stat. Letters*, 28, 165-171.
- ZOLOTAREV, V. (1983). *One-dimensional Stable Distributions, Translations of Mathematical Monographs*, Vol. 65, American Mathematical Society, Providence, Rhode Island.

UNIVERSITY OF ILLINOIS, URBANA-CHAMPAIGN