BEBR

FACULTY WORKING PAPER NO. 1487

Tail Behavior of Regression Estimators and their Breakdown Points

Xuming He
Jana Jureckova
Roger Koenker
Stephen Portnoy

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois, Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 1487

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

August 1988

Tail Behavior of Regression Estimators and Their Breakdown Points

Xuming He
Department of Statistics

Jan Jureckova
Charles University, Czechoslovakia

Roger Koenker, Professor
Department of Economics

Stephen Portnoy
Department of Statistics

This research was initiated while Jureckova was Miller Visiting Professor at the University of Illinois. Support of NSF Grants SES-8707169 and KMS 88-02555 is also gratefully acknowledged.

TAIL BEHAVIOR OF REGRESSION ESTIMATORS AND THEIR BREAKDOWN POINTS*

Xuming He
Department of Statistics
University of Illinois
Champaign, Illinois 61820

Jana Jurečková
Department of Probability and Statistics
Charles University
Prague 8, Czechoslovakia

Roger Koenker
Department of Economics
University of Illinois
Champaign, Illinois 61820

Stephen Portnoy
Department of Statistics
University of Illinois
Champaign, Illinois 61820

August 1988

ABSTRACT

Following Jurečková (1981) we introduce a finite-sample measure of performance of regression estimators based on tail behavior. The least squares estimator is studied in detail, and we find that it may achieve good tail performance under strictly Gaussian conditions. However, the tail performance of the least-squares estimator is found to be extremely poor in the case of heavy-tailed error distributions or when leverage points are present. Further analysis of the least-squares estimator with light-tailed errors indicates the strong influence of the design matrix in determining tail performance.

Turning to the tail behavior of various robust estimators of the parameters of the linear model, we focus on tail performance under heavy (algebraic) tailed errors. The l_1-estimator is seen to be a leading case: we find a simple characterization of its tail behavior in terms of the design configuration and show that a broad class of M-estimators have the same performance.

Perhaps most significantly, it is shown that our finite-sample measure of tail performance is, for heavy tailed error distributions, essentially the same as the finite sample concept of breakdown point introduced by Donoho and Huber(1983). This finding provides an important probabilistic interpretation of the breakdown point and clarifies the role of tail behavior as a quantitative measure of robustness. This link is further explored for high-breakdown regression estimators including Rousseeuw's (1982) least-median-of-squares estimator.

---

* This research was initiated while Jurečková was Miller Visiting Professor at the University of Illinois. Support of NSF Grants SES-8707169 and DMS 88-02555 is also gratefully acknowledged.

1. Introduction

Several authors, including Bahadur(1967) and Sievers(1978), have studied measures of performance for estimators $T_n = T_n(X_1,...,X_n)$ of a location parameter, θ, based on the tail probability $P_\theta(|T_n-\theta|>a)$ for a fixed a as $n\rightarrow\infty$. In the location model with independent observations $X_1,\ldots,X_n$ from an absolutely continuous, symmetric distribution function $F(x-\theta)$ with density $f(z) = f(-z) > 0$, $z\in\mathbf{R}^1$, Jurečková (1981) considered the measure of performance

$$B(a,T_n) = \frac{-\log P_\theta(|T_n-\theta|>a)}{-\log(1-F(a))} \tag{1.1}$$

for fixed n as $a\rightarrow\infty$, and showed that for any [reasonable] translation equivariant T_n,

$$1 \leq \liminf_{a\rightarrow\infty} B(a,T_n) \equiv \underline{B} \leq \limsup_{a\rightarrow\infty} B(a,T_n) \equiv \overline{B} \leq n. \tag{1.2}$$

Further, it was shown that both bounds can be attained by the sample mean $T_n = \overline{X}_n$. For F's with exponential tails, $\overline{X}_n$ attains optimal tail performance with the log of the probability of $\{|T_n-\theta|>a\}$ tending to zero n times faster than the log of the probability that a single observation from F exceeds a. While for F's with algebraic tails, the probability that $|\overline{X}_n-\theta|>a$ tends to zero no faster than the probability of a single observation from F exceeding a. This striking lack of robustness of $\overline{X}_n$, i.e. the sensitivity of its tail performance to the tail behavior of F, leads to the question: Are there other estimators with good tail behavior over a broad class of possible F's?

For L-estimators of location of the form,

$$T_n = \sum_{i=1}^{n} c_i X_{(i)}$$

where $X_{(1)},...,X_{(n)}$ are the order statistics of the random sample, $X_1,\cdots,X_n$; $c_i\geq 0$, $i=1,...,n$; $\sum c_i=1$; and $c_i=c_{n-i+1}=0$ for $i=0,1,...,k$, $0\leq k<n/2$. Jurečková (1979) proved that

$$k+1\leq\liminf_{a\rightarrow\infty} B(a,T_n)\leq\limsup_{a\rightarrow\infty} B(a,T_n)\leq n-k. \tag{1.3}$$

Thus for the sample median with n odd, $\lim_{a\rightarrow\infty} B(a,T_n) = (n+1)/2$. Furthermore, for any

Huber-type M-estimator, T_n, (defined as a solution to the equation $\sum_{i=1}^{n}\psi(X_i - t)$ for a nondecreasing, odd function ψ such that $\psi(x)=\psi(k)$ for $x \geq k$, for some $k>0$), has the same tail performance as the median for F's with either exponential or algebraic tails. Note that this holds for Huber estimators with fixed scale--a sufficiently poor estimator of scale could wreck havoc here.

The foregoing results seem to suggest a relationship between tail performance of location estimators and the finite-sample version of breakdown point of estimators introduced by Donoho and Huber(1983). The latter concept, originally suggested by Hampel (1968), has played a central role in recent work on robust estimation and testing since it provides an appealing yet tractable quantitative assessment of robustness. The finite-sample replacement version of the breakdown point ϵ_n^* of an estimator T_n is the following. Let $x^0 = (x_1, \ldots, x_n)$ denote an initial sample, and let x^m be a "contaminated" sample constructed by replacing m arbitrary elements of x^0 with arbitrary values. The breakdown point of T_n at the sample x^0 is then, $\epsilon_n^* = m^*/n$, where m^* is the least integer m such that $\sup_{x^m}\|T_n(x^m) - T_n(x^0)\| = \infty$, i.e. the smallest number of observations which, if replaced by arbitrary values, could drive T_n beyond all bounds. The following result clarifies the relationship between tail performance and breakdown point for a large class of location estimators.

THEOREM 1.1. *Suppose $T_n(X_1, \ldots, X_n)$ is a location equivariant estimator of θ such that T_n is nondecreasing in each argument X_i. Then for any symmetric, absolutely continuous F, with density $f(z) = f(-z) > 0$, $z \in \mathbb{R}^1$.*

$$m^* \leq \liminf_{a \to \infty} B(a, T_n) \leq \limsup_{a \to \infty} B(a, T_n) \leq n - m^* + 1$$

The key point of the proof is the following:

LEMMA 1.1. *Under the conditions of Theorem 1.1, there exists a constant A such that*

$$x_{(m^*)}-A \le T_n \le x_{(n-m^*+1)}+A$$

Proof. By equivariance,

$$T_n(x_1,\ldots,x_n) = T_n(x_{(1)}-x_{(m)},\ldots,x_{(m-1)}-x_{(m)},0,+,+,\ldots,+) + x_{(m)}$$

$$\ge x_{(m)}+T_n(x_{(1)}-x_{(m)},\ldots,x_{(m-1)}-x_{(m)},0,0,0,\ldots,0)$$

where $m=m^*$ and + denotes any positive number. By the definition of m^*, $|T_n|$ with only m^*-1 possible outliers is bounded, say by A. Hence, $T_n>x_{(m)}-A$. The other inequality may be established similarly. ■

Proof. (of Theorem 1.1.) From the lemma,

$$P_\theta(T_n-\theta>a) = P_0(T_n>a)$$

$$\ge P_0(x_{(m)}>a+A)$$

$$\ge (1-F(a+A))^{n-m-1}.$$

Thus,

$$B(a,T_n) = \frac{-\log 2P_0(T_n>a)}{-\log(1-F(a))} \le (n-m+1)\frac{\log 2+\log(1-F(a+A))}{\log(1-F(a))}.$$

Letting $a\to\infty$ we get $\limsup_{a\to\infty} B(a,T_n)<n-m+1$. On the other hand $\liminf_{a\to\infty} B(a,T_n)\ge m$ follows from,

$$P_0(T_n>a) \le P_0(x_{n-m+1}>a+A) = \binom{n}{m}(1-F(a+A))^m. \qquad ■$$

Remarks: Note that Theorem 2.1 holds for both exponential and algebraic tails of F. The conditions on T_n are satisfied for M-estimators with monotone ψ, L-estimators with positive weight functions, but not, for example for redescending M-estimators or the least median of squares. If T_n has high breakdown, like the sample median, with $\epsilon_n^*\to\frac{1}{2}$, then $\limsup_{a\to\infty} B(a,T_n) \le n/2$ for either error distribution type. This seems to suggest that highly robust methods necessarily sacrifice performance in light tailed circumstances. Finally, the high breakdown estimators are seen to satisfy a minimax property: they maximize least favorable tail performance over the two distribution types described in Jurečková (1981) and defined in the next section.

The intimate connection between finite-sample breakdown and finite-sample tail performance is developed further in subsequent sections. Section 2 introduces a measure of tail performance for regression estimators and discusses some possible alternative measures. Section 3 contains a detailed analysis of the tail performance of the least squares estimator. Section 4 investigates the tail behavior of various robust estimators of the parameters of the linear model. Some concluding remarks on the relationship between breakdown and tail performance appear in the final section.

2. A Measure of Tail Performance for Regression Estimators

We now turn to the linear model,

$$Y = X\beta + e \tag{2.1}$$

where Y is an n vector of random responses, X is a *fixed* $(n \times p)$ design matrix of rank p, with rows $x_i{:}i=1,\dots,n$, β is an unknown p-dimensional parameter vector and e is an n-vector of independent errors, with common absolutely continuous distribution function F. Assume throughout that

$$0<F(z)<1, \quad F(z)+F(-z) = 1, \quad z \in \mathbf{R}^1. \tag{2.2}$$

We wish to estimate the parameter β without precise knowledge of the shape of F. Consider estimators T_n of β satisfying the affine (or regression) equivariance condition,

CONDITION A. $T_n(Y_1+x_1 b, \cdots, Y_n+x_n b) = T_n(Y_1, \dots, Y_n)+b, \quad b \in \mathbf{R}^p.$

To extend the measure of tail performance (1.1) used for scalar location estimators to the p-dimensional regression context we propose

$$B(a,T_n) = \frac{-\log P_\beta(\max_i |x_i(T_n-\beta)|>a)}{-\log(1-F(a))}. \tag{2.3}$$

Obviously, we would like the probability in the numerator (of a discrepancy of a between some $\hat{Y}_i$ and EY_i) to tend to zero as quickly as possible as $a \to \infty$. But, as in the case of location estimators, this rate is inherently controlled by the tail behavior of the error observations. Thus we should expect from good estimators that $B(a,T_n)$ is reasonable high, but since we can not hope to control $B(a,T_n)$ for all a, we study the tail behavior of T_n, i.e. the limiting behavior of $B(a,T_n)$ as $a \to \infty$.

It is crucial in the sequel to distinguish between two broad classes of tail behavior for the underlying error distribution F. Following Jurečková (1981), a distribution F will be called Type I (exponentially tailed) if

$$\lim_{a\to\infty} \frac{-\log(1-F(a))}{ba^r} = 1$$

for some $b>0$, and $r>0$; and it will be called Type II (algebraically tailed) if

$$\lim_{a\to\infty} \frac{-\log(1-F(a))}{m\log a} = 1$$

for some $m>0$.

The tail performance bound (1.2) extends to the regression model if we impose a mild further condition on T_n that there exists at least one nonnegative and one nonpositive residual $r_i = Y_i - x_i T_n$, $i = 1, \cdots, n$. Under this condition, $y_{(1)} > a$ implies $\hat{y}_{(n)} > a$ and $y_{(n)} < -a$ implies $\hat{y}_{(n)} < -a$, thus, using the symmetry of F,

$$\begin{aligned} P_0(|\hat{y}|_{(n)} > a) &\geq 2P_0(y_{(1)} > a) \\ &= 2(1 - F(a))^n \end{aligned}$$

and hence

$$\limsup_{a\to\infty} B(a, T_n) \leq n .$$

However, as subsequent results will illustrate, achieving this upper bound may be limited to the case of the sample mean in the location model with exponential tails. A lower bound of 1 may be derived for errors of Type II, under somewhat more stringent conditions, but for F of Type I, it is easy to construct examples for which limsup $B(a, T_n)$ is less than 1. In fact, consider a simple linear regression with $n=3$, $x_1=0$, $x_2=1$, and $x_3=2$, and let T_n be determined by the first two observations. Then, if $|y_1| \leq \epsilon$ and $|y_2| \geq b$ it follows that $\max_i |x_i T_n| = x_3|T_n| \geq 2b/\epsilon$. Hence, $P(\max|x_i T_n| \geq a) \geq P(|y_1| \leq \epsilon)\, P(|y_2| \geq a\epsilon/2) \geq cP(y_2 \geq a\epsilon/2)$. So, for Type I distributions, $\bar{B} \leq (\epsilon/2)^r$.

The measure of performance (2.3) is only one of many possible criteria of tail performance for the p-dimensional regression estimator, and one is naturally led to ask whether results are sensitive to the specific form of the criteria. This question is partially answered by the following:

THEOREM 2.1 *Let* Γ *be the class of all* $\gamma = \gamma_X : \mathbf{R}^p \to \mathbf{R}$ *which are strictly positive (for non-zero arguments), continuous, and linearly homogeneous (i.e.,* $\gamma(cb) = c\gamma(b)$ for $c>0$ *). For any* $\gamma \in \Gamma$, *define the tail criterion*

$$B_\gamma(T_n, a) = \frac{-\log P\{\, \gamma(T_n - \beta) \geq a \,\}}{-\log(1 - F(a))} \tag{2.4}$$

and define $\underline{B} = \liminf_{a\to\infty} B(a, T_n)$ *and* $\bar{B} = \limsup_{a\to\infty} B(a, T_n)$. *Then for any* $\gamma_1 \in \Gamma$ *and* $\gamma_2 \in \Gamma$, *there is a constant,* c, *such that for all* b, $\gamma_1(b) \leq c\gamma_2(b)$. *As a consequence, for any Type II distribution,*

$$\underline{B}_{\gamma_1} = \underline{B}_{\gamma_2} \quad \textit{and} \quad \bar{B}_{\gamma_1} = \bar{B}_{\gamma_2}$$

Proof. For $\gamma \in \Gamma$, define

$$\bar{c}(\gamma) = \sup_{\|b\|=1} \gamma(b) \;<\; +\infty$$

$$\underline{c}(\gamma) = \inf_{\|b\|=1} \gamma(b) \;>\; 0\,.$$

Then, for any γ_1 and γ_2,

$$\gamma_1(b) \le \bar{c}(\gamma_1)\,\|b\| \le \frac{\bar{c}(\gamma_1)}{\underline{c}(\gamma_2)}\,\gamma_2(b) = c^*\,\gamma_2(b)$$

where $c^* = \bar{c}(\gamma_1)/\underline{c}(\gamma_2)$; and, hence,

$$P\{\, \gamma_1(T_n - \beta) \ge a \,\} \le P\{\, \gamma_2(T_n - \beta) \ge c^* a \,\} .$$

As a consequence, letting $a^* = c^* a$,

$$B_{\gamma_1}(T_n, a) \ge B_{\gamma_2}(T_n, a^*) \left(\frac{-\log P\{e_i \ge a^*\}}{-\log P\{e_i \ge a\}} \right) . \tag{2.5}$$

The result follows since the roles of γ_1 and γ_2 can be interchanged and since the limit of the ratio of logs of probabilities in (2.5) tends to one for type II distributions. ■

Remark. For type I distributions, the ratio of logarithms of probabilities in (2.5) tends to $(c^*)^r$, and so tail behavior depends on the choice of γ .

For γ defined in terms of residuals, it is possible to define a corresponding M-estimator whose tail behavior is determined by one-dimensional tail behavior. Using such an estimator, one can obtain tail behavior with $\underline{B} \ge [(n-p+1)/2]$ for type II distributions. Such robustness results will discussed further in Section 4.

LEMMA 2.1. *Consider the criteria* γ *defined by* $\gamma_X(b) = \gamma^*(|Xb|) \equiv \gamma^*(|x_1 b|, \cdots, |x_n b|)$ *where* $\gamma^*: R_+^n \to R_+$. *Let* γ^* *and* $\tilde{\gamma}$ *be two such functions satisfying*

(i) $\gamma^*(r)$ *is non-decreasing in each coordinate of* r ,

(ii) $\gamma^*(r + s) \le c\,(\tilde{\gamma}(r) + \tilde{\gamma}(s))$ *for some constant* $c = c(X)$,

(iii) There exists an estimator T_n^* *minimizing* $\tilde{\gamma}(\,|Xb - Y|\,)$ *over* b .

Then, for any type II distribution,

$$P\{\, \gamma^*(|X(T_n^* - \beta|) \ge a \,\} \le P\{\, \tilde{\gamma}(|e|) \ge \frac{a}{2c} \,\} .$$

Proof: Using the triangle inequality and the conditions above,

$$P\{\gamma^*(|X(T_n^* - \beta)|) \geq a\} \leq P\{\gamma^*(|XT_n^* - Y)| + |X\beta - Y|) \geq a\}$$

$$\leq P\{c\,\tilde{\gamma}(|XT_n^* - Y)| + c\,\tilde{\gamma}(|X\beta - Y|) \geq a\}$$

$$\leq P\{2c\,\tilde{\gamma}(|X\beta - Y)|) \geq a\}$$

$$= P\{\tilde{\gamma}(|e|) \geq \frac{a}{2c}\} . \quad \blacksquare$$

This result can be used to show that there exists an estimator achieving $\underline{B} \geq [(n-p+1)/2]$ in type II cases. This bound may be the best possible in general, and it corresponds exactly to the best breakdown bound possible for affine invariant estimators (see Rousseeuw and LeRoy (1987, p. 125)).

COROLLARY. *Let* $\gamma^*(r) = r_{(p+1)}$ *(the* $(p+1)^{st}$ *smallest* r_i*), and with* $k = [(n-p+1)/2]$ *let* $\tilde{\gamma} = r_{(n-k+1)}$ *(so that* T_n^* *minimizes the corresponding* k^{th} *largest absolute residual). Then* $\underline{B}(T_n^*) \geq k$.

Proof. Condition (i) is clear. For condition (ii), note that if $(r+s)_{(p+1)} = a$ then $r_i \geq a/2$ or $s_i \geq a/2$ for at least (n-p) indices. Whence, either $r_i \geq a/2$ for k indices or $s_i \geq a/2$ for k indices. Therefore,

$$(r+s)_{(p+1)} \leq 2(r_{(n-k+1)} + s_{(n-k+1)}) .$$

Condition (iii) follows from general results on S-estimators, so the Lemma holds. Furthermore, for some constant c ,

$$P\{\,|e|_{(n-k+1)} \geq a\,\} \leq c\,(P\{e_i \geq a\})^k \quad ,$$

and the result follows from (2.4) and Theorem 2.1. $\blacksquare$

3. Tail Behavior of the Least Squares Estimator

The tail performance of the classical least-squares estimator suffers from the same sensitivity to the tail behavior of F as its counterpart from the location model the sample mean.

THEOREM 3.1 *Let T_n be the least squares estimator of β in the model (2.1) with F satisfying condition (2.2). Let* $\bar{h} = \max_i h_{ii} = \max_i x_i(X'X)^{-1}x_i'$.

(i) If F is of Type I with $1<r\leq 2$, then

$$\bar{h}^{1-r} \leq \liminf_{a\to\infty} B(a,T_n) \leq \limsup_{a\to\infty} B(a,T_n) \leq \bar{h}^{-r}.$$

(ii.) If F is of Type I with $r=1$, then

$$\bar{h}^{-1/2} \leq \liminf_{a\to\infty} B(a,T_n) \leq \limsup_{a\to\infty} B(a,T_n) \leq \bar{h}^{-1}.$$

(iii.) If F is normal, then

$$\lim_{a\to\infty} B(a,T_n) = \bar{h}^{-1}.$$

(iv.) If F is of type II, then

$$\lim_{a\to\infty} B(a,T_n) = 1.$$

Proof. Let $H = X(X'X)^{-1}X'$ be the projection (hat) matrix corresponding to X, and suppose that $\bar{h} = h_{11}$. Recall that $0<\bar{h}\leq 1$ and $\hat{Y}_i = x_i T_n = h_i' Y$, so we may write,

$$P_\beta(\max_i |x_i(T_n-\beta)|>a) = P_0(\max_i |h_i' Y|>a)$$

$$\geq P_0(h_1' Y>a) \geq P_0(\bar{h}Y_1>a, h_{12}Y_2\geq 0, \cdots, h_{1n}Y_n\geq 0)$$

$$\geq P_0(Y_i>a/\bar{h})(\tfrac{1}{2})^{n-1} = (1-F(a/\bar{h}))(\tfrac{1}{2})^{n-1}.$$

Hence,

$$\limsup_{a\to\infty} B(a,T_n) \leq \limsup_{a\to\infty} \frac{-\log(1-F(a/\bar{h}))}{-\log(1-F(a))}. \tag{3.1}$$

for F of Type I this further implies,

$$\limsup_{a\to\infty} B(a,T_n) \leq \limsup_{a\to\infty} \frac{b(a/\bar{h})^r}{ba^r} = \bar{h}^{-r}, \tag{3.2}$$

which gives the upper bounds in (i) and (ii), respectively. For F of Type II, (3.1) implies,

$$\limsup_{a\to\infty} B(a,T_n) \le \limsup_{a\to\infty} \frac{m\log(a/\bar{h})}{m\log a} = 1. \tag{3.3}$$

And, since the least-squares estimator has at least one positive and one negative residual, (iv) follows.

On the other hand, if F is of type I with $1<r\le 2$, then, by Markov's inequality for any $\epsilon\in(0,1)$,

$$P_\beta(\max_i |x_i(T_n-\beta)|>a) = P_0(\max_i |\hat{Y}_i|>a)$$

$$\le E_0\exp\{(1-\epsilon)b\bar{h}^{1-r}(\max_i|\hat{Y}_i|)^r\}\exp\{-(1-\epsilon)b\bar{h}^{1-r}a^r\}, \tag{3.4}$$

so if,

$$0 < E_0\exp\{(1-\epsilon)b\bar{h}^{1-r}(\max_i|\hat{Y}_i|)^r\} \le C_\epsilon < \infty \tag{3.5}$$

then

$$-\log P_0(\max_i|\hat{Y}_i|>a) \ge -\log C_\epsilon + (1-\epsilon)b\bar{h}^{1-r}a^r$$

and the lower bound in (i) follows. To prove (3.5) we may write

$$(max_i|\hat{Y}_i|)^r = \max_i |h_iY|^r \le \max_i(\|h_i\|_s\|Y\|_r)^r$$

$$\le \max_i(\textstyle\sum h_{ik}^2)^{r/s}\sum|Y_k|^r \le \bar{h}^{r-1}\sum|Y_k|^r$$

where $s = r/(r-1) > 2$. Hence,

$$E_0\exp\{(1-\epsilon)b\bar{h}^{1-r}(\max_i|\hat{Y}_i|)^r\} \le E_0\exp\{(1-\epsilon)b\textstyle\sum|Y_k|^r\}$$

$$\le (E_0\exp\{(1-\epsilon)b|Y_1|^r\})^n. \tag{3.6}$$

Using an integration by parts,

$$0 < E_0\exp\{(1-\epsilon)b|Y_1|^r\} = -2\int_0^\infty \exp\{(1-\epsilon)by^r\}d(1-F(y))$$

$$\leq 2\int_0^K \exp\{(1-\epsilon)by^r\}dF(y) + 2\exp\{(1-\epsilon)bK^r\}(1-F(K))$$

$$+ 2\int_K^\infty r(1-\epsilon)by^{r-1}(1-F(y))\exp\{(1-\epsilon)by^r\}dy$$

$$\leq 2\int_0^K \exp\{(1-\epsilon)by^r\}dF(y) + 2(1-F(K))\exp\{(1-\epsilon)bK^2\}$$

$$+ 2\int_K^\infty r(1-\epsilon)by^{r-1}\exp\{-(\epsilon/2)by^r\}dy \leq C_\epsilon < \infty.$$

for K such that $1-F(y) \leq \exp(-(1-\epsilon/2)by^r)$ for $y > K$. This gives the lower bound in (i) and (iii), respectively. Analogously, if $r=1$, then,

$$P_0(\max_i |\hat{Y}_i|>a) \leq E_0\exp\{(1-\epsilon)b\bar{h}^{-1/2}\max_i |\hat{Y}_i|\}\exp\{-(1-\epsilon)b\bar{h}^{-1/2}a\}$$

and

$$\max_i |\hat{Y}_i| = \max_i |h_i Y| \leq \bar{h}^{-1/2}\sum |Y_k|.$$

which gives the lower bound in (ii). Finally, if F is normal $N(0,\sigma^2)$ then $\hat{Y}-X\beta$ has an n-dimensional normal distribution $N(0,\sigma^2 H)$. Hence,

$$P(\max_i |\hat{Y}_i| > a) \geq P_0(h_1 Y>a) = 1-\Phi(a\sigma^{-1}\bar{h}^{-1/2})$$

and this give the upper bound in (iii). ∎

Remarks. In the one-sample location model $\bar{h} = 1/n$ and thus (iii) specializes to Theorem 2.2 of Jurečková (1981) but in the linear model with leverage points $\bar{h}$ may be near one and consequently tail performance would be very poor. On the other hand, for F of Type II, the tail behavior of the LSE is always extremely poor. In effect, its tail performance with n observations is no better than that with a single observation, that is, here the tail performance is the same as the breakdown. In the balanced p-sample problem, $\bar{h} = p/n$, while for stochastic designs $\bar{h} = O_p(p/n)$ under some regularity conditions, so the tail-performance bound n/p

in the Gaussian case is suggested. In heavy tailed cases much better performance is possible with robust methods of estimation to which we now turn our attention.

4. Tail Performance of Robust Estimators of the Linear Model

Given the poor tail performance of the least-squares estimator under heavy tailed (Type II) conditions, it is useful to know whether better performance is possible from various robust estimators. Since any amount of Type II contamination of a Type I distribution will yield a Type II distribution, (Jurečková (1981)), the tail performance of regression estimators under Type II conditions appears to be an useful quantitative assessment of robustness.

An upper bound on the tail performance of L-estimators of the regression parameter based on regression quantiles (Koenker and Bassett(1978) and Koenker and Portnoy(1987)) is based on the following result.

LEMMA 4.1 *Suppose in the linear model (2.1), the design matrix, X, contains an intercept, i.e. $Xb = 1_n$ for some b, and the error distribution F satisfies (2.2), then for any θ^{th} regression quantile, $\bar{B} \le [\min\{\theta, 1-\theta\}n] + 1$.*

Proof. If $\max_i |\hat{y}_i| < |y|_{(k-1)}$, then there exist at least k strictly positive or k strictly negative residuals, but by Theorem 3.4 of Koenker and Bassett(1978), $k \le \min\{\theta, 1-\theta\}n$. The result then follows for either type of error distribution. ■

THEOREM 4.1. *Let T_n be an estimator of the form,*

$$T_n = \int_{\alpha_0}^{\alpha_1} J(t)\hat{\beta}(t)dt$$

where $\hat{\beta}(t)$ is the regression quantile process, and J is a non-negative function which integrates to one on $[\alpha_0, \alpha_1] \subset (0,1)$. If F is either Type I or Type II, then $\limsup_{a \to \infty} B(a, T_n) \le [\min\{\alpha_0, 1-\alpha_1\}n] + 1$.

Proof. By the convexity of the measure of performance $\gamma(b) = \max_i |x_i b|$ we have $P(\gamma(T_n) > a) \le P(\max_j \gamma(\hat{\beta}_j) > a)$ where the max is over the distinct regression quantiles on the interval $[\alpha_0,\alpha_1]$. The result then follows from the preceeding lemma. ■

Thus we find that the tail performance of the l_1 estimator may be as good as $(n+1)/2$, as for example in the case of sample median. However, the lower bound on the tail performance of the l_1 estimator is much more informative. To pursue this, we define m_* to be the largest integer m such that for any subset M of $N = \{1,2,...,n\}$ of size m,

$$\inf_{\|b\|=1} \frac{\sum_{i\in N\setminus M} |x_i b|}{\sum_{i\in N} |x_i b|} > ½.$$

($N\setminus M$ will denote the complement of M with respect to N.) In the special case of scalar regression through the origin m_*+1 is the smallest integer k such that for some subset K, of size k, $\sum_{i\in K} |x_i| > ½ \sum_{i\in N} |x_i|$. When the x_i's are equally spaced on [0,1], for example, this m_*/n tends to $1-1/\sqrt{2} \cong .29289$ as $n\to\infty$.

LEMMA 4.2 *For the l_1 estimator, $\underline{B} \geq m_*+1$ for F of Type II.*

Proof It suffices to show that if all but $m \leq m_*$ of the y's are bounded by 1, then the l_1-estimator will be uniformly bounded. By the triangle inequality,

$$\sum_N |y_i - x_i b| \geq \sum_{N\setminus M} |x_i b| - \sum_{N\setminus M} |y_i| - (\sum_M |x_i b| - \sum_M |y_i|)$$

$$= \sum_{N\setminus M} |x_i b| - \sum_M |x_i b| + \sum_N |y_i| - 2\sum_{N\setminus M} |y_i|.$$

By the definition of m_*, there exists $c>½$ such that $\sum_{N\setminus M} |x_i b| \geq c\sum_M |x_i b|$ and $\sum_M |x_i b| \leq (1-c)\sum_N |x_i b|$ for all subsets M of size $m \leq m_*$. Therefore,

$$\sum_N |y_i - x_i b| - \sum_N |y_i| \geq (2c-1)\sum_N |x_i b| - 2\sum_{N\setminus M} |y_i|$$

So if $|y_i| \leq 1$ for $i\in N\setminus M$, there exists a constant C such that if $\|b\| > C$ then $\sum_N |y_i - x_i b| - \sum_N |y_i| \geq 0$.

The proof may be completed by noting that the l_1 estimator is scale-equivariant so $\max_i |x_i T_n| \le C\,|y|_{(n-m_*)}$ where $|y|_{(n-m_*)}$ is the m_*-th largest absolute y. ■

Lemma 4.2 may be extended to a broad class of "l_1-*type*" estimators characterized by ρ functions such that for some $K>0$ and all u,

$$\left|\ \rho(u) - |u|\ \right| \le K \ . \tag{4.1}$$

In fact, all such l_1-*type* estimators have exactly the same tail behavior for a wide class of tail criteria and for both type I and type II distributions.

LEMMA 4.3. *Let* $\hat\beta$ *be an M-estimator minimizing* $\sum_{i=1}^{n} \rho(y_i - x_i b)$, *where* ρ *satisfies (4.1). Let* $\hat\beta_1$ *denote the* l_1 *estimator. Then there is a constant c such that*

$$\|\hat\beta - \hat\beta_1\| \le \frac{2nK}{c} \ . \tag{4.2}$$

Proof. Without loss of generality assume that $\hat\beta_1$ is unique; for otherwise the vector y could be perturbed by a very small amount so that $\hat\beta_1$ is unique and the residuals (and hence the objective function for ρ also) changes by a bounded amount.

Now, consider the objective function for the l_1 estimator and let $h = \{\, i : x_i \hat\beta_1 = y_i \,\}$. From Koenker-Bassett (1978), the directional derivative of the objective function at $\hat\beta_1$ in direction b (with $\|b\| = 1$) is

$$-\sum_{i \notin h} \mathrm{sgn}(y_i - x_i \hat\beta_1) x_i b + \sum_{i \in h} |x_i b| > 0 \qquad \text{for all } b \ .$$

For each b of norm 1, define the finite set $E(b)$ and the function $g(b)$ as follows:

$$E(b) = \{\, \epsilon_i = \pm 1,\ i \notin h : -\sum_{i \notin h} \epsilon_i x_i b + \sum_{i \in h} |x_i b| > 0 \,\}$$

$$g(b) = \min_{\epsilon \in E(b)} \left\{ -\sum_{i \notin h} \epsilon_i x_i b + \sum_{i \in h} |x_i b| \right\} \ .$$

It follows that for each b, there is a neighborhood about b on which g is continuous and $g(b)>0$. Let $c \equiv \inf\{ g(b): \|b\| = 1 \}$. Then c is a positive constant (depending only on the design matrix). Therefore, using (4.1), for $\| \beta - \hat{\beta}_1 \| \geq 2nK/c$,

$$\sum_{i=1}^{n} \rho(y_i - x_i\beta) \geq \sum_{i=1}^{n} |y_i - x_i\beta| - nK$$

$$> \sum_{i=1}^{n} |y_i - x_i\hat{\beta}_1| + \frac{2nK}{c} c - nK$$

$$\geq \sum_{i=1}^{n} \rho(y_i - x_i\hat{\beta}_1) \quad .$$

Hence, (since $\hat{\beta}$ minimizes its objective function, $\| \hat{\beta} - \hat{\beta}_1 \| \leq 2nK/c$. ■

THEOREM 4.2. *Following Theorem 2.1, let $\gamma \in \Gamma$ be a norm so that in addition to positive, linear homogeneity,*

$$\gamma(a + b) \leq (\gamma(a) + \gamma(b)) \text{ for all } a, b \in \mathrm{R}^p \quad . \tag{4.3}$$

Then $\underline{B}_\gamma(\hat{\beta}) = \underline{B}_\gamma(\hat{\beta}_1)$ and $\bar{B}_\gamma(\hat{\beta}) = \bar{B}_\gamma(\hat{\beta}_1)$, for any M-estimator whose ρ-function satisfies (4.1) for both type I and type II distributions. In particular, $\underline{B}(\hat{\beta}) \geq m_* + 1$ for F of Type II.

Proof. From (4.3)

$$-\gamma(\hat{\beta} - \hat{\beta}_1) \leq \gamma(\hat{\beta} - \beta) - \gamma(\hat{\beta}_1 - \beta) \leq \gamma(\hat{\beta} - \hat{\beta}_1) \quad .$$

It follows from the proof of Theorem 2.1 that

$$|\gamma(\hat{\beta} - \beta) - \gamma(\hat{\beta}_1 - \beta)| \leq c\|\hat{\beta} - \hat{\beta}_1\| \leq c^* .$$

Therefore, the tail behavior $B(\hat{\beta}, a)$ differs from $B(\hat{\beta}_1, a)$ by a factor $-\log P\{e \geq a - c^*\}/-\log P\{e \geq a\}$, which tends to 1 (as $a \to \infty$) for both type I and type II distributions. ■

Remark: Note that $\gamma(b) \equiv \max_i |x_i b|$ satisfies the hypotheses of theorem 4.2.

The following result establishes that this lower bound on the tail performance of l_1-type estimators is essentially the same as their (Donoho-Huber) finite-sample breakdown point.

THEOREM 4.3 *For l_1 type estimators, $m_*+1 \le m^* \le m_*+2$.*

Proof. In fact, it has already been shown above that if there are only m_* outliers, l_1 type estimators will not break down. Thus, $m^* \ge m_*+1$.

By the definition of m_*, there exists a subset M of size m_*+1 and a vector $\|b_0\| = 1$ such that $\sum_{N\backslash M} |x_i b_0| \le \sum_M |x_i b_0|$. Thus, for $m = m_*+2$ there exists a subset M such that $\sum_{N\backslash M} |x_i b_0| < \sum_M |x_i b_0|$ (with strict inequality). Let $\eta(b) = \sum_M |x_i b| - \sum_{N\backslash M} |x_i b|$. Clearly, $\eta(b_0) > 0$, and $\eta(b_c) = \eta(cb) = c\eta(b)$. Suppose $y_i = 0$ for $i \in N\backslash M$ and $y_i = cx_i b_0$ for $i \in M$. Then,

$$\sum |y_i - x_i b_c| = \sum_{N\backslash M} |x_i b_c| = \sum_M |x_i b_c| - c\eta(b_0) = \sum |y_i| - c\eta(b_0).$$

On the other hand, for any bounded β,

$$\sum |y_i - x_i\beta| \ge \sum |y_i| - \sum_N |x_i\beta| \ge \sum |y_i| - n(max_i |x_i|)\|\beta\|.$$

With sufficiently large c, we have $\sum |y_i - x_i b_c| < \sum |y_i - x_i\beta|$ for all bounded β. This implies there is a breakdown with m_*+2 outliers. The extension to l_1-type estimators follows from Lemma 4.3. ■

Remark. This result strengthens the close link between tail performance of estimators under Type II errors and their finite-sample breakdown point. This connection is further developed below in the discussion of high-breakdown regression estimators.

There seems to be a common misapprehension about the l_1 estimator that for fixed x_{ij} the l_1 method is very robust while for random x_{ij} it is more fragile. Indeed, Donoho and Huber(1983) remark that the breakdown point of the l_1 estimator is ½ when there is "corruption only in y". However, as the preceeding result illustrates, the breakdown point, and therefore the tail-performance, of the l_1 estimator can be quite poor even for "fixed" designs. Even in the relatively favorable case of uniformly spaced x's the breakdown point is roughly .3 in the through-the-origin model.

While l_1-type estimators can have rather poor tail performance, and consequently low breakdown points, we shall show that the high breakdown estimators for regression recently introduced by Rousseeuw(1984) and Rousseeuw and Yohai(1984) possess good tail performance as measured by (2.3) when the error distribution is of Type II. We will focus on the least median of squares (LMS) estimator proposed by Rousseeuw(1984) which solves

$$\underset{b \in \mathbb{R}^p}{Min}\ \underset{i}{median}\{(y_1 - x_i b)^2, \cdots, (y_n - x_n b)^2\}.$$

The breakdown point analysis of the LMS estimator is carefully carried out in Rousseeuw and Leroy (1987). Following their terminology we shall say that the observations $\{(y_i, x_i)\ i = 1, n\}$ are in general position if any p of them give a unique determination of $b(h) = X(h)^{-1} y(h)$. Some slight modifications of their argument for contamination of both y and X leads to the following result for *fixed* design points.

LEMMA 4.3 *For the LMS estimator with observations in general position* $m^* = [n/2] - p + 2$.

The lower bound on the tail performance of the LMS estimator without restriction to observations in general position is closely related.

LEMMA 4.4 *For F of Type II, the LMS estimator has* $\underline{B} \geq [n/2] - k + 1$, *where k is the smallest integer such that* $\inf_{\|b\|=1} |x_i b|_{(k)} > 0$.

Proof. Since the LMS estimator is scale equivariant, it suffices to show that for some constant C,

$$\max |x_i T_n| \leq C\, |y|_{(n-[n/2]+k)}$$

The order statistic $|x_i b|_{(k)}$ is continuous in b so $|x_i b|_{(k)} > c > 0$ for all $\|b\| = 1$. Or, $|x_i b|_{(k)} > c\ \|b\|$ for all b.

Now, if there are only $[n/2]-k$ possible outliers and all the other $n-[n/2]+k$ observations (y's) are bounded by 1, we must show that the estimator T_n is uniformly bounded by, say K. This is true since for at least $n-[n/2]+k$ residuals, $|y_i-x_iT_n| \geq c\|T_n\|-1$, which implies that median $|y_i-x_iT_n| \geq 1 \geq$ median $|y_i|$ if $\|T_n\| \geq 2/c$. Therefore the bound on $\|T_n\|$ can be $K=2/c$. ■

If the observations are in general position, then $k = p$, and combining Lemmas 3.3 and 3.4 we have $\underline{B} \geq m^*-1$. In fact, a more careful analysis enables one to strengthen Lemma 3.4 and thus to prove

THEOREM 4.4. *For the LMS estimator with observations in general position* $\underline{B} \geq m^* = [n/2]-p+2$.

Proof. (Sketch) Suppose initially, $|y_i| \leq 1$ for all $i=1,2,\ldots n$, then for some constant C, $\|T_n\| \leq C$. Set $M = \sup_{|y_i| \leq 1} \max_i |y_i-x_iT_n|$. This is also finite. Let V denote a $p-1$-dimensional subspace in $\mathbf{R}^p$ and let $S(t)$ be the set of all x_i whose distance to V is no larger than t. The quantity $\delta = \frac{1}{2}\inf\{t>0 \mid \#S(t) \geq p\}$. Clearly, δ depends only on the design points. If the observations are in general position, it can be shown that $\delta > 0$.

Finally, essentially the same geometric argument given in Rousseeuw and Leroy(1987, pp. 118-119), implies that $\|T_n\| \leq 3C+2M/\delta$ for T_n obtained by using $[n/2]+p-1$ "good" observations, i.e. $|y_i| \leq 1$. ■

Remark. When the observations are not in general position, it can be shown that $m^* \leq [n/2]-k+2$. In that case we still have that $\underline{B} \geq m^*-1$.

The lower bound for the tail performance of the LMS estimator may be extended to the general S-estimators of Rousseeuw and Yohai (1984), which minimize an estimate of scale $s(r_1, \cdots, r_n)$ derived from

$$n^{-1}\sum\rho(r_i/s)=\rho_0$$

where $r_i = y_i-x_ib$. Using the fact that

$$\alpha_0 \text{median}|r_i| \le s(r_1, \cdots, r_n) \le \alpha_1 \text{median}|r_i|$$

for some positive constants α_0 and α_1, and from the proof of Theorem 4.4 it can be shown that $\underline{B} \ge m^* - 1$ for S-estimators when the observations are in general position (see He (1988) for details).

An upper bound on the tail performance of the LMS estimator for error distributions of Type II is given in the following result.

THEOREM 4.6 *For F of Type II the LMS estimator has* $\bar{B} \le [n/2]+1$.

Proof. Suppose n is odd, and set $m = (n+1)/2$. We begin by showing that for sufficiently large A, $0 < A < a$,

$$P(|y|_{(m)} > a, |y|_{(m-1)} > A) \le P(|y|_{(m)} > a, |y|_{(m-1)} \le A). \tag{4.4}$$

In fact, for some combinatoric constant c_n,

$$P(|y|_{(m)} > a, |y|_{(m-1)} > A) \le c_n (2-2F(a))^m (2-2F(A))$$

and

$$P(|y|_{(m)} > a, |y|_{(m-1)} \le A) = c_n (2-2F(a))^m (2F(A)-1)^{m-1}.$$

So for fixed n, and A sufficiently large (4.4) holds. Next, we will suppose $\|x_i\| \le 1$ for all $i = 1,2,\ldots,n$ and show that if $|y|_{(m)} > a$ and $|y|_{(m-1)} \le A$ then $\|T_n\| \ge \frac{1}{2}|y|_{(m)} - A$. Note that for bounded $\|b\|$, $\text{median}_i |y_i - x_i b| \ge |y|_{(m)} - \|b\|$. Now, choose b_0 such that for i corresponding to $|y|_{(m)}$, $r_i = \frac{1}{2}|y|_{(m)}$ and $\|b_0\| = \frac{1}{2}|y|_{(m)}$. When $x_{i1} \equiv 1$, this can be accomplished by setting $b_0 = (\frac{1}{2}|y|_{(m)}, 0, \cdots, 0)$. For this b_0, and i such that $|y_i| \le |y|_{(m-1)}$,

$$|y_i - x_i b| \le |y|_{(m-1)} - \|x_i\|\,\|b_0\|.$$

$$\le A + \tfrac{1}{2}|y|_{(m)}.$$

Hence, including $|y|_{(m)}$, $|y_i - x_i b_0| \le \frac{1}{2}|y|_{(m)} + A$ for m subscripts and

$$\text{median}|y_i - x_i b_0| \le \tfrac{1}{2}|y|_{(m)} + A.$$

Thus, if $\frac{1}{2}|y|_{(m)}+A < |y|_{(m)}-\|b\|$ or, $\|b\| < \frac{1}{2}|y|_{(m)}-A$, the median residual cannot be minimized at T_n; and consequently, $\|T_n\| \geq \frac{1}{2}|y|_{(m)}-A$. Finally, from (4.4),

$$P(|y|_{(m)}>a) = P(|y|_{(m)}>a, |y|_{(m-1)}>A) + P(|y|_{(m)}>a, |y|_{(m-1)}\leq A)$$

$$\leq 2P(|y|_{(m)}>a, |y|_{(m-1)}\leq A) \leq 2P(\|T_n\| \geq a/2-A).$$

and it follows that $\bar{B} \leq (n+1)/2$ from the Type II behavior of F.

For n even, we take the median of the n ordered residuals $\{r_1^2, \cdots r_n^2\}$ to be $\frac{1}{2}(r_{n/2}^2 + r_{1+n/2}^2)$. Let $z_1=|y|_{(n/2)}$, $z_2=|y|_{1+(n/2)}$ and $z_3 = \max_{i\leq n/2}|y_{(i)}|$. And argument similar to that for (4.4) shows that

$$P(z_1 > a) \leq 2P(z_1>a, z_3 \leq A)$$

for sufficiently large A. It remains to show that when $z_1>a$ and $z_3\leq A$ (as a becomes large enough) the LMS estimator satisfies $\|T_n\| > a/4$. For $\|b\| > a/4$, we have $\text{median}\{r_i^2\} > \frac{1}{2}(z_1^2 + z_2^2) + \|b\|^2-\|b\|(z_1+z_2)$. On the other hand, if we choose $b_0 = (\frac{1}{2}z_1, 0, \cdots, 0)$, then $\text{median}\{r_i^2\} \leq \frac{1}{2}(\frac{1}{4}z_1^2 + (z_2^2-\frac{1}{2}z_1)^2)$ or $\frac{1}{2}(z_1^2+z_2^2)-(\frac{1}{4}z_1^2+\frac{1}{2}z_1z_2)$ One can then show that if $\|b\| > a/4$, then $(\frac{1}{4}z_1^2+\frac{1}{2}z_1z_2) \geq \|b\|(z_1z_2)$ which implies that $\text{median}\{r_i^2\}$ cannot be minimized at T_n. Since $P(|y|_{(n/2)}>a) \leq 2P(\|T_n\|>a/4)$ implies $\bar{B} \leq 1+n/2$, the proof is complete. ∎

Remark. The theorem also holds if T_n minimizes the median absolute residual instead of the median squared residual. A further extension can also be made to T_n which minimize the k-th largest absolute residual $|r|_{(n-k)}$. In that case, $\bar{B} \leq k+1$. A special case is considered above in the Corollary to Lemma 2.1 where $k = [(n-p+1)/2]$. Combining the lower bound there with the upper bound here, we have in this case $[(n-p+1)/2] \leq \underline{B} \leq \bar{B} \leq [(n-p+1)/2]+1$

5. Conclusion

The results of the preceeding section emphasize the close link, suggested in the introduction, between the tail performance of estimators and their finite-sample breakdown points.

Theorem 1.1 clarifies this relationship for a broad class of location estimators. In the regression context, the breakdown point of l_1-type estimators is seen to be essentially the same as the lower bound on their tail performance under heavy tailed error conditions. For the least median of squares estimator, and indeed for the broad class of S-estimators as well as for the least squares estimator, we find that tail performance is bounded below by the breakdown point under Type II error conditions. Others, Donoho and Huber (1983), Hampel, *et al.* (1986), and Rousseeuw and Leroy (1987), have made a persuasive case that the breakdown point of estimators is an important "figure of merit" in the assessment of quantitative robustness. The conjunction of breakdown and tail performance affords, in our view, a rewarding new window on the robustness scene.

REFERENCES

Bahadur, R.R. (1967). Rates of convergence of estimators and test statistics. *Annals of Mathematical Statistics*, 38, 303-324.

Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point. *A Festschrift for Eric Lehmann*, P.J. Bickel, K.A. Doksum, J.L. Hodges, (eds.) Wadsworth: Belmont, CA.

Hampel, F.R., (1968) *Contributions to the Theory of Robust Estimation*, Ph.d. Thesis, Department of Statistics, University of California, Berkeley.

He, X., (1988) *Contributions to the Theory of Statistical Breakdown*, Ph.d. Thesis, Department of Statistics, University of Illinois,

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986) *Robust Statistics: The Approach Based on Influence Functions*, Wiley: New York

Jurečková, J. (1981) Tail behavior of location estimators, *Annals of Statistics*, 9, 578-585.

Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica*, 46, 33-50.

Koenker, R. and Portnoy, S. (1987). L-estimation for linear models. *Journal of the American Statistical Association*, 82, 851-857.

Rouseeuw, P.J. (1984) Least median of squares regression *Journal of the American Statistical Association*, 79, 871-880.

Rouseeuw, P.J. and Leroy , A. (1987) *Robust Regression and Outlier Detection*, Wiley: New

York.

Rouseeuw, P.J. and Yohai, V. (1984) Robust regression by means of S-estimators, in *Robust and Nonlinear Time-Series analysis,* J. Franke, W. Hardle, and R.D. Martin (eds.) Springer: New York.

Sievers, G.L. (1978). Estimation of location: a large deviation comparison. *Annals of Statistics, 6, 610-618.*