응용통계연구 제 6권 1호
1993년 3월 pp 79 - 93

Adaptive L-Estimation for Regression Slope under Asymmetric Error Distributions[1)]

Sang Moon Han[2)]

Abstract

We consider adaptive L-estimation of estimating slope parameter in regression model. The proposed estimator is simple extension of trimmed least squares estimator proposed by Ruppert and Carroll. The efficiency of the proposed estimator is especially well compared with usual least squares estimator, least absolute value estimator, and M-estimators designed for asymmetric distributions under asymmetric error distributions.

1. INTRODUCTION

In the last thirty years a large amount of work has been done in the area of robustness. A large number of procedures have been proposed as alternatives to the classical least squares procedure. Of these perhaps the main classes are the so-called M-, L- and R-classes(see e.g. Huber(1981), Hogg(1979), Hogg, et.al(1988)). M-estimators generalized in a very natural way from the location model to the regression model(e.g. Huber(1973)). R-estimators were generalized by Jureckova(1971), Jaeckel(1972), Adichie(1974), Hettmansperger and McKean(1977).

L-estimators were first proposed for the regression case by Bickel(1973). Although his estimators had the right asymptotic behaviour they have complex forms and also difficult to compute. Moreover they are not invariant with respect to a reparameterization of the design.

Koenker and Bassett(1978) suggested an alternative approach to L-estimators that is not based on an ordering of the residuals from a preliminary fit. They define so called regression quantiles as M-estimators with a particular defining check function and then use these to define regression trimmed means in a very natural way. Although they concentrated on trimmed means, more general

1) This research was supported by the 1990 Seoul City Research Fund.
2) Department of Computer Science and Statistics, Seoul City University, Jeonnong Dong 8-3 Dongdaemoon Ku, Seoul.

L-estimators can be defined in a straightforward way. Their estimators also have the right asymptotic behaviour as well as being invariant wrt reparameterization of the design. In addition they can be computed easily using modifications of L_1-programs. These estimators are the first practically useful examples of regression L-estimators.

Now we mention briefly about regression quantiles and trimmed least squares directly related to our work. Consider the regression model

$$\underset{\sim}{y}=X\underset{\sim}{\beta}+\underset{\sim}{z} \tag{1.1}$$

where $\underset{\sim}{y} = (y_1,\cdots,y_n)'$, X is $n\times p$ matrix of known constants whose i-th row is $\underset{\sim}{x}_i'$, $\underset{\sim}{\beta} = (\beta_0,\beta_1,\cdots,\beta_{p-1})'$ is a vector of unknown parameters, and $\underset{\sim}{z} = (z_1,\cdots,z_n)'$ is a vector of independent, identically distributed random variables with unknown distribution function F.

The basis of the definition of regression quantiles is the fact that the ordinary sample quantiles(order statistics) for the location model may be found as M-estimators with check function

$$P_\theta(x)=\begin{cases}\theta x & x\geq 0\\ (\theta-1)x & x<0\end{cases} \tag{1-2}$$

where $\theta\ \varepsilon\ (0,1)$. This can easily be generalized to the model (1.1). Let $K(\theta)$ denote the θ-th regression quantile, then $K(\theta)$ solves the minimization problem

$$\min_{\underset{\sim}{b}\in R^p}\sum_{i=1}^{n}P_\theta(y_i-\underset{\sim}{x}_i'\underset{\sim}{b}) \tag{1.3}$$

Using this, Koenker and Bassett(1978) suggested the following trimmed mean. For $0 < p_1 < p_2 < 1$ let $K(p_1)$, $K(p_2)$ denote the p_1-th and p_2-th regressing quantiles. Define

$$a_i=\begin{cases}1 & \text{if } \underset{\sim}{x}_i'K(p_1)\leq y_i\leq \underset{\sim}{x}_i'K(p_2)\\ 0 & \text{otherwise.}\end{cases} \tag{1.4}$$

The regression trimmed mean $L(\underset{\sim}{p})$ is then the least squares estimator based on those observations with $a_i=1$ i.e.

$$L(\underset{\sim}{p})=(X'AX)^{-}X'A\underset{\sim}{y} \tag{1.5}$$

with A = diag(a_i) and X the design matrix with rows $\underset{\sim}{x}_i{}'$,$i=1,\cdots,n$ and $\underset{\sim}{p}=(p_1, p_2)'$.

Suppose X contains an intercept and $n^{-1}X'X \rightarrow$ Q as $n \rightarrow \infty$, with Q positive definite and ξ_1 and ξ_2 be respectively p_1-th and p_2-th quantile of underlying distribution F. Under certain technical assumptions, Ruppert and Carroll(1980) showed that

$$\sqrt{n}(L(\underset{\sim}{p})-\underset{\sim}{\beta}-\underset{\sim}{\delta}(\underset{\sim}{p})) \quad \overset{D}{\rightarrow} \quad N(0,\ \sigma^2(\underset{\sim}{p},F)Q^{-1})$$

with $\overset{D}{\rightarrow}$ denoting convergence in distribution ,

$\underset{\sim}{\delta}(\underset{\sim}{p})=((p_2-p_1)^{-1}\int_{\xi_1}^{\xi_2} xdF(x),0,\cdots,0))'$ and $\sigma^2(\underset{\sim}{p},F)$ the asymptotic variance of the p_1-th and $(1-p_2)$-th trimmed mean in the location case. These results will be used for asymptotics for our estimator.

2. Motivation and large sample properties of the proposed estimator

Before starting this section, we introduce some notation and assumptions which are imposed for all lemmas, theorems and corollary in this section. Although $\underset{\sim}{y}$, X, and $\underset{\sim}{z}$ in (1.1) depend on n, this is not made explicit in the notation. Let $\underset{\sim}{e}=(1,0,\cdots,0)'$ be $(p\times 1)$, and let I_p be the $(p\times p)$ identity matrix. Whenever r is scalar, $\underset{\sim}{r}=r\underset{\sim}{e}$. For $0 < p_i < 1$, define $\xi_i = F^{-1}(p_i)$. Let $N_p(\underset{\sim}{\mu},\Sigma)$ denote the p-variate normal distribution with mean vector $\underset{\sim}{\mu}$ and variance-covariance matrix Σ. We also make the following assumptions about the family Λ of distributions in what follows.

A1. F has a continuous density f and $f(x)>0$ for all $F\in\Lambda$.

A2. Let $\underset{\sim}{x}_i=(x_{i1},x_{i2},\cdots,x_{ip})'$ be the i-th row of X and $x_{i1}=1$, $i=1,2,\cdots,n$ and $\sum_{i=1}^{n} x_{ij}=0$, $j=2,3,\cdots,p$.

A3. $\lim_{n\to\infty}\left(\max_{j\le p, i\le n} n^{-1/2}|x_{ij}|\right)=0$.

A4. There exists a positive definite matrix Q such that $\lim_{n\to\infty} n^{-1}(X'X)=Q$.

A5. Assuming $\hat{\beta}_0$ be preliminary fit, then $\sqrt{n}(\hat{\beta}_0-\beta-\theta\mathbf{e})=0_p(1)$ for some constant θ .

Motivation of our estimator is as follows ; Assume $0<p_0<p_1<\cdots<p_k=q_0<1$. Moreover, let $K(p_i)$ be the corresponding p_i-th regression quantiles. Then for $i=1,2,\cdots,n$, define

$$a_1=\begin{cases}1 & \text{if } \mathbf{x}_i'K(p_0)<y_i\le \mathbf{x}_i'K(p_1)\\ 0 & \text{if otherwise}\end{cases}$$
$$a_2=\begin{cases}1 & \text{if } \mathbf{x}_i'K(p_1)<y_i\le \mathbf{x}_i'K(p_2)\\ 0 & \text{if otherwise}\end{cases} \tag{2.1}$$
$$\vdots$$
$$a_k=\begin{cases}1 & \text{if } \mathbf{x}_i'K(p_{k-1})<y_i\le \mathbf{x}_i'K(p_k)\\ 0 & \text{if otherwise .}\end{cases}$$

Let $L(\mathbf{p}_1)$, $L(\mathbf{p}_2)$, $\cdots$, $L(\mathbf{p}_k)$ be the least squares estimators based on those observations with $a_1=1, a_2=1,\cdots,a_k=1$ respectively. That is,

$$\begin{aligned} L(\mathbf{p}_1) &= (X'A_1X)^- X'A_1\mathbf{y} ,\\ L(\mathbf{p}_2) &= (X'A_2X)^- X'A_2\mathbf{y} ,\\ &\vdots \\ L(\mathbf{p}_k) &= (X'A_kX)^- X'A_k\mathbf{y} ,\end{aligned} \tag{2.2}$$

where $A_i=\text{diag}(a_i)$ for $i=1,2,\cdots,k$ and X design matrix with rows $\mathbf{x}_i'$ for $i=1,2,\cdots,n$.

Then our estimator has the following form :

$$B_k=w_1L(\mathbf{p}_1)+w_2L(\mathbf{p}_2)+\cdots+w_kL(\mathbf{p}_k), \text{ with } \sum_{i=1}^{k}w_i=1 .$$

Therefore, our slope estimator can be obtained by just deleting the intercept part from $L(\mathbf{p}_1),\cdots,L(\mathbf{p}_k)$. Let us denote these $(p-1)$ dimension estimators as $L_0(\mathbf{p}_1),\cdots,L_0(\mathbf{p}_k)$. Then our slope estimator has the following form :

$$C_k = w_1 L_0(\underline{p}_1) + w_2 L_0(\underline{p}_2) + \cdots + w_k L_0(\underline{p}_k), \text{ with } \sum_{i=1}^{k} w_i = 1 . \tag{2.3}$$

For practical application, with k=2, it is quite reasonable to give nearly equal weights to w_1 and w_2 if the true underlying distribution is similar to normal, but we want to give more weight to w_1 or w_2 if the true underlying distribution is skewed to the left or right respectively . We will explain the method of assigning the weights $w_1, w_2, \cdots, w_k$ later. However, before that we state the necessary theorems using the result of Ruppert and Carroll(1980).

Theorem 2.1 Fix k and $p_0, p_1, \cdots, p_k$ such that $p_1 - p_0 = p_2 - p_1 = \cdots = p_k - p_{k-1} = q_k$ then

$$\sqrt{n}[B_k - \underline{\beta} - \sum_i w_i \underline{\delta}(\underline{p}_i)] \xrightarrow{D} N_p(0, (\underline{w}' \Sigma \underline{w}) Q^{-1})$$

where $\underline{w} = (w_1, w_2, \cdots, w_k)'$, and for $i = 1, 2, \cdots, k$ and $i \le j \le k$

$$\underline{\delta}(\underline{p}_i) = \left(q_k^{-1} \int_{\xi_{i-1}}^{\xi_i} x dF(x), 0, \cdots, 0 \right)'$$

$\Sigma = (\sigma_{ij})_{k \times k}$, with

$$\sigma_{ii} = q_k^{-2} \left\{ -2 \int_{\xi_{i-1}}^{\xi_i} x F(x) dx + 2\xi_i \int_{\xi_{i-1}}^{\xi_i} F(x) dx - [\int_{\xi_{i-1}}^{\xi_i} F(x) dx]^2 \right\},$$

$$\sigma_{ij} = \sigma_{ji} = q_k^{-2} \left\{ \xi_j \int_{\xi_{i-1}}^{\xi_i} F(x) dx - \xi_{j-1} \int_{\xi_{i-1}}^{\xi_i} F(x) dx - \int_{\xi_{i-1}}^{\xi_i} F(x) dx \int_{\xi_{j-1}}^{\xi_j} F(x) dx \right\} .$$

Proof. From theorm 3 of Ruppert and Carroll(1980), we have the asymptotic expansion

$$\sqrt{n}[L(\underline{p}_j) - \underline{\beta} - \underline{\delta}(\underline{p}_j)] = Q^{-1} \sqrt{n} \left\{ \sum_{i=1}^{n} x_i [\Phi_j(z_i) - E\Phi_j(z_i)] \right\} + o_p(1)$$

where for $j = 1, 2, \cdots, k$,

$$\begin{aligned} \Phi_j(z) &= \xi_{j-1}/q_k && \text{if } z < \xi_{j-1} \\ &= z/q_k && \text{if } \xi_{j-1} \le z \le \xi_j \\ &= \xi_j/q_k && \text{if } z > \xi_j . \end{aligned}$$

Therefore

$$\sqrt{n}(B_k - \underset{\sim}{\beta} - \sum_{i=1}^{k} w_i \underset{\sim}{\delta}(\underset{\sim}{p_i})) = Q^{-1}\sqrt{n}\left\{\sum_{i=1}^{n} x_i \sum_{r=1}^{k} w_r[\Phi_r(z_i) - E\Phi_r(z_i)]\right\} + o_p(1) \quad . \tag{2.4}$$

If we let $\Phi_k^*(z_i) = \sum_{r=1}^{k} w_r[\Phi_r(z_i) - E\Phi_r(z_i)]$ then in order to guarantee asymptotic p-variate normality of the right hand side of above equation, it suffice to show that for $\underset{\sim}{c} \in R^p$, $W_n = \underset{\sim}{c}' Q^{-1} n^{1/2} \sum_{i=1}^{k} x_i \Phi_k^*(z_i)$ has asympototically univariate nomal.

Let $W_{in} = \underset{\sim}{c}' Q^{-1} n^{-1/2} x_i \Phi_k^*(z_i)$. Then

$$E(W_{in}) = 0$$
$$Var(W_{in}) = \underset{\sim}{c}' Q^{-1} n^{-1} x_i Var\{\Phi_k^*(z_i)\} x_i' Q^{-1} \underset{\sim}{c}$$

Hence $E(W_n)=0$, $Var(W_n) \to Var\{\Phi_k^*(z_i)\} \underset{\sim}{c}' Q^{-1} \underset{\sim}{c} = (\underset{\sim}{w}' \Sigma \underset{\sim}{w}) \underset{\sim}{c}' Q^{-1} \underset{\sim}{c}$ as $n \to \infty$ where Σ is variance-covariance of $(\Phi_1(z), \Phi_2(z), \cdots, \Phi_k(z))$. By Lindeberg C.L.T., it suffice to show that for all $\varepsilon > 0$,

$$\frac{1}{VarW_n} \sum_{i=1}^{k} \int_{(|W_{in}| > \varepsilon\sqrt{VarW_n})} W_{in}^2 dF_{in} \to 0 \text{ as } n \to \infty \tag{2.5}$$

Where F_{in} be distribution function of W_{in}. Note by our construction, $|\Phi_k^*(z)| \leq k$ for some positive number k. Then the left-hand side of (2.5) is bounded by

$$\frac{k^2}{VarW_n} \sum_{i=1}^{n} \int_{\{\max|W_{in}| > \varepsilon\sqrt{VarW_n}\}} n^{-1} \underset{\sim}{c}' Q^{-1} x_i \, x_i' Q^{-1} \underset{\sim}{c} \; dF_{in}$$
$$\leq \frac{k^2}{VarW_n} \sum_{i=1}^{n} \underset{\sim}{c}' Q^{-1} (n^{-1} x_i \, x_i') Q^{-1} \underset{\sim}{c} P(\max|W_{in}| > \varepsilon\sqrt{VarW_n}) \; .$$

Note by condition A3,

$$P(\max|W_{in}| > \varepsilon\sqrt{VarW_n}) = P(\max\sqrt{n}|c' Q^{-1} x_i \Phi^*(z_i)| > \varepsilon\sqrt{VarW_n}) \to 0 \text{ as } n \to \infty \; . \tag{2.6}$$

Conbining the result (2.6), the right hand of (2.5) goes to 0 as $n \to \infty$. From the above result, the right hand side of (2.4) has asymptotically p-variate normal. By straight forward calculation, it has mean vector $\underline{0}$ and variance-covariance matrix $(\underline{w}'\Sigma\underline{w})Q^{-1}$ and Σ is characterized by variance-covariance of $\Phi_i(z)$, $i=1,2,\cdots,k$. Also by straightforward calculation, σ_{ii} which is the variance of $\Phi_i(z)$ and the covariance σ_{ij} of $\Phi_i(z)$ and $\Phi_j(z)$ can be obtained as this the theorem states.

From theorem 2.1, we deduce the fact that $\sqrt{n}(C_k-\underline{\beta}_0) \xrightarrow{D} N_{p-1}(0,(\underline{w}'\Sigma\underline{w})Q_0^{-1})$, where $\underline{\beta}_0$ is obtained from $\underline{\beta}$ by deleting first component and Q_0^{-1} is obtained from Q^{-1} by deleting first row and column.

Next we want to find minimizing weight $\underline{w}$ of the asymptotic variance of C_k. But usually it is very difficult to estimate the asymptotic variance of C_k. So we adopt the similar idea of Johns(1974) to approximate variance-covariance structure Σ and then find minimizing weight $\underline{w}$. We approximate, for $i=1,2,\cdots,k$, as follows;

$$\sigma_{ii} \sim q_k^{-2} b_i(1-b_i)d_i^2 \tag{2.7}$$

$$\sigma_{ij} \sim q_k^{-2} b_i(1-b_j)d_i d_j \tag{2.8}$$

where $d_i=\xi_i-\xi_{i-1}$, $b_i=p_{i-1}+\frac{1}{2}q_k$. Then the approximated variance-covariance structure of Σ is given as follows ;

$$B=q_k^{-2}D\begin{bmatrix} b_1(1-b_1) & b_1(1-b_2) & \cdots & b_1(1-b_k) \\ b_2(1-b_1) & b_2(1-b_2) & \cdots & b_2(1-b_k) \\ \vdots & \vdots & & \vdots \\ b_1(1-b_k) & b_2(1-b_k) & \cdots & b_k(1-b_k) \end{bmatrix} D$$

Next we want to minimize the following quantity :

$$\min_{\underline{w}'\underline{1}=1}(\underline{w}'B\underline{w})Q_0^{-1}$$

where $\underline{1}$ = (1, 1, ..., 1)'. Because Q_0^{-1} is fixed, we want to minimize $\underline{w}'B\underline{w}$ under the constraint $\underline{w}'\underline{1}=1$. By a straightforward Lagrange Multiplier argument, we establish that

$$\frac{\partial}{\partial \underline{w}}[\underline{w}'B\underline{w}-\lambda(\underline{w}'\underline{1}-1)]=2B\underline{w}-\lambda\underline{1}=0$$
$$\frac{\partial}{\partial \lambda}[\underline{w}'B\underline{w}-\lambda(\underline{w}'\underline{1}-1)]=\underline{w}'\underline{1}-1=0 \ . \qquad (2.9)$$

If we solve (2.9), we have $\lambda=2(\underline{1}'B^{-1}\underline{1})^{-1}$ and this gives $\underline{w}=(\underline{1}'B^{-1}\underline{1})^{-1}B^{-1}\underline{1}$ Using this minimizing weight, we get

$$\min_{\underline{w}'\underline{1}=1}\underline{w}'B\underline{w}=(\underline{1}'B^{-1}\underline{1})^{-1} \ .$$

With the minimizing weight given above, we define our slope estimator as given by (2.3). If we define

$$e_1=1/d_1(b_2/b_1d_1-1/d_2)q_k,$$
$$e_i=1/d_i(2/d_i-1/d_{i-1}-1/d_{i+1}), \quad i=2,3,\cdots,k-1 \ ,$$
$$e_k=1/d_k(-1/d_k-(1-b_{k-1})/(1-b_k)d_k)q_k,$$

then by straightforward calculation, we have the minimiging weight as

$$w_i=e_i/\left\{\sum_{i=1}^{k}e_i\right\} \ , \ i=1,2,\cdots,k \ .$$

Let e_i^* be a consistent estimator of e_i by simply substituting d_i for corresponding consistent estimator $\hat{d}_i$. Then if we define $w_i^*=e_i^*/\left\{\sum_{i=1}^{k}e_i^*\right\}$, $(i=1,2,\cdots,k)$, then our estimator is of the form

$$\hat{C}_k=w_1^*L_0(\underline{p}_1)+w_2^*L_0(\underline{p}_2)+\cdots+w_k^*L_0(\underline{p}_k) \ . \qquad (2.10)$$

Finally we want to show that $\sqrt{n}(\hat{C}_k-\underline{\beta_0})$ and $\sqrt{n}(C_k-\underline{\beta_0})$ has same limiting distrisution. In order to do that we need consistent estimator of w_i. Next theorem is very useful in the construction of consistent estimators $w_1^*,\cdots,w_k^*$ from residuals of the preliminary fit.

Theorem 2.2 Let $0<p_0<p_1<\cdots,<p_k=q_0<1$ and ξ be the np_i-th ordered residuals from the preliminary fit $\hat{\underline{\beta}}_0$ which satisfies assumption A5. Then, for $i=1,2,\cdots,k$, $\hat{d}_i$ is consistent estimator of d_i.

Proof. Using Lemma 1 of Ruppert and Carroll(1980), we have for $i=1,2,\cdots,k$,

$$\sqrt{n}(\hat{\xi}_i-\xi_i)=[f(\xi_i)]^{-1}n^{-1/2}\sum_{i=1}^{n}\phi_{p_i}(z_i-\xi_i)-e'n^{1/2}(\hat{\underline{\beta}}_0-\underline{\beta})+0_p(1)$$

and (2.11)

$$\sqrt{n}(\hat{\xi}_{i-1}-\xi_{i-1})=[f(\xi_{i-1})]^{-1}n^{-1/2}\sum_{i=1}^{n}\phi_{p_i-1}(z_i-\xi_{i-1})-e'n^{1/2}(\hat{\underline{\beta}}_0-\underline{\beta})+0_p(1)$$

where $\phi_\theta(x)=\theta-I(x<0)$. Therefore by subtracting two equations in (2.11), we have

$$\sqrt{n}(\hat{d}_i-d_i)=[(f(\xi_i)]^{-1}n^{-1/2}\sum_{i=2}^{n}\phi_{p_i}(z_i-\xi_i)- \qquad (2.12)$$
$$[(f(\xi_{i-1})]^{-1}n^{-1/2}\sum_{i=1}^{n}\phi_{p_{i-1}}(z_i-\xi_{i-1})+0_p(1)\ .$$

By C.L.T., each of the first two terms of the right hand side of (2.12) has a limiting normal with finite variance. Therefore for $i=1,2,\cdots,k$, we have

$$\hat{d}_i-d_i=o_p(1)\ .$$

From theorem 2.2, we have a consistent estimator of the difference of the two population quantiles of the underlying distribution. This enables us to construct consistent estimator $\hat{C}_k$ given by (2.10).

Corollary 2.1 $\sqrt{n}(C_k-\underline{\beta}_0)$ and $\sqrt{n}(\hat{C}_k-\underline{\beta}_0)$ have the same limiting distribution.

Proof. By construction of C_k and $\hat{C}_k$, we get

$$\sqrt{n}[(C_k-\underline{\beta}_0)-(\hat{C}_k-\underline{\beta}_0)]=\sum_{i=1}^{k}(w_i-w_i^*)\sqrt{n}L_0(p_i)\ . \qquad (2.13)$$

If we note that $w_i^*\xrightarrow{p}w_i$ and $\sqrt{n}L_0(p_i)$ is bounded in distribution, $\sqrt{n}(C_k-\underline{\beta}_0)$ and $\sqrt{n}(\hat{C}_k-\underline{\beta}_0)$ have the same limiting distribution.

Remark In this section we show that theorem 2.1 leads to the basic conclusion about the constructed estimators

1. The intercept estimate is asymptotically unbiased if F is symmetric about zero, $p_0 = 1 - p_k$, and symmetric weights are given;
2. The slope estimates are asymptotically unbiased even if F is asymmetric ;
3. The asymptotic covariance matrix Σ depends on the choice of and regression quantiles $K(p_0), K(p_1), \cdots, K(p_k)$ and in general it will be very difficult to estimate and in our construction of slope estimator $\hat{C}_k$, we use approximating technique similar to Johns.

3. Monte Carlo study for slope estimators

In the pilot Monte Carlo study our slope estimator $\hat{C}_k$ for k=2 case called RHH is not well worked under symmetric F compared with existing robust M-estimators designed for symmetric. But RHH works pretty good compared with existing estimators under asymmetric F.

We also study for k=3 and k=4 case. But in all cases the performance of our estimator becomes poorer as k becomes large due to overadaptation. So our Monte Carlo study concentrates on k=2 and asymmetric F. Even RHH using K(0.5) as preliminary fit works pretty good under asymmetric F, we find prelimary fit is very important to improve efficiency within our range of study. So we use different preliminary fit from K(0.5) to get high efficiency within the range of our study. Therefore we need a different version of σ_{ii} and σ_{ij} to get a little bit different version of estimator RHH.

If we omit the condition $p_1 - p_0 = p_2 - p_1 = q_2$, we get the following variance -covariance structure of our two block estimator called SHH which is given as follows ;

$$\sigma_{ii} = (r_i)^{-2} \ \{2\xi_i \int_{\xi_{i-1}}^{\xi_i} F(x)dx - [\int_{\xi_{i-1}}^{\xi_i} F(x)dx]^2\} \ ,$$

$$\sigma_{ij} = \sigma_{ji} = \quad (r_i r_j)^{-1}[-\xi_{i-1} \int_{\xi_{i-1}}^{\xi_i} F(x)dx + \xi_i \int_{\xi_{i-1}}^{\xi_i} F(x)dx - \int_{\xi_{i-1}}^{\xi_i} F(x)dx \int_{\xi_{i-1}}^{\xi_i} F(x)dx] \ ,$$

for $1 \le i \le j \le 2$, where $r_1 = p_1 - p_0$ and $r_2 = p_2 - p_1$. Let $b_1 = p_0 + 0.5 r_1$, $b_2 = p_1 + 0.5 r_2$,

$d_1=\xi_1-\xi_0$, and $d_2=\xi_2-\xi_1$. Then, by similar method as Johns(1974), we get $\sigma_{11}\sim r_1^{-2}b_1(1-b_1)d_1^2$, $\sigma_{12}\sim(r_1r_2)^{-1}b_1(1-b_2)d_1d_2$ and $\sigma_{22}\sim r_2^{-2}b_2(1-b_2)d_2^2$. If we define $V_1=b_1(1-b_1)d_1^2$, $V_2=b_2(1-b_2)d_2^2$ and $C=b_1(1-b_2)d_1d_2$, we easily obtain the $\underline{w}=(w_1,w_2)'$ that minimizes

$$\underline{w}'\begin{pmatrix}(r_1)^{-2}V_1 & (r_1r_2)^{-1}C\\ (r_1r_2^{-1}C & (r_2)^{-1}V_2\end{pmatrix}\underline{w}Q_0^{-1}\ ,\ \underline{w}'\underline{1}=1\ . \tag{3.1}$$

The solution of (3.1) is given by

$$w_1=(r_1^2V_2-r_1r_2C)/(r_2^2V_1-2r_1r_2C+r_1^2V_2)$$

and

$$w_2=(r_2^2V_1-r_1r_2C)/(r_2^2V_1-2r_1r_2C+r_2^2V_2)\ .$$

By substituting the consistent estimatiors $\hat{d}_1$ and $\hat{d}_2$ of d_1 and d_2 from a good preliminary fit, we get the consistent estimators $\hat{\underline{w}}=(\hat{w}_1, \hat{w}_2)'$.

For SHH, we use $K(0.275)$ instead of L_1 as our preliminary fit to get the ordered residuals. The reason for the use of $K(0.275)$ instead of L_1 is that our pilot Monte Carlo study shows that $K(0.275)$ has a much more reliable slope estimator than L_1 when the underlying distributions are skewed to the right. We use 5% trimming, say p_0=.05 p_1=0.275, p_2=0.95 and sample sizes n=20, 40, 80 respectively. The first column of the design matrix X are equal to 1, and the elements of the second column (x_i) are taken to be typical normal deviates, namely $\Phi^{-1}(i/n+1)$, where Φ^{-1} is the inverse of the standard normal cumulative distribution function Φ. The second column of the design matrix X was standardized so that $\sum x_i^2=1$. The error distributions were generated from four different distributions, ranging from extreme right skewed distribution (like $\chi^2(1)$) to moderate right skewed distribution (like $\chi^2(8)$). Using these scheme cited above, we study the estimated mean square error(MSE) for β_1 under a simple regression model

$$Y_i=\beta_0+\beta_1X_i+Z_i$$

assuming $\beta_0=\beta_1=0$.

If LS_1 is the L.S. slope estimator using observations such that $\underline{x_i}'K(0.05) \le y_i \le \underline{x_i}'K(0.275)$ and LS_2 is the L.S. slope estimator using observations such that $\underline{x_i}'K(0.255) \le y_i \le \underline{x_i}'K(0.95)$, then SHH is defined as

$$\text{SHH} = \frac{\hat{w}_1}{(\hat{w}_1 + \hat{w}_2)} LS_1 + \frac{\hat{w}_2}{(\hat{w}_1 + \hat{w}_2)} LS_2 \ . \tag{3.2}$$

All the calculations needed for regression quantiles follow the algorithm proposed by Barrodale and Roberts(1974). In this study, for comparing with SHH, we include L.S., L_1, $K(.25)$ slope estimator, say, Q25 and RHH. We also include two M-estimators called SBIW and SHT designed for asymmetric error distributions which are proposed by Hogg, et al(1988). We briefly mention these estimators as follows ;

Define MAD = median $|\, r_i - \text{median}(r_i)\,|$, $i = 1, 2, \cdots, n$, where $r_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$, β_0 and β_1 are the $K(.275)$ estimates.

If $z_i = 0.6745 r_i$ /MAD, they use the weight for SBIW as

$$w_i = [1 - (z + 0.6)^2]^2$$

and SHT as

$$w_i = [1 + 5(z + 0.5)^2]^{-1}$$

to calculate these M-estimates. Then 2000 samples were simulated at n=20, n=80 and 4000 samples were simulated at n=40. We summarize the results in Table 1. For sample sizes of 20 and 40, SHH performs best on average efficiency. However, with n=80, SBIW performs best based on average efficiency and average rank. In this comparison, we found out that least squares estimator and L_1 perform extremely bad with very right skewed distributions, like $\chi^2(1)$ and $\chi^2(2)$. For example, for n=80, the efficiency of LS is just 1.4% and 12.2% with underlying distributions, $\chi^2(1)$ and $\chi^2(2)$, respectively. Finally, all estimators are far better than least squares in this overall study. This holds true even for the smaller sample sizes of 20. Least squares is clearly the best estimator when the underlying distribution is close to the normal, but as the distribution gets a little skewed, our estimates SHH or others desigred for asymmetric error distributions do a much better job. This is particularly true for larger sample sizes, like n=80.

Table 1
MSE's and Relative Efficiencies of Estimators
Skewed Right Distributions

	n	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(4)$	$\chi^2(8)$	Ave.Eff.
LS	20	2115 5.4	3961 27.6	7748 59.2	15347 88.7	45.2
	40	1892 2.3	3922 15.8	8403 42.4	15962 76.7	34.3
	80	2034 1.4	3902 12.2	8084 41.5	15886 74.7	
L1	20	1163 9.9	3904 28.0	9638 47.6	22705 60.0	36.4
	40	1034 4.3	3819 15.8	10514 33.9	22746 53.8	27.0
	80	986 2.8	3980 12.0	10045 33.4	23474 50.5	24.7
Q25	20	222 51.8	1585 69.1	5886 77.9	17344 78.5	69.3
	40	150 29.3	1363 45.6	5545 64.3	16708 73.2	53.1
	80	127 22.0	1269 37.5	5583 60.1	16831 70.5	47.5
SBIW	20	115 100	1110 98.6	4998 91.7	17200 79.2	92.4
	40	44 100	621 100	3819 93.4	15160 80.7	93.5
	80	28 100	476 100	3358 100	13988 84.8	96.2
SHT	20	141 81.6	1095 100	4668 98.2	14989 90.8	92.7
	40	66 66.7	693 89.6	3961 90.0	14335 85.3	82.9
	80	43 65.1	505 94.3	3665 91.6	14111 84.1	83.8
RHH	20	414 27.8	1912 57.3	6930 66.2	14419 94.4	61.4
	40	210 21.0	1192 52.1	5399 66.0	12261 99.7	59.7
	80	166 16.9	1042 46.7	4240 79.2	12307 96.4	59.1
SHH	20	124 92.7	1102 99.4	4585 100	13614 100	98.0
	40	53 83.0	665 93.4	3566 100	12227 100	94.1
	80	34 82.4	560 85.0	3393 98.9	11864 100	91.6

In each cell, the first value is 10^3(MSE of the estimator), the standard error of which is usually 2 to 4 percent of that value. The second value is the percent efficiency relative to the estimator having the lowest estimated MSE for that distribution.

References

[1] Adichie, J. N. (1974), "Rank score comparision of several regression parameters." *Annals of Statistics* **2**, 396-402.

[2] Barrodale, I., and Roberts, F. D. K. (1974), "Solution of an overdetermined system of equations in the L_1 Norm." *Communications of the Association for Computing Machinery* **17**, 407-415.

[3] Bickel P. J. (1973), "On some analogues to linear combinations of ordered statistics in the linear model." *Annals of Statistics* **1**, 597-616.

[4] Hettmansperger, T. P., and Mckean J. W. (1977), "A robust alternative based on ranks to least squares in analyzing linear models" *Technometrics* **19**, 274-284.

[5] Hogg, R. V. (1979), "Statistical Robustness : One view of its use in applications today." *The American Statistician* **33**, 108-115.

[6] Hogg, R. V., Bril, G. K., S. M., Han, L., Yuh(1988), "An Argument for Adaptive Robust Estimation." *Probability and Statistics, Essays in Honor of Graybill, F. A.,* North Holland. 135-148.

[7] Huber, P. J. (1973), "Robust regression : Asymptotics, conjectures and Monte Carlo." *Annals of statistics* **1**, 799-821.

[8] Huber, P. J. (1981), *Robust Statistics,* John Wiley.

[9] Jaeckel, L. A. (1972), "Estimating regerssion coefficient by minimizing dispersion of the residuals." *Annals of Mathematical Statistics* **43**, 1449-1458.

[10] Johns, M. V. (1974), "Nonparametric Estimation of Location." *Journal of the American Statistical Association* **69**, 453-460.

[11] Jureckova, J. (1971), "Nonparametric estimate of regression coefficients." *Annals of Mathematical Statistics* **42**, 1328-1338.

[12] Koenker, R., Bassett, G. (1978), "Regression Quantiles." *Econometrica* **46**, 33-50.

[13] Ruppert, D., Carroll, R. J. (1980), "Trimmed least squares estimation in the linear model." *Journal of the American Statistical Association* **75**, 828-838.

비대칭 오차모형하에서의 회귀기울기에 대한 적합된 L-추정법[1]

한 상 문[2]

요 약

회귀모형에 있어서의 Ruppert와 Carroll의 절사 회귀 추정법을 확장하여 회귀 분위수에 의한 두개의 부분으로 관측치를 분할하여 각 부분마다 가중치를 달리 부여하는 방법으로 적합된 L-추정법을 제안하였다.

이 제안된 L-추정법은 특히 비대칭인 오차분포하에서 좋은 효율을 가지고 있었다.

---

[1] 본 연구는 1990년도 서울시 학술진흥연구비에 의해 수행되었음.
[2] (130-743) 서울특별시 동대문구 전농동 8-3 서울시립대학교 전산통계학과