The Annals of Statistics
1978, Vol. 6, No. 3, 603–609

WEIGHTED MEDIAN REGRESSION ESTIMATES[1]

BY FRIEDRICH-WILHELM SCHOLZ

University of Washington

In the simple linear regression problem $\{Y_i = \alpha + \beta x_i + e_i \; i = 1, \cdots, n, e_i$ i.i.d. $\sim F$ continuous, $x_1 \leqq \cdots \leqq x_n$ known, α, β unknown$\}$ we investigate the following type of estimator: To each $s_{ij} = (Y_j - Y_i)/(x_j - x_i)$ with $x_i < x_j$ attach weight w_{ij} and as estimator for β consider the median of this weight distribution over the s_{ij}. A confidence interval for β is found by taking certain quantiles of this weight distribution. The asymptotic behavior of both is investigated and conditions for optimal weights are given.

1. Introduction and summary. In the simple linear regression problem $\{Y_i = \alpha + \beta x_i + e_i \; i = 1, \cdots, n, e_i$ i.i.d. $\sim F$ continuous, $x_1 \leqq \cdots \leqq x_n$ known, α, β unknown parameters$\}$ the following estimator for β was introduced by Theil (1950) and more generally investigated by Sen (1968): Compute $s_{ij} = (Y_j - Y_i)/(x_j - x_i)$ for $x_i < x_j$ and use as estimator $\tilde{\beta} =$ median $\{s_{ij} : x_i < x_j\}$. Sen found the asymptotic relative efficiency (ARE) of $\tilde{\beta}$ relative to the ordinary least squares estimator $\hat{\beta}$. This ARE is largest, say e_0, whenever the x's are equally spaced or whenever the x's are distributed over only two distinct values in which case the problem reduces to the two-sample location problem.

Jaeckel (1972), when considering the above problem, proposed as estimator of β that value of b which minimizes some dispersion measure of the residuals $Y_i - bx_i$. In a special case he arrives at the following variant of the estimator $\tilde{\beta}$: Order the s_{ij} and attach weights w_{ij} to each s_{ij} and take as estimator $\tilde{\beta}'$ the median of this weight distribution over the s_{ij}. Under some restrictive conditions and using weights proportional to $x_j - x_i$, Jaeckel obtains that $\tilde{\beta}'$ has ARE e_0 with respect to $\hat{\beta}$ regardless of the design of the x's.

In this paper we investigate the asymptotic behavior of Jaeckel's variant under weaker conditions for general w_{ij}. This includes the Theil estimator as well ($w_{ij} \sim$ const.). It is shown that one cannot improve on e_0. Conditions on the weights are given under which e_0 is achieved; however, the optimal weights are not unique, as already the design of equally spaced x's shows, where both $\tilde{\beta}$ and $\tilde{\beta}'$ achieve e_0. It is further shown that any two such weighted median estimators which achieve e_0 will be asymptotically equivalent.

In the case of constant weights Sen also investigated the asymptotic behavior of confidence intervals for β based on the s_{ij} as suggested by Theil, and his efficiency results parallel those of the estimator. Confidence intervals for β were

Received November 1976; revised September 1977.

[1] This research was partially supported by NSF grant MCS75-08557 A01.

AMS 1970 *subject classifications.* Primary 62G05, 62G15, 62G20, 62J05; Secondary 62G35.

Key words and phrases. Estimation, confidence interval, linear regression, efficiency.

not given by Jaeckel and are provided here for general weights. The efficiency results match those of the estimation case.

2. Definition and asymptotic behavior of the weighted median regression estimator. Let $Y_i = \alpha + \beta x_i + e_i$ $i = 1, 2, \cdots, n$, $e_1, \cdots, e_n$ be independent identically distributed random variables with continuous distribution function F. $x_1 \leqq \cdots \leqq x_n$ are known constants with $b_n{}^2 = \sum_i (x_i - \bar{x})^2 > 0$, α and β are unknown parameters, of which we will estimate β. For $x_i < x_j$ consider the pairwise slopes

$$s_{ij} = (Y_j - Y_i)/(x_j - x_i) = \beta + (e_j - e_i)/(x_j - x_i) .$$

Let $w = \{w_{ij} : i \leqq j\}$ be a set of weights with the following properties: $w_{ij} \geqq 0$, $w_{ij} = 0$ whenever $x_i = x_j$ and $\sum_{i \leqq j} w_{ij} = 1$. We extend the definition of w_{ij} to all $1 \leqq i, j \leqq n$ as follows:

$$w_{ij} = -w_{ji} \qquad \text{for} \quad i \geqq j .$$

Define $G(t) = \sum_{i<j} w_{ij} I_{[s_{ij} \leqq t]}$ with

$$\begin{aligned} I_{[s_{ij} \leqq t]} &= 1 \qquad \text{if} \quad s_{ij} \leqq t \\ &= 0 \qquad \text{else.} \end{aligned}$$

$G(t)$ is the distribution function of the probability distribution of weights w_{ij} over the s_{ij}, $i < j$. Let

$$\begin{aligned} \beta_1 &= \inf \{t : G(t) \geqq .5\} \\ \beta_2 &= \sup \{t : G(t) \leqq .5\} . \end{aligned}$$

Corresponding to the weights w define the estimator $\tilde{\beta}_w$ of β by $\tilde{\beta}_w = (\beta_1 + \beta_2)/2$.

Although asymptotic considerations would require an additional index n on the x_i, w_{ij}, s_{ij}, etc., we omit it for notational convenience.

The asymptotic normality of $\tilde{\beta}_w$ will be established under the following set (C) of conditions:

(C)

C1: $H(t) = \int F(t + x)\, dF(x)$ has a positive derivative at $t = 0$; denote it by $H'(0)$.

C2: With $W_n = \sum_{i=1}^n (\sum_{j=1}^n w_{ij})^2$, assume

$$\begin{aligned} n \textstyle\sum_{i=1}^n \sum_{j=1}^n w_{ij}^2 / W_n &= O(1) \qquad \text{as} \quad n \to \infty \\ \sup_{1 \leqq j \leqq n} (\textstyle\sum_i |w_{ij}|)^2 / W_n &= o(1) . \end{aligned}$$

C3: With $b_n{}^2 = \sum_{i=1}^n (x_i - \bar{x})^2$ assume that

$$\rho_n = \textstyle\sum_{i<j} w_{ij} (x_j - x_i) / (W_n^{\frac{1}{2}} b_n)$$

stays bounded away from zero as $n \to \infty$; further assume that

$$\max_{1 \leqq j \leqq n} |x_j - \bar{x}| / c_n = o(1) \quad \text{as} \quad n \to \infty \qquad \text{where} \quad c_n = b_n \rho_n .$$

COMMENTS. For $w_{ij} = c(x_j - x_i)$ $i \leqq j$, $\tilde{\beta}_w$ is the estimator obtained by

Jaeckel. One easily checks $n \sum_{ij} w_{ij}^2/W_n = 2$ and $\sup_j (\sum_i |w_{ij}|)^2/W_n \leqq n^{-1} + \sup_j (x_j - \bar{x})^2/b_n^2$. Thus (C) is satisfied whenever C1 and C3 hold, since $\rho_n \leqq 1$. For $w_{ij} = c$ whenever $x_i < x_j$, β_w is the estimator proposed by Theil and Sen. If there are $k > 1$ distinct values $t_1 < \cdots < t_k$ among the x's, denote by u_i the multiplicity of t_i among the x's. One easily checks that

$$n \sum_{ij} w_{ij}^2/W_n = 3(1 - \sum_i (u_i/n)^2)/(1 - \sum_i (u_i/n)^3) \leqq 3$$

and

$$\sup_j (\sum_i |w_{ij}|)^2/W_n = 3 \max_j (n - u_j)^2/(n^3 - \sum_i u_i^3) = o(1)$$

iff $\min_i (n - u_i) \to \infty$ as $n \to \infty$. Hence (C) is equivalent to C1, C3 and $\min_i (n - u_i) \to \infty$ as $n \to \infty$.

THEOREM 1. *For a given set of weights $w = \{w_{ij} : i \leqq j\}$ which satisfies* (C) *we have*

$$c_n(\tilde{\beta}_w - \beta) \to_L N(0, (12H'(0)^2)^{-1}) .$$

PROOF.

$$\begin{aligned} & P(\textstyle\sum_{i<j} w_{ij} I_{[e_j - e_i \leq t_{nij}]} \geqq .5 + \max_{i<j} w_{ij}) \\ (1) \qquad & \quad \leqq P(c_n(\tilde{\beta}_w - \beta) \leqq t) \leqq P(\textstyle\sum_{i<j} w_{ij} I_{[e_j - e_i \leq t_{nij}]} \geqq .5) \\ & \qquad \text{with} \quad t_{nij} = t(x_j - x_i)/c_n . \end{aligned}$$

We will show that

$$(2) \qquad (12)^{\frac{1}{2}}(T_n - ET_n)/W_n^{\frac{1}{2}} \to_L N(0, 1) \qquad \text{as} \quad n \to \infty$$

where $T_n = \sum_{i<j} w_{ij} U_{ij}$ with $U_{ij} = I_{[e_j - e_i \leq t_{nij}]}$. Proceeding with the projection method of Hájek (1968), we will approximate $S_n = T_n - ET_n$ by $\hat{S}_n = \sum_{k=1}^n E(S_n \mid e_k)$ yielding $E(S_n - \hat{S}_n)^2 = \operatorname{Var} S_n - \operatorname{Var} \hat{S}_n$. Lengthy but straightforward manipulations show that

$$\begin{aligned} (3) \qquad \operatorname{Var} S_n - \operatorname{Var} \hat{S}_n &= \textstyle\sum_{i<j} w_{ij}^2 [E \operatorname{Var} (U_{ij} \mid e_i) + E \operatorname{Var} (U_{ij} \mid e_j)] \\ &\leqq \textstyle\sum_{i<j} w_{ij}^2/2 . \end{aligned}$$

Further one shows easily that $\operatorname{Var} \hat{S}_n = W_n/12 + \sum_i \delta_{in}$ with $|\delta_{in}| \leqq 4 \sup_x |F(x + \Delta_n) - F(x)| \cdot (\sum_j |w_{ij}|)^2$ where $\Delta_n = \max_{i<j} |t_{nij}| = o(1)$ as $n \to \infty$ by C3.

Hence

$$(4) \qquad \operatorname{Var} \hat{S}_n = \tfrac{1}{12} W_n + o(1) \textstyle\sum_i (\sum_j |w_{ij}|)^2 = \tfrac{1}{12} W_n (1 + o(1)) \quad \text{by C2} .$$

(3), (4) and C2 yield

$$\begin{aligned} (5) \qquad E(S_n - \hat{S}_n)^2/\operatorname{Var} \hat{S}_n &= (\operatorname{Var} S_n - \operatorname{Var} \hat{S}_n)/\operatorname{Var} \hat{S}_n \\ &\leqq \textstyle\sum_{i,j} w_{ij}^2/(4 \operatorname{Var} \hat{S}_n) = o(1) . \end{aligned}$$

Denoting the distribution of $Y_{nj} = E(S_n \mid e_j)/(\operatorname{Var} \hat{S}_n)^{\frac{1}{2}}$ by F_{nj} and observing

$$Y_{nj}^2 \leqq 12(\textstyle\sum_i (|w_{ij}|)^2/W_n(1 + o(1)) ,$$

we find by C2:

$$\forall \eta > 0 \quad \textstyle\sum_j \int_{|y| > \eta} y^2 \, dF_{nj} \leqq 12 \sup_j (\sum_i |w_{ij}|)^2/(W_n \eta^2 (1 + o(1))) = o(1) ;$$

hence by the Lindeberg–Feller central limit theorem

$$\hat{S}_n/(\text{Var}\,\hat{S}_n)^{\frac{1}{2}} \to_L N(0, 1) \quad \text{as} \quad n \to \infty \tag{6}$$

which with (4) and (5) yields (2).

Since $\max_{i<j} w_{ij}^2/W_n = o(1)$ by C2, (1) yields $\lim_{n\to\infty} P(c_n(\hat{\beta}_w - \beta) \leqq t) = \Phi(\lim_{n\to\infty} (ET_n - .5)(12)^{\frac{1}{2}}/W_n^{\frac{1}{2}})$ provided the limit on the right exists. Φ is the standard normal distribution function. Now $ET_n - .5 = A_n + B_n$ with $B_n = \sum_{i<j} w_{ij} t_{nij} H'(0)$ and

$$\begin{aligned}|A_n| &= \left|\sum_{i<j} w_{ij} \left(\frac{H(t_{nij}) - H(0)}{t_{nij}} - H'(0)\right) t_{nij}\right| \\ &\leqq o(1) \sum_{i<j} w_{ij}(x_j - x_i)|t|/c_n \quad \text{by C1}\end{aligned}$$

thus $|A_n|/W_n^{\frac{1}{2}} = o(1)$ and $\lim_{n\to\infty} (ET_n - .5)(12)^{\frac{1}{2}}/W_n^{\frac{1}{2}} = H'(0)t \cdot (12)^{\frac{1}{2}}$. This concludes the proof.

3. Confidence intervals for β. From Section 2 we know the following: if $G_n(t) = \sum_{i<j} w_{ij} I_{[s_{ij} \leq t]}$ then $\sigma_n^2 = \text{Var}\, G_n(\beta) = \frac{1}{12} W_n(1 + o(1))$, $EG_n(\beta) = .5$ and

$$(G_n(\beta) - .5)/\sigma_n \to_L N(0, 1) \quad \text{as} \quad n \to \infty\,.$$

Define $G_n^{-1}(u) = \inf\{t : G_n(t) \geqq u\}$ for $u \in (0, 1)$, and let $S_1 = G_n^{-1}(.5 + z_{\alpha/2}\sigma_n)$ and $S_2 = G_n^{-1}(.5 + z_{1-\alpha/2}\sigma_n)$ where $z_u = \Phi^{-1}(u)$.

Then for large n we have

$$1 - \alpha \cong P(z_{\alpha/2} \leqq (G_n(\beta) - .5)/\sigma_n < z_{1-\alpha/2}) = P(S_1 \leqq \beta \leqq S_2)\,,$$

i.e., we may consider $[S_1, S_2]$ as a large sample confidence interval for β with approximate confidence level $1 - \alpha$. S_1 and S_2 are particular ordered slopes s_{ij}.

THEOREM 2. *For weights w satisfying conditions* (C) *we have* $c_n(S_2 - S_1) \to_P z_{1-\alpha/2}/(3^{\frac{1}{2}}H'(0))$.

PROOF. As in the proof of Theorem 1 we obtain by C1 and C3:

$$E[G_n(\beta + t/c_n) - G_n(\beta)]/W_n^{\frac{1}{2}} = t \cdot H'(0)(1 + o(1)) \quad \text{as} \quad n \to \infty$$

and by C2 and C3:

$$\text{Var}\,[(G_n(\beta + t/c_n) - G_n(\beta))/W_n^{\frac{1}{2}}] = o(1) \quad \text{as} \quad n \to \infty$$

hence $\forall\, t \in R$

$$(G_n(\beta + t/c_n) - G_n(\beta))/W_n^{\frac{1}{2}} \to_P tH'(0) \quad \text{as} \quad n \to \infty\,. \tag{7}$$

In (7) the left side is nondecreasing in t, the limit on the right is a nondecreasing and continuous function of t. Using the stochastic version of Pólya's theorem, we can conclude

$$\sup_{|t| \leq k} |(G_n(\beta + t/c_n) - G_n(\beta))/W_n^{\frac{1}{2}} - tH'(0)| \to_P 0 \quad \text{as} \quad n \to \infty\,. \tag{8}$$

(4) and (6) of the previous section yield that $c_n(S_i - \beta)$ $i = 1, 2$ are asymptotically

normally distributed, hence stay bounded in probability. This and (8) imply that $(G_n(S_i) - G_n(\beta))/W_n^{\frac{1}{2}} = c_n(S_i - \beta) \cdot H'(0) + R_{ni}$ $i = 1, 2$ with $R_{ni} \to_P 0$ as $n \to \infty$, which together with

$$|G_n(S_1) - .5 - z_{\alpha/2}\sigma_n|/W_n^{\frac{1}{2}} \leqq \sup_{i<j} w_{ij}/W_n^{\frac{1}{2}} = o(1)$$

and

$$|G_n(S_2) - .5 - z_{1-\alpha/2}\sigma_n|/W_n^{\frac{1}{2}} = o(1) \qquad \text{as} \quad n \to \infty$$

yields the assertion of the theorem.

4. Optimal weights. Since Theorem 1 allows us to approximate the distribution of $\tilde{\beta}_w$ by a normal distribution with mean β and variance $((12)^{\frac{1}{2}}H'(0)b_n\rho_n)^{-2}$ it is of interest to note that

$$\text{(9)} \qquad \rho_n \leqq 1 \quad \text{with equality iff} \quad (\textstyle\sum_j w_{1j}, \cdots, \sum_j w_{nj}) = \lambda(\bar{x} - x_1, \cdots, \bar{x} - x_n) \quad \text{for some} \quad \lambda \in R\,.$$

This follows from $\sum_{i<j} w_{ij}(x_j - x_i) = \sum_i ((\bar{x} - x_i) \sum_j w_{ij}) \leqq b_n W_n^{\frac{1}{2}}$. That the optimal weights, which achieve $\rho_n = 1$, are not unique is quite evident from (9). How different the optimal weights can be is shown in the special case where $x_i = i$ $i = 1, \cdots, n$; here $w = \{c(j - i) : i \leqq j\}$ and $w' = \{c' : i \leqq j\}$ represent optimal weights. This raises the question whether $(\tilde{\beta}_w + \tilde{\beta}_{w'})/2$ would be an improved estimator. The following theorem answers this question in the negative.

THEOREM 3. *Let $w_k = \{w_{ijk} : i \leqq j\}$ $k = 1, 2$ be two weight systems satisfying condition* (C) *and $\rho_{nk} = 1$ $k = 1, 2$ where ρ_{nk} corresponds to w_k in the fashion of* C3. *Then*

$$b_n(\tilde{\beta}_{w_1} - \tilde{\beta}_{w_2}) \to_P 0 \qquad \text{as} \quad n \to \infty\,,$$

i.e., *the two estimators are asymptotically equivalent.*

PROOF. Let $G_{kn}(t) = \sum_{i<j} w_{ijk} I_{[s_{ij} \leqq t]}$ $k = 1, 2$, and set $G_{kn}(\beta + t_k/c_{nk}) = T_{kn}$ $k = 1, 2$ with $t_k \in R$, $c_{nk} = \rho_{nk} b_n$.

From the proof of Theorem 1 we have

$$\sigma^2(T_{kn}) = \tfrac{1}{12} \textstyle\sum_{i=1}^n (\sum_{j=1}^n w_{ijk})^2 (1 + o(1)) \quad k = 1, 2\,,$$

and after straightforward manipulations, one arrives at

$$\text{(10)} \qquad |\mathrm{Cov}\,(T_{1n}, T_{2n}) - \tfrac{1}{12} \textstyle\sum_i (\sum_j w_{ij1} \sum_j w_{ij2}) - \tfrac{1}{3} \sum_{i<j} w_{ij1} w_{ij2}| \leqq o(1) \sum_i (\sum_j |w_{ij1}| \sum_j |w_{ij2}|)\,.$$

Note that $\sum_i (\sum_j |w_{ij1}| \sum_j |w_{ij2}|)/(\sigma(T_{1n})\sigma(T_{2n}))$ stays bounded by C2, and

$$\tfrac{1}{12} \textstyle\sum_i (\sum_j w_{ij1} \sum_j w_{ij2})/(\sigma(T_{1n})\sigma(T_{2n})) = 1 + o(1) \qquad \text{by} \quad \rho_{n1} = \rho_{n2} = 1 \quad \text{and (9)}\,.$$

Since further $\sum_{i<j} w_{ij1} w_{ij2}/(\sigma(T_{1n})\sigma(T_{2n})) = o(1)$ by C2 we can conclude from (10) that

$$\mathrm{Cov}\,(T_{1n}, T_{2n})/(\sigma(T_{1n})\sigma(T_{2n})) = 1 + o(1)$$

which implies

$$(T_{1n} - ET_{1n})/\sigma(T_{1n}) - (T_{2n} - ET_{2n})/\sigma(T_{2n}) \to_P 0 \qquad \text{as} \quad n \to \infty .$$

From the proof of Theorem 1 we conclude that

$$\left(\frac{T_{1n} - ET_{1n}}{\sigma(T_{1n})}, \frac{T_{2n} - ET_{2n}}{\sigma(T_{2n})}\right) \to_L N(\binom{0}{0}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix})$$

which in turn implies the same limit law for

$$(12)^{\frac{1}{2}} H'(0) \cdot (b_n(\tilde{\beta}_{w_1} - \beta), b_n(\tilde{\beta}_{w_2} - \beta)) .$$

This concludes the proof.

REMARK. In the equally spaced design $x_i = i$ $i = 1, 2, \cdots, n$, we have the following two optimal weighting designs: $w_{ij} = (j - i) \cdot 6/(n(n^2 - 1))$, $i < j$ and $w'_{ij} = 2/(n(n - 1))$, $i < j$. In the spirit of Hodges (1967) one can examine the tolerance of $\tilde{\beta}_w$ and $\tilde{\beta}_{w'}$ for extreme values of the y-observations. One easily finds that the tolerance of $\tilde{\beta}_w$ is about 20.6% ($= 100 \cdot (1 - (.5)^{\frac{1}{3}})$%) and the tolerance of $\tilde{\beta}_{w'}$ is about 29.3% ($= 100 \cdot (1 - (.5)^{\frac{1}{2}})$%). On these grounds it appears that $\tilde{\beta}_{w'}$ is to be preferred although both estimators are asymptotically equivalent by the previous theorem.

For a given weight system w the computation of $\tilde{\beta}_w$ involves as many slopes s_{ij} as there are positive w_{ij}. For ease of computation it would therefore be of interest to find a weight system w_0 satisfying (C) such that a maximal number of weights are zero. For example in the case of a symmetric design, i.e., x_i are symmetric around $\bar{x}$, one might try $w_{i,n-i+1} = \bar{x} - x_i$ for $i = 1, \cdots, k$, $k = [n/2]$ and $w_{ij} = 0$ else for $i \leqq j$, then $\rho_n = 1$. However, C2 is not satisfied since $n \sum_i \sum_j w_{ij}^2/W_n = n$. In fact it can be shown that $b_n(\tilde{\beta}_w - \beta) \to_L N(0, (8H'(0)^2)^{-1})$, i.e., the asymptotic variance is increased by a factor of 1.5 over the asymptotic variance obtained for optimal weights satisfying (C). However, the lower bound of the ARE of this estimator compared with the ordinary least squares estimator of β is still bounded below by $.864 \cdot \frac{2}{3} = .576$, a price that one might be willing to pay for a quickly computed and fairly robust estimator of β. The author has investigated the use of appropriate linear combinations of two such "quick" estimators, corresponding to two "quick" weighing designs. The efficiency was somewhat improved but not much. Finally we can state that the efficiency results concerning the estimators carry over to the confidence intervals as well.

REFERENCES

HÁJEK, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

HODGES, J. L. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 163–186.

JAECKEL, L. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43** 1449–1458.

SEN, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.* **63** 1379–1389.

THEIL, H. (1950). A rank invariant method of linear and polynomial regression analysis, I, II, III. *Proc. Kon. Nederl. Akad. Wetensch. A.* **53** 386–392, 521–525, 1397–1412.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195