The Annals of Mathematical Statistics
1969, Vol. 40, No. 5, 1784–1790

A ROBUST POINT ESTIMATOR IN A GENERALIZED REGRESSION MODEL

By P. V. Rao and J. I. Thornby

University of Florida

1. Introduction and summary. The objective of the present paper is to define a robust point estimator of the parameter β in the model

$$y_j = \alpha + g_j(\beta) + z_j, \qquad j = 1, 2, \cdots, n, \tag{1.1}$$

where α and β are unknown parameters, $g_1, g_2, \cdots, g_n$ are real-valued functions of real variable satisfying suitable conditions and $z_1, z_2, \cdots, z_n$ are independent identically distributed random variables having a distribution function belonging to a specified class.

An important special case of (1.1) is the regression model obtained by taking $g_j(\beta) = \beta x_j, j = 1, 2, \cdots, n$, where the x's are known constants. Robust point estimators of β in this case have already been given by Adichie [1], who followed the method of Hodges and Lehmann and by Brown and Mood [7], who considered the so-called "median" estimators. The estimator presented in this paper, which is also a Hodges-Lehmann type estimator, therefore, provides a third alternative for the regression model.

In Section 2 of this paper, a robust point estimator for β in the model (1.1) under some suitable regularity conditions on g_j and z_j is defined. In Section 3, a simple computational technique for the calculation of this estimator is given. A small sample property of the estimator is given in Section 4, and in Section 5, asymptotic normality is established under some regularity conditions. In Section 6, some special cases of the model are considered.

2. The model and the estimator. Let $y_1, y_2, \cdots, y_n$ be independent random variables with distribution function:

$$P_{\alpha\beta}(y_j \leqq y) = F[y - \alpha - g_j(\beta)], \qquad j = 1, 2, \cdots, n, \tag{2.1}$$

where $P_{\alpha\beta}(\ )$ indicates that the probability is computed for parameter values α and β, $g_1, \cdots, g_n$ are known functions of β that are not all equal, and, for all $i < j$, the functions $g_{ij}(\beta) = g_i(\beta) - g_j(\beta)$ are either all never-increasing or all never-decreasing. Without loss of generality, we may assume that the functions g_{ij} are never-decreasing, the other case being obtained by simply replacing the subscript j by $n - j + 1$ in $y_j, j = 1, 2, \cdots, n$. The parameter β will be assumed to belong to a bounded interval Ω of the real line.

Now for each pair $1 \leqq i < j \leqq n$, and $a \in \Omega$ let,

$$\begin{aligned} Z_{ij}(a) &= 1 \quad \text{if } y_i - y_j \leqq g_{ij}(a), \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{2.2}$$

Received 5 August 1968; revised 18 April 1969.

Put $h_n(a) = \sum_{1 \leqq i<j \leqq n} Z_{ij}(a)$, and define β_n^* and β_n^{**} by

$$\beta_n^* = \sup_{a \in \Omega} \{a : h_n(a) < \tfrac{1}{4}n(n-1)\},$$
$$\beta_n^{**} = \inf_{a \in \Omega} \{a : h_n(a) > \tfrac{1}{4}n(n-1)\}.$$

Then $\hat{\beta}_n$ defined by

$$\hat{\beta}_n = \tfrac{1}{2}(\beta_n^* + \beta_n^{**}), \tag{2.3}$$

will be taken as a point estimator for β.

Clearly, $Z_{ij}(a)$ is never-decreasing in a for all $i < j$ so that $h_n(a)$ is also never-decreasing. This implies that β_n^*, β_n^{**} and $\hat{\beta}_n$ are well defined and that

$$\beta_n^* \leqq \hat{\beta}_n \leqq \beta_n^{**}. \tag{2.4}$$

3. Computation of $\hat{\beta}_n$. A simple method for computing the estimator $\hat{\beta}_n$ is given in Theorem 1.

THEOREM 1. *Let y_j and $g_j(j = 1, 2, \cdots, n)$ be as in (2.1) and for each $i < j$ define*[1]

$$a_{ij} = \inf_{a \in \Omega} \{a : y_i - y_j \leqq g_{ij}(a)\}, \qquad 1 \leqq i < j \leqq n. \tag{3.1}$$

Then $\hat{\beta}_n$ is the median of the set of $\frac{1}{2}n(n-1)$ numbers a_{ij}, $1 \leqq i < j \leqq n$.

PROOF. The proof is analogous to the proof of equation (4.1) of Hodges and Lehmann [6].

As an application of Theorem 1, consider the problem of computing $\hat{\beta}_n$ for Graybill's data given in [1]. If we take $g_j(\beta) = x_j\beta$, where $x_1 > x_2 > \cdots > x_7$, we can easily see that g_{ij} is never-decreasing for all $i < j$ and that $a_{ij} = (y_i - y_j)/(x_i - x_j)$. Hence $\hat{\beta}_n$ is the median of the numbers $(y_i - y_j)/(x_i - x_j)$; $1 \leqq i < j \leqq 7$, i. e.

$$\hat{\beta}_n = \mathrm{Med}_{i<j} \{(y_i - y_j)/(x_i - x_j)\},$$

which can be easily seen to be equal to 4.00, a value equal to Adichie's [1] estimate of β for the same data.

4. Median unbiasedness. In general, the estimator $\hat{\beta}_n$ will neither be unbiased, nor possess invariance and symmetry properties similar to those given in Lemmas 4.1 and 4.2 of [1]. However, median unbiasedness of $\hat{\beta}_n$ can be established for many cases as can be seen from Theorem 2.

THEOREM 2. *Let F be continuous and $P_\beta(\ )$ indicate that the probability is computed with parameter value β fixed. Then,*

$$\tfrac{1}{2}\{1 - P_\beta[h_n(\beta) = \mu_n]\} \leqq P_\beta(\hat{\beta}_n \leqq \beta) \leqq \tfrac{1}{2}\{1 + P_\beta[h_n(\beta) = \mu_n]\},$$

where $\mu_n = \frac{1}{4}n(n-1)$. If $n(n-1) \not\equiv 0 \bmod 4$, then

$$P_\beta(\hat{\beta}_n \leqq \beta) = \tfrac{1}{2},$$

so that β is a median of $\hat{\beta}_n$.

[1] If $y_i - y_j > g_{ij}(a)$ for all $a \in \Omega$, then we put $a_{ij} = \sup \Omega$.

PROOF. The proof is analogous to the proof of Lemma 5 of Hodges and Lehmann [6].

It must be pointed out that even in the case where $n(n-1) = 0 \bmod 4$, $P_\beta[h_n(\beta) = \mu_n]$ will be typically small, so that $\hat{\beta}_n$ is at least "approximately" median unbiased.

5. Asymptotic normality. We shall now consider the large sample distribution of the estimator $\hat{\beta}_n$ defined by (2.2). Theorem 3 gives a set of conditions under which $\hat{\beta}_n$ will have an asymptotic normal distribution.

THEOREM 3. *In* (2.1) *assume the following*:

(I) *F is absolutely continuous with an absolutely continuous, square integrable density f.*

(II) (a) *There exists a function G such that $\lim_{\epsilon\to 0} G(\epsilon) = 0$ and for all $i < j$, a and u, $|g_{ij}(a+u) - g_{ij}(a)| \leqq G(u)$, and*

(b) *For the parameter value β, there exists a constant $\Delta(\beta)$ satisfying,*

$$\lim_{n\to\infty} 6n^{-\frac{3}{2}} \sum_{i<j} [g_{ij}(\beta + un^{-\frac{1}{2}}) - g_{ij}(\beta)] = u\Delta(\beta) \quad \text{for all} \quad u.$$

Then

$$\lim_{n\to\infty} P_\beta(n^{\frac{1}{2}}(\hat{\beta}_n - \beta)/K(\beta) \leqq u) = \Phi(u), \tag{5.1}$$

where Φ is the normal distribution function with mean 0 and variance 1 and

$$K(\beta) = [\Delta(\beta) \int_{-\infty}^{+\infty} f^2(x)\, dx]^{-1}. \tag{5.2}$$

The proof of Theorem 3 will depend on the following lemma. Let $h_n(a)$ be defined as in Section 2 and $\mu_{\beta n}(a)$ and $\sigma^2_{\beta n}(a)$ denote the mean and variance, respectively, of $h_n(a)$ for the parameter value β.

LEMMA 1. *Under the assumptions of Theorem* 3, *the statistic*

$$\xi_n(u) = [h_n(\beta + un^{-\frac{1}{2}}) - \mu_{\beta n}(\beta + un^{-\frac{1}{2}})]/\sigma_{\beta n}(\beta) \tag{5.3}$$

is asymptotically normal with mean 0 *and variance* 1, *and*

$$\lim_{n\to\infty} \alpha_n(u) = -u[K(\beta)]^{-1}, \tag{5.4}$$

where,

$$\alpha_n(u) = [\mu_{\beta n}(\beta) - \mu_{\beta n}(\beta + un^{-\frac{1}{2}})]/\sigma_{\beta n}(\beta).$$

PROOF. We may write,

$$\begin{aligned}\xi_n(u) &= \{h_n(\beta) - \mu_{\beta n}(\beta)\}/\sigma_{\beta n}(\beta) + \{H_n(\beta + un^{-\frac{1}{2}}) - E_\beta[H_n(\beta + un^{-\frac{1}{2}})]\}/\sigma_{\beta n}(\beta) \\ &= \eta_n + \tau_n, \quad \text{say,}\end{aligned}$$

where,

$$H_n(\beta + un^{-\frac{1}{2}}) = h_n(\beta + un^{-\frac{1}{2}}) - h_n(\beta),$$

and $E_\beta(\ )$ stands for expectation with parameter value β. It is well known ([3] page 241) that η_n is asymptotically normal with mean 0 and variance 1. Hence our

proof is complete ([2] page 254) if we show that plim $\tau_n = 0$. We shall establish this by showing $V_\beta(\tau_n) \to 0$ as $n \to \infty$, where $V_\beta(\ )$ stands for the variance with parameter value β.

Since $\sigma^2_{\beta n}(\beta) \sim n^3/36$ ([3], page 241), it follows that

$$V_\beta(\tau_n) \sim 36n^{-3}V_\beta[H_n(\beta + un^{-\frac{1}{2}})].$$

For $u \geqq 0$, writing,

$$H_n(\beta + un^{-\frac{1}{2}}) = \sum_{i<j} H_{ij}(\beta + un^{-\frac{1}{2}}),$$

where

$$(5.5) \qquad H_{ij}(\beta + un^{-\frac{1}{2}}) = 1 \quad g_{ij}(\beta) < y_i - y_j \leqq g_{ij}(\beta + un^{-\frac{1}{2}}),$$
$$= 0 \quad \text{otherwise;}$$

we get

$$(5.6) \quad V_\beta(H_n) = \sum_{i<j} V_\beta(H_{ij}) + 2\sum_{i<j<k} \mathrm{Cov}_\beta(H_{ij}, H_{ik}) + 2\sum_{i<j<k} \mathrm{Cov}_\beta(H_{ij}, H_{jk}) + 2\sum_{i<j<k} \mathrm{Cov}_\beta(H_{ik}, H_{jk}),$$

where $H, H_{ij}, \cdots$ etc., have obvious meanings. Hence,

$$V_\beta(H_n) \sim cn\sum_{i<j} P_\beta[g_{ij}(\beta) < y_i - y_j \leqq g_{ij}(\beta + un^{-\frac{1}{2}})]$$
$$= cn\sum_{i<j} \int_{-\infty}^{+\infty} [F(y + g_{ij}(\beta + un^{-\frac{1}{2}}) - g_{ij}(\beta)) - F(y)]\, dF(y)$$

where c is a generic constant. Expanding the integrand above in a Taylor series we obtain,

$$V_\beta(H_n) \sim cn\sum_{i<j}[g_{ij}(\beta + un^{-\frac{1}{2}}) - g_{ij}(\beta)] \int_{-\infty}^{+\infty} f(y + \gamma_{ij}(y))\, dF(y),$$

where

$$(5.7) \qquad |\gamma_{ij}(y)| \leqq |g_{ij}(\beta + un^{-\frac{1}{2}}) - g_{ij}(\beta)|.$$

By (5.7) and II (a) of Theorem 3, we have $\lim_{n\to\infty} \gamma_{ij}(y) = 0$ uniformly in i, j and y. This fact, together with the absolute continuity of f and II (b) of Theorem 3, implies

$$V_\beta(H_n) \sim cn^{\frac{5}{2}}u[K(\beta)]^{-1},$$

so that $V_\beta(H_n) = o(n^3)$. Thus $V_\beta(\tau_n) = o(1)$ if $u \geqq 0$. That the same result will hold if $u < 0$ can be easily seen by making obvious modification in the definition of H_{ij} given in (5.5). Hence the proof of the first part of Lemma 1 is complete.

To prove the second part, note that

$$-\alpha_n(u) = 6n^{-\frac{5}{2}}\sum_{i<j} P_\beta[g_{ij}(\beta) < y_i - y_j \leqq g_{ij}(\beta + un^{-\frac{1}{2}})],$$

which tends to $u[K(\beta)]^{-1}$ as can be seen from the proof of the first part. This completes the proof of Lemma 1.

Proof of Theorem 3. Since $\beta_n^* \leqq \hat{\beta}_n \leqq \beta_n^{**}$, we have

$$(5.8) \qquad P_\beta(\beta_n^{**} \leqq a) \leqq P_\beta(\hat{\beta}_n \leqq a) \leqq P_\beta(\beta_n^* \leqq a)$$

for all a. Also, $h_n(a) > \mu_{\beta n}(\beta) \Rightarrow \beta_n^{**} \leqq a$ and $\beta_n^{*} \leqq a \Rightarrow h_n(a) \geqq \mu_{\beta n}(\beta)$. Hence from (5.8) we get

$$P_\beta[h_n(a) > \mu_{\beta n}(\beta)] \leqq P_\beta[\hat{\beta}_n \leqq a] \leqq P_\beta[h_n(a) \geqq \mu_{\beta n}(\beta)]. \tag{5.9}$$

By taking $a = \beta + un^{-\frac{1}{2}}$ in (5.9) it is easy to see that

$$P_\beta[\xi_n(u) > \alpha_n(u)] \leqq P_\beta[n^{\frac{1}{2}}(\hat{\beta}_n - \beta) \leqq u] \leqq P_\beta[\xi_n(u) \geqq \alpha_n(u)].$$

Theorem 3 now follows from Lemma 1.

6. Some particular cases of the model. Some interesting particular cases of the general model defined in Section 2 are obtained by taking (i) $g_j(\beta) = x_j\beta$, (ii) $g_j(\beta) = K|x_j - \beta|$, (iii) $g_j(\beta) = K\beta^{x_j}$, $\beta \geqq 1$ and (iv) $g_j(\beta) = K \log (\beta + x_j)$, $\beta > 0$, where $x_1, x_2, \cdots, x_n$ and K are known real numbers. It is clear that (i) gives the regression model whereas the model given by (ii) will arise, for example if $y_j - \alpha$ is the time required by a particle traveling at a constant velocity K to travel to the point x_j from an unknown origin β. Case (iii) is an example of exponential regression frequently used in the analysis of experiments with fertilizers to express the relationship with crop yield y_j and the amount x_j of fertilizer applied to the crop. Finally, (iv) is an example of dose response function usually used in biomedical research.

In this section we shall give properties of the estimator $\hat{\beta}_n$ of β in cases (i) and (ii). The problem of estimation with models (iii) and (iv) will not be taken up here since they do not present any new difficulty.

Case (i). Let $x_1 \geqq x_2 \geqq \cdots \geqq x_n$ and

$$g_j(\beta) = x_j\beta. \tag{6.1}$$

Then $g_{ij}(\beta) = \beta(x_i - x_j)$ is a never-decreasing function of β for all $i < j$, so that a robust point estimator for β is given by (2.2). In fact, as pointed out in Section 3, this estimator $\hat{\beta}_n$ is given by the median of the a_{ij} where $a_{ij} = (y_i - y_j)/(x_i - x_j)$, $i < j = 1, 2, \cdots, n$. In order to study the asymptotic distribution of $\hat{\beta}_n$ using Theorem 3, we have to ensure that our model satisfies condition II(a) of the theorem. One method of doing this is by introducing a spacing function to determine the values of the independent variables x_j as done by Hill [5]. Let p be a strictly increasing continuous function on [0, 1] and put

$$x_j = p(1 - j/n), \qquad j = 1, 2, \cdots, n. \tag{6.2}$$

The function p, called the spacing function, will determine the values of x_j for every n. In applications, p will be typically linear, say, $p(t) = A + Bt$.

With x_j defined according to (6.2) it is easy to see that

$$|g_{ij}(\beta + u) - g_{ij}(\beta)| = |u|(x_i - x_j) \leqq |u|[p(1) - p(0)],$$

for all $i < j$, so that condition II(a) of Theorem 3 is satisfied with $G(u) = |u|[p(1) - p(0)]$. Straightforward calculation will show that

$$\lim_{n\to\infty} 6n^{-3} \sum_{i<j} [g_{ij}(\beta + un^{-\frac{1}{2}}) - g_{ij}(\beta)] = 6u \int_0^1 (2t - 1)p(t)\, dt. \tag{6.3}$$

Therefore, by Theorem 3, $\hat{\beta}_n$ has an asymptotic normal distribution with mean β and variance $n^{-1}[6 \int_{-\infty}^{+\infty} f^2(x)\,dx \int_0^1 (2t-1)p(t)\,dt]^{-2}$. For the special case $p(t) = A + Bt$, the asymptotic variance of $\hat{\beta}_n$ reduces to $n^{-1}[B \int_{-\infty}^{+\infty} f^2(x)\,dx]^{-2}$.

In [1] Adichie proposed a class of estimators for β in the regression model and demonstrated asymptotic normality under somewhat more restrictive assumptions than those needed for Theorem 3 of this paper. An important member of Adichie's class of estimators is an estimator $\tilde{\beta}_n$ based on Wilcoxon's 2-sample statistic. If the spacing function p is linear, then by taking $\psi(u) = u$ in (5.1) of [1] it follows that the asymptotic variance of $\tilde{\beta}_n$ is same as that of $\hat{\beta}_n$. Thus, in this particular case, $\tilde{\beta}_n$ and $\hat{\beta}_n$ are asymptotically equivalent, even though $\hat{\beta}_n$ is computationally simpler. That $\hat{\beta}_n$ and $\tilde{\beta}_n$ turn out to be asymptotically equivalent for linear spacing functions is not surprising if one observes that while $h_n(\beta)$ is essentially the Kendall τ, the statistic T_2 (see (2.9) in [1]) on which $\tilde{\beta}_n$ is based is Spearman ρ, and that τ and ρ are asymptotically equivalent ([4], page 61).

CASE (ii). Let $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ and

$$g_j(\beta) = K|x_j - \beta|, \qquad j = 1, 2, \cdots, n,$$

where K is known. Without loss of generality, we may take $K = 1$, so that we have

$$g_j(\beta) = |x_j - \beta|, \qquad 1 \leqq j \leqq n. \tag{6.4}$$

Simple calculations will show that the g_j defined by (6.4) do satisfy the conditions required by the model (2.1), and that for all u,

$$|g_{ij}(\beta + u) - g_{ij}(u)| \leqq 2|u|,$$

so that condition II(a) of Theorem 3 is satisfied with $G(u) = 2|u|$.

Now, let $\lambda_n(x)$ denote the sample distribution function of $x_1, x_2, \cdots, x_n$. Then putting $n_1 = n\lambda_n(\beta)$, $n_2 = n[1 - \lambda_n(\beta + un^{-\frac{1}{2}})]$ and $n_0 = n - n_1 - n_2$, we can show, after some simple algebra, that

$$\begin{aligned}
&6n^{-\frac{3}{2}} \textstyle\sum_{i<j} [g_{ij}(\beta + un^{-\frac{1}{2}}) - g_{ij}(\beta)] \\
&\quad = 12n_1n_2n^{-2}u + 12n^{-\frac{3}{2}}[n_1 \textstyle\sum_{j=1}^{n_0} (x_{n_i+j} - \beta) + n_2 \sum_{j=1}^{n_0} (\beta + un^{-\frac{1}{2}} - x_{n_1+j}) \\
&\qquad + \textstyle\sum_{1 \leqq i<j<n_0} (x_{n_1+j} - x_{n_1+i})] \\
&\quad = 12\lambda_n(\beta)[1 - \lambda_n(\beta + un^{-\frac{1}{2}})]u + o(1).
\end{aligned}$$

Hence, if there exists a function λ such that $\lim_{n\to\infty} \lambda_n(\beta) = \lambda(\beta)$ and λ is continuous in some neighborhood of β, then (6.4) satisfies condition II(b) of Theorem 3 with $\Delta(\beta) = 12\lambda(\beta)[1 - \lambda(\beta)]$. Therefore, if λ with the specified properties exists, then $\hat{\beta}_n$ has asymptotically a normal distribution with mean β and variance $n^{-1}\{12\lambda(\beta)[1 - \lambda(\beta)] \int f^2(x)\,dx\}^{-2}$.

REFERENCES

[1] ADICHIE, J. N. (1967). Estimates of regression parameters based on rank tests. *Ann. Math. Statist.* **38** 894–904.

[2] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.
[3] FELLER, W. (1962). *An Introduction to Probability Theory and its Applications*. (2nd ed.) **1** Wiley, New York.
[4] HÁJEK, J., and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.
[5] HILL, B. M. (1962). A test of linearity versus convexity of a median regression curve. *Ann. Math. Statist.* **33** 1096–1123.
[6] Hodges, J. L. Jr., and Lehmann, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.
[7] Mood, A. M. (1950). *Introduction to the Theory of Statistics*. McGraw-Hill, New York.