The Annals of Statistics
1981, Vol. 9, No. 2, 438–442

BROWN-MOOD TYPE MEDIAN ESTIMATORS FOR SIMPLE REGRESSION MODELS

BY D. G. KILDEA

C.S.I.R.O., Melbourne

In the context of simple linear regression we investigate a class of estimators for the parameters which contains the Brown-Mood estimators. We derive the asymptotic efficiency properties of the class of slope estimators and find the most efficient slope estimator. This estimator is shown to have efficiency properties analogous to those of other median estimators for slope found in the literature. The equations which define the optimal estimators are median analogues of the least squares normal equation.

1. Introduction. The general method of Brown-Mood estimation was proposed by Brown and Mood (1951) and Mood (1950), as an intuitive analogue of the sample median, suitable for estimating parameters of linear regression models.

For the case of simple linear regression Hill (1962) has studied the theoretical properties of the estimators. Hill's results cast doubt on the status of Brown-Mood estimators as median analogues of least squares estimators—see Adichie (1967).

We will show here how the Brown-Mood estimator may be modified to produce a natural median analogue of the least squares estimator.

Section 2 introduces a class of modified Brown-Mood estimators and Section 3 establishes conditions for the class of slope estimators to be asymptotically normal. Section 4 contains the proof of a technical lemma needed in Section 3. Section 5 derives the most efficient member of the class of slope estimators and Section 6 discusses the case of regression through the origin.

2. Modified Brown-Mood estimators.

WEIGHTED MEDIAN AND WEIGHTED MEAN. If $X_1, \cdots X_n$ are random variables (rv's), $a_1, \cdots a_n$, are nonnegative constants, define a weighted empirical distribution (df) pointwise by $F_n(s) = [\sum_G a_i I(X_i \leq s)](\sum_G a_i)^{-1}$ where G is some subset of $\{1, 2, \cdots n\}$ for which $\sum_G a_i \neq 0$ and $I(\ )$ is the set indicator function. Define the following symbols:

$$\text{med.}(X_i, a_i, G) \quad \text{is the smallest median of } F_n(\ ). \tag{2.1}$$

$$\text{mean}(X_i, a_i, G) \quad \text{is} \int_{-\infty}^{\infty} s\, dF_n(s). \tag{2.2}$$

(When mentioning specific sets G we will occasionally, for example, abbreviate "$\{i: i \leq n\}$" to "$i \leq n$".)

Note that the symbol mean (X_i, a_i, G) is just a weighted mean, and, by analogy, we will call the symbol in (2.1) a weighted median. This terminology has been used previously by Jaeckel (1972).

We have chosen to define the weighted median as the smallest median of the underlying empirical df for convenience. All results still hold if any other median of the underlying weighted empirical df is defined as the weighted median.

Received May, 1978; revised April, 1979; December, 1979.

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

Key words and phrases. Brown-Mood estimator, median analogue, simple linear regression, regression through the origin, asymptotically optimal estimator.

Before introducing the modified Brown-Mood estimators we formally introduce the simple linear regression model. We assume we have n pairs of observations (y_i, c_i), $i = 1, 2, \cdots n$, with each pair satisfying $y_i = \alpha + \beta c_i + \varepsilon_i$, where $\{c_i\}$ are known constants, α and β are unknown parameters, and $\{\varepsilon_i\}$ are unobservable errors.

Now let c^* be the smallest median of the numbers $\{c_i; i = 1, 2, \cdots n\}$, and let $\bar{c}$ be their mean, then the Brown-Mood estimator (α', β') of (α, β) is defined by

$$\text{med.}(y_i - \alpha' - \beta' c_i, 1, c_i \leq c^*) = \text{med.}(y_i - \alpha' - \beta' c_i, 1, c_i > c^*) = 0. \tag{2.3}$$

We can rewrite (2.3) as

$$(2.4)\quad \begin{aligned} &\text{(A)} \quad \text{med.}(y_i - \alpha' - \beta' c_i, 1, i \leq n) = 0, \\ &\text{(B)} \quad \text{med.}(y_i - \beta' c_i, 1, c_i \leq c^*) = \text{med.}(y_i - \beta' c_i, 1, c_i > c^*). \end{aligned}$$

We can write the normal equations which define the least squares estimator $(\hat{\alpha}, \hat{\beta})$ in the same form using (2.2) to get

$$(2.5)\quad \begin{aligned} &\text{(A)} \quad \text{mean}(y_i - \hat{\alpha} - \hat{\beta} c_i, 1, i \leq n) = 0, \\ &\text{(B)} \quad \text{mean}(y_i - \hat{\beta} c_i, |c_i - \bar{c}|, c_i \leq \bar{c}) = \text{mean}(y_i - \hat{\beta} c_i, c_i - \bar{c}, c_i > \bar{c}). \end{aligned}$$

A comparison of (2.5) with (2.4) suggests that (2.4)(A) is a natural median analogue of (2.5)(A), but (2.4)(B) differs from (2.5)(B) not only in the mean and median operations but also with respect to weightings and groupings of residuals. Hence we allow the equation (2.4)(B) to be modified as follows.

Let S and U be two disjoint nonempty subsets of $\{1, 2, \cdots n\}$ such that

$$(i \in S \text{ and } j \in U) \Rightarrow c_i < c_j. \tag{2.6}$$

Let $\{s_i : i \in S\}$ and $\{u_i : i \in U\}$ be two sets of nonnegative numbers (weights) for which $\sum s_i > 0$, $\sum u_i > 0$. Then our class of modified Brown-Mood estimators is defined by

$$(2.7)\quad \begin{aligned} &\text{(A)} \quad \text{med.}(y_i - \tilde{\alpha} - \tilde{\beta} c_i, 1, i \leq n) = 0, \\ &\text{(B)} \quad \text{med.}(y_i - \tilde{\beta} c_i, s_i, S) = \text{med.}(y_i - \tilde{\beta} c_i, u_i, U). \end{aligned}$$

Since $g(t) = \text{med}(y_i - tc_i, s_i, S) - \text{med}(y_i - tc_i, u_i, U)$ is piecewise linear with derivative at least $\min_U c_i - \max_S c_i$, which exceeds 0 by (2.6), it follows that $(\tilde{\alpha}, \tilde{\beta})$ defined by (2.7) exists and is unique.

Calculation of $\tilde{\beta}$ (and hence $\tilde{\alpha}$), can be carried out exactly by evaluating $g(t)$ at the finite number of points $\{(y_j - y_i)(c_j - c_i)^{-1} : (i, j \in U) \text{ or } (i, j \in S)\}$. Although tedious this is programmable.

3. Asymptotic normality of the class of estimators $\tilde{\beta}$. We shall suppress the dependence on n of all quantities in this section except for $\tilde{\beta}$ which we shall write as $\tilde{\beta}_n$. We need the following technical lemma whose proof we defer until Section 4.

Let G and $\{a_i\}$ be as given in the definition of weighted median. Let m be the cardinality of G and $\{b_n\}$ a sequence of positive numbers. Assume that $m \to \infty$, as $n \to \infty$.

Consider the following conditions:

(3.1) $\{\varepsilon_i\}$ are i.i.d. with df $F(\ )$ and density $f(\ )$ continuous in a neighbourhood of 0. Further $F(0) = ½$ and $f(0) > 0$.

$$\limsup_{n\to\infty}[b_n^2(\textstyle\sum_G c_i^2)] < \infty, \tag{3.2}$$

$$\lim_{n\to\infty}(\max_G a_i^2)(\textstyle\sum_G a_i^2)^{-1} = \lim_{n\to\infty}(\max_G c_i^2)(\textstyle\sum_G c_i^2)^{-1} = 0. \tag{3.3}$$

Define:

$$
\begin{aligned}
&\tilde{\theta}_n(t, \mathbf{a}, G) = \text{med.}(\varepsilon_i - tc_i b_n, a_i, G),\\
(3.4)\qquad &\theta_n(t, \mathbf{a}, G) = -tb_n(\textstyle\sum_G a_i c_i)(\sum_G a_i)^{-1},\\
&\sigma_n^2(\mathbf{a}, G) = [4f^2(0)]^{-1}(\textstyle\sum_G a_i^2)(\sum_G a_i)^{-2}.
\end{aligned}
$$

LEMMA 3.1. *If conditions* (3.1) *to* (3.3) *hold, then*

$$(3.5)\qquad [\sigma_n(\mathbf{a}, G)]^{-1}[\tilde{\theta}_n(t, \mathbf{a}, G) - \theta_n(t, a, G)] \to_D N(0, 1) \qquad \textit{as } n \to \infty \textit{ for each fixed } t.$$

We can now state and prove our main theorem. Let the cardinalities of S and U both $\to \infty$ as $n \to \infty$. We need the following definition and condition:

$$(3.6)\qquad V_n^2 = [\sigma_n^2(\mathbf{s}, S) + \sigma_n^2(\mathbf{u}, U)][(\textstyle\sum_U u_i c_i)(\sum_U u_i)^{-1} - (\sum_S s_i c_i)(\sum_S s_i)^{-1}]^{-2},$$

$$(3.7)\qquad \limsup_{n\to\infty}[V_n^2(\textstyle\sum_{i=1}^n c_i^2)] < \infty.$$

THEOREM 3.1. *Given that* (3.1) *holds, that* (3.3) *holds for* $\{\{s_i : i \in S\}, \{c_i : i \in S\}\}$ *and* $\{\{u_i : i \in U\}, \{c_i : i \in U\}\}$ *and that we also have condition* (3.7), *then it follows that*

$$(3.8)\qquad V_n^{-1}(\tilde{\beta}_n - \beta) \to_D N(0, 1) \qquad \textit{as } n \to \infty.$$

PROOF OF THEOREM 3.1. We can assume without loss of generality that $\alpha = \beta = 0$. Recall $g(\ )$ defined below (2.7). From its monoticity we have

$$(3.9)\qquad \Pr(V_n^{-1}\tilde{\beta}_n > t) = \Pr(g(tV_n) < 0).$$

Hence, by Lemma 3.1 and (3.6), and using the independence of the two parts of $g(\ )$,

$$\Pr(V_n^{-1}\tilde{\beta}_n > t) = \Pr(M_n(t) < -t)$$

where $M_n(t) \to_D N(0, 1)$ as $n \to \infty$, for each t. Hence the proof is complete.

4. Proof of Lemma 3.1. We need the following notation

$$
\begin{aligned}
S_n(s; t) &= \textstyle\sum_G a_i I(\varepsilon_i - tc_i b_n \le s)\\
U_n(s; t) &= [S_n(s; t) - ES_n(s; t)]\mathrm{Var}^{-1/2} S_n(s; t).
\end{aligned}
$$

We write $\theta_n(t, \mathbf{a}, G)$ and $\sigma_n(\mathbf{a}, G)$ as simply θ_n and σ_n respectively in this section.

We have

$$
\begin{aligned}
(4.1)\qquad \Pr\{[\tilde{\theta}_n(t, \mathbf{a}, G) - \theta_n]\sigma_n^{-1} > s\} &= \Pr[S_n(s\sigma_n + \theta_n; t) < \tfrac{1}{2}\textstyle\sum_G a_i]\\
&= \Pr[U_n(s\sigma_n + \theta_n; t) < r_n],
\end{aligned}
$$

where

$$(4.2)\qquad r_n = [\tfrac{1}{2}\textstyle\sum_G a_i - ES_n(s\sigma_n + \theta_n; t)][\mathrm{Var}\, S_n(s\sigma_n + \theta_n; t)]^{-1/2}.$$

Now, by using (3.1) and (3.2) we obtain

$$(4.3)\qquad \tfrac{1}{2}\textstyle\sum_G a_i - ES_n(s\sigma_n + \theta_n; t) = -\tfrac{1}{2}s(\sum_G a_i^2)^{1/2} + o\big((\sum_G a_i^2)^{1/2}\big) \qquad \text{as } n \to \infty,$$

$$(4.4)\qquad \mathrm{Var}\, S_n(s\sigma_n + \theta_n; t) = \tfrac{1}{4}\textstyle\sum_G a_i^2 + o\big(\sum_G a_i^2\big) \qquad \text{as } n \to \infty.$$

Both $\lim_{n\to\infty}\theta_n = 0$ and $\lim_{n\to\infty}\sigma_n = 0$ follow from (3.2) and (3.3) and some simple arguments after it is established that $b_n \max|c_i| \to 0$ as $n \to \infty$). Hence $\lim_{n\to\infty} r_n = -s$. Also because $U_n(s; t)$ is a weighted sum of bounded, independent rv's, and because of (3.3) and (4.4), it follows that $U_n(s; t) \to_D N(0, 1)$ as $n \to \infty$. The result then follows immediately.

5. An optimal $\tilde{\beta}$. Our criterion for optimality in this section will be minimum asymptotic variance.

THEOREM 5.1. *Let the hypotheses of Theorem* 3.1 *hold. Then V_n^2 of* (3.6) *is minimized if*

$$s_i = |c_i - \bar{c}|, \quad S = \{i: c_i \le \bar{c}\} \quad \textit{and} \quad u_i = c_i - \bar{c}, \quad U = \{i: c_i > \bar{c}\} \tag{5.1}$$

the minimum value of V_n^2 being

$$[4f^2(0)]^{-1}[\textstyle\sum_{i=1}^n (c_i - \bar{c})^2]^{-1}. \tag{5.2}$$

PROOF OF THEOREM 5.1. V_n^2 is proportional to $(\sum_{j=1}^n L_j^2)(\sum_{j=1}^n L_j c_j)^{-2}$ where

$$\begin{aligned} L_j &= (\textstyle\sum_U u_i)s_j \qquad &\text{for } j \in S,\\ L_j &= -(\textstyle\sum_S s_i)u_j \qquad &\text{for } j \in U. \end{aligned} \tag{5.3}$$

Define two n dimensional vectors $\mathbf{L} = (L_1, \cdots, L_n)^T$ and $\mathbf{c} = (c_1, \cdots, c_n)^T$. Then (5.3) implies that any $\mathbf{L}$ must lie on the hyperplane in n dimensions defined by $\sum_{i=1}^n L_i = 0$. Hence minimizing V_n^2 is equivalent to finding a vector $\mathbf{L}$ on this hyperplane whose angle with the vector $\mathbf{c}$ is minimized. Hence the optimal vector must be proportional to the projection of $\mathbf{c}$ on the hyperplane. The result then follows because we must have all $\{s_i\}$ and $\{u_i\}$ nonnegative and the projection of c into the hyperplane is $(c_1 - \bar{c}, c_2 - \bar{c}, \cdots, c_n - \bar{c})^T$.

Substituting (5.1) in (2.7) we get the following estimating equations.

$$\begin{aligned} &\text{(A)} \qquad \text{med.}(y_i - \tilde{\alpha} - \tilde{\beta}c_i, 1, i \le n) = 0,\\ &\text{(B)} \qquad \text{med.}(y_i - \tilde{\beta}c_i, |c_i - \bar{c}|, c_i \le \bar{c}) = \text{med.}(y_i - \tilde{\beta}c_i, c_i - \bar{c}, c_i > \bar{c}). \end{aligned} \tag{5.4}$$

Thus (5.4) can be seen to be a natural median analogue of the least squares normal equations (2.5). Further the asymptotic minimum variance given in (5.2) is the same as that of other median analogues found in the literature (Adichie (1967), Bickel (1973) and Kraft and Van Eeden (1972)).

6. Regression through the origin. We shall briefly indicate the modifications required to estimate β when α is known a priori. This "regression through the origin" model was not discussed by Brown and Mood. Without loss of generality we will take α known to be 0.

We shall write "$\sum_{c_i<0}$" as "$\sum_1$", and "$\sum_{c_i>0}$" as "$\sum_2$". For this model the least squares normal equation becomes

$$(\textstyle\sum_1 |c_i|)\text{mean}(y_i - \beta c_i, |c_i|, c_i < 0) = (\textstyle\sum_2 c_i)\text{mean}(y_i - \beta c_i, c_i, c_i > 0). \tag{6.1}$$

This suggests that we should take a median estimating equation for β that is slightly more general than (2.7)(B), namely

$$K_s \text{med}(y_i - \tilde{\beta}c_i, s_i, S) = K_u \text{med}(y_i - \tilde{\beta}c_i, u_i, U), \tag{6.2}$$

where $(i \in S \Rightarrow c_i < 0)$ and $(i \in U \Rightarrow c_i > 0)$.

Theorem 3.1 generalizes immediately to this situation. For fixed sequences of weights $\{s_i\}$ and $\{u_i\}$, the values of K_s and K_u which give minimum asymptotic variance are

$$\begin{aligned} K_s &= (\textstyle\sum_S s_i)(\sum_S s_i c_i)(\sum_S s_i^2)^{-1},\\ K_u &= (\textstyle\sum_U u_i)(\sum_U u_i c_i)(\sum_U u_i^2)^{-1}. \end{aligned} \tag{6.3}$$

The minimum asymptotic variance of $\tilde{\beta}_n$ is

$$[4f^2(0)]^{-1}(\textstyle\sum_{i=1}^n c_i^2)^{-1}. \tag{6.4}$$

This is proportional to the least squares asymptotic variance with the same constant of proportionality as for simple linear regression. The optimal weights are

$$s_i = |c_i|, \quad S = \{i: c_i < 0\} \quad \text{and} \quad u_i = c_i, \quad U = \{i: c_i > 0\}. \tag{6.5}$$

The optimal estimating equation is once again a natural median analogue of (6.1):

$$(\textstyle\sum_1 |c_i|)\operatorname{med}(y_i - \tilde{\beta}c_i,\ c_i,\ c_i < 0) = (\textstyle\sum_2 c_i)\operatorname{med}(y_i - \tilde{\beta}c_i,\ c_i,\ c_i > 0). \tag{6.6}$$

The median estimator for this model which is given in Bickel (1973) is thus seen to be a one-step solution to (6.6)—see Bickel's equation (2.13).

ACKNOWLEDGMENT. This work forms part of the author's Ph.D. thesis submitted to LaTrobe University. Thanks are due to Dr. B. M. Brown and Dr. N. G. Becker for their encouragement, guidance and patience. Special thanks are due to an Associate Editor whose careful criticism of an earlier version has greatly improved the presentation of the results. He also located an error in Lemma 3.1.

REFERENCES

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

[2] BICKEL, P. J. (1973). On some analogues to linear combinations of order statistics in the linear model. *Ann. Statist.* **1** 597–616.

[3] BROWN, G. W. and MOOD, A. M. (1951). On median tests for linear hypotheses. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 159–166. Univ. of Calif.

[4] HILL, BRUCE MARVIN. (1962). A test of linearity versus convexity of a median regression curve. *Ann. Math. Statist.* **33** 1096–1123.

[5] JAECKEL, LOUIS A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43** 1449–1458.

[6] KRAFT, C. and VAN EEDEN, C. (1972). Linearized rank estimates and signed rank estimates for the general linear hypothesis. *Ann. Math. Statist.* **43** 42–57.

[7] MOOD, ALEXANDER MCFARLANE. (1950). *Introduction to the Theory of Statistics.* McGraw-Hill, New York.

C.S.I.R.O.
DIVISION OF MATHEMATICS AND STATISTICS
MELBOURNE, VICTORIA 3205
AUSTRALIA