The Annals of Statistics
1983, Vol. 11, No. 3, 921–932

MINIMUM DISTANCE ESTIMATION IN A LINEAR REGRESSION MODEL

BY H. KOUL[1] AND T. DEWET

Michigan State University and Rhodes University

This paper discusses a class of minimum distance Cramer-Von Mises type estimators of the slope parameter in a linear regression model. These estimators are obtained by minimizing an integral of squared difference between weighted empiricals of the residuals and their expectations with respect to a large class of integrating measures. The estimator corresponding to the weights proportional to the design variable is shown to be asymptotically efficient within the class at a given error distribution. The paper also discusses the asymptotic null distribution of a class of minimum Cramer-Von Mises type goodness-of-fit test statistics.

1. Introduction. Minimum distance (m.d.) estimation methods in the one sample problem have recently received considerable attention in the literature, e.g. see Beran (1977, 1978), Boos (1981), Parr and Schucany (1981), Parr and DeWet (1981), Millar (1981), among others. For more detailed references see the bibliography by Parr (1981). The most usual distance statistics used in the literature are the Cramer-Von Mises type statistics. One of the reasons for this is that the corresponding m.d. estimators are asymptotically normal. More recently, Millar (1981) has shown that these estimators are also asymptotically minimax and robust. In this paper we provide suitable analogues of the Cramer-Von Mises type m.d. estimators in a linear regression model.

Consider the linear regression model

$$Y_{ni} = x_{ni}\beta + \varepsilon_{ni}, \quad 1 \le i \le n, \tag{1.1}$$

where ε_{ni}, $1 \le i \le n$, are independent identically distributed (i.i.d.) random variables (r.v's) with known d.f. F, $x_{n1}, \cdots, x_{nn}$ are known constants and β is the parameter of interest. We are interested in seeking m.d. estimators of β, using Cramer-Von Mises type statistics, that will have properties similar to those of the one sample location parameter estimators. One natural thing to do is to construct the empirical d.f. based on the residuals $\{Y_{ni} - x_{ni}b, 1 \le i \le n\}$ and find b that minimizes the Cramer-Von Mises type statistics between this empirical d.f. and the error d.f. F. Theorem 3.2 below says that such estimators are not asymptotically as efficient as those obtained by using a certain weighted empirical process. To introduce this process, let $\mathbf{d}_n = (d_{n1}, \cdots, d_{nn})$ be a vector of real numbers and define a class of weighted empirical processes, one corresponding to each vector $\mathbf{d}_n$, by

$$V_d(y, b) = \sum d_{ni} I(Y_{ni} \le y + x_{ni}b), -\infty < y, b < +\infty. \tag{1.2}$$

Summations run from 1 to n throughout this paper. Note that $\{d_{ni}\}$ need not be non-negative.

The process that arises naturally in the model (1.1) is V_x, the V_d-process with $d_{ni} = x_{ni}(\sum x_{ni}^2)^{-1/2}$, $1 \le i \le n$. Observe that if F is continuous then the process $\{V_x(y, 0), -\infty < y < +\infty\}$ completely summarizes the data given in model (1.1) with probability one. The role played by this process is at least as important to the regression model as that of the process $\{V_1(y, 0), -\infty < y < +\infty\}$ in the one sample problem. Here $V_1 = V_d$ with $d_{ni} \equiv n^{-1/2}$. The V_x-process arises naturally in the least squares, M, L, and R estimators of β.

Received July 1981; revised December 1982.

[1] A revision of this paper was done while this author was visiting Poona University. His research, in part, was supported by NSF Grant No. MCS 8201291.

AMS 1980 *subject classification.* Primary 62F10; secondary 62J05, 62F03.

Key words and phrases. Weighted empiricals, goodness-of-fit tests, efficient estimators.

(See Koul, 1979, 1977, for a further discussion of this.) We are thus led naturally to the following definition of a class of estimators of β.

Observe that if β is true then $EV_d(y, \beta) = (\sum d_{ni})F(y)$. Therefore, define

$$T_d(b) = \int_{-\infty}^{\infty} \{V_d(y, b) - (\sum d_{ni})F(y)\}^2 \, dH(y) \tag{1.3}$$

where H is as in (3.4) below. Define $\hat{\beta}_d$ by the relation

$$\inf_b T_d(b) = T_d(\hat{\beta}_d). \tag{1.4}$$

Write $\hat{\beta}_1$ and $\hat{\beta}_x$ for $\hat{\beta}_d$ when $d_{ni} = n^{-1/2}$ and $d_{ni} = x_{ni}(\sum x_{ni}^2)^{-1/2}$, $1 \le i \le n$, respectively.

In this paper we study some finite and large sample properties of the class of estimators $\{\hat{\beta}_d\}$. We also study the asymptotic null distribution of a goodness-of-fit test statistic $T_x(\hat{\beta}_x)$ for testing hypotheses about the error distributions.

Section 2 contains some finite sample properties of $\{\hat{\beta}_d\}$. Section 3 discusses the asymptotic distribution of $\{\hat{\beta}_d\}$. In particular, Theorem 3.2 contains the result that $\hat{\beta}_x$ has the smallest asymptotic variance among those $\hat{\beta}_d$ for which $\mathbf{d}_n = (d_{n1}, \cdots, d_{nn})$ satisfies

$$\liminf (\sum x_{ni}^2)^{-1/2} |\sum d_{ni} x_{ni}| \ge \eta > 0, \quad \sum d_{ni}^2 = 1, \quad \max_{1\le i\le n} d_{ni}^2 \to 0$$
$$\text{and} \quad d_{ni}x_{ni} \le 0 \quad \text{or} \quad d_{ni}x_{ni} \ge 0 \quad \text{for} \quad 1 \le i \le n,$$

for a fairly large class of error distributions F and integrating measures H. The same section discusses asymptotic efficiency properties of $\hat{\beta}_x$ for various H. Remark 3.7 gives an extension of $\hat{\beta}_x$ to the multiple linear regression model with the known error d.f.

Admittedly model (1.1) is quite restrictive from the practical point of view because it assumes known intercept and completely known error distribution. One main reason for restricting attention to this model has been to reveal the importance of V_x-process in the m.d. estimation problem as clearly as possible. If, instead, one assumes the model

$$Y_{ni} = \alpha + \beta x_{ni} + \sigma e_i, \quad 1 \le i \le n, \tag{1.5}$$

where $\{e_i\}$ are i.i.d. with a known d.f. and (α, β, σ) are the parameters of interest, then one can also define m.d. estimators of (α, β, σ) using suitably modified V_1 and V_x processes. This is elaborated upon in Remark 3.8. The results one can obtain here extend those mentioned by Boos (1981) for the one sample scale-location model.

In Section 4 we propose and study the asymptotic null distribution of $T_x(\hat{\beta}_x)$ as a goodness-of-fit test for the error distribution of the model (1.1). The asymptotic null distribution of this statistic is the same as that of its analogue in the one sample location model. Remark 4.3 discusses goodness-of-fit tests for model (1.5). Section 5 contains most of the technical details for doing the asymptotic theory of Sections 3 and 4.

One of the members of $\{\hat{\beta}_d\}$ when $d_{ni} \propto x_{ni} - \bar{x}_n$, $1 \le i \le n$ and $H(y) \equiv y$ was studied by Williamson (1979). He showed that this estimator is asymptotically equivalent to the Wilcoxon type R estimator. Generally the class of estimators $\{\hat{\beta}_x(H)\}$, H satisfying conditions of Theorem 3.1 below, does not have any connection with M or R estimators of β for finite n. However, as is pointed out in the Remark 3.3, asymptotically this class is related to a certain class of M and R estimators but has no relation with the least squares estimators.

During the course of the writing of this paper the authors became aware of an alternative approach of m.d. estimation in regression by Millar (1982). He in fact deals with a much more general problem of independent, not identically distributed r.v.'s where each d.f. is indexed by a parameter. In the case of the regression model, his approach is different from the one being proposed here. For further discussion on this see Remark 3.5 below.

NOTATIONAL REMARK. In what follows, the index i in the maximum runs from 1 to n and all limits are taken as $n \to \infty$, unless mentioned otherwise. For any real numbers $\{d_{ni}\}$

let $\tau_d^2 = \sum d_{ni}^2$. By $o_p(1)$ ($O_p(1)$) we mean a sequence of r.v.'s that converges to zero (stays bounded) in probability. Often in a proof or a discussion we shall drop the suffix n from various underlying quantities for the sake of convenience. Thus d_i will stand for d_{ni} etc.

For any function g and h on $\mathscr{R} \times \mathscr{R}$ to $\mathscr{R}$, let $|g_s|_H^2$ denote $\int g^2(y, s)\, dH(y)$ and $|g_s - h_t|_H^2$ denote $\int \{g(y, s) - h(y, t)\}^2 dH(y)$ for any real numbers s and t.

2. Finite sample properties of $\hat{\beta}_d$. In order to discuss various properties of $\hat{\beta}_d$, we need to assume

(2.1a) $$d_i x_i \geq 0, \quad 1 \leq i \leq n,$$

or

(2.1b) $$d_i x_i \leq 0, \quad 1 \leq i \leq n,$$

and that F satisfies (3.6i) below. Let

(2.2) $$U_d(y, b) = V_d(y, b) - (\textstyle\sum d_i)F(y).$$

By the Cauchy-Schwarz inequality we have

(2.3) $$T_d(b) \geq \left\{\int U_d(y, b) f^{1/2}(y)\, dH(y)\right\}^2 \Big/ \left(\int f/dH\right) = L_d^2(b) \Big/ \left(\int f/dH\right), \quad \text{say.}$$

Note that under (2.1a)((2.1b))$L_d(b)$ is a nondecreasing (non-increasing) function of b. Therefore, by (2.3), T_d is bounded below by a nonnegative function which is nonincreasing on $(-\infty, b_0)$ and non-decreasing on $[b_0, \infty)$ for some finite b_0. This observation ensures that $\hat{\beta}_d$, though it may not be uniquely defined, can be uniquely defined as an average of the two quantities at which $T_d(b)$ is minimized for the first time and for the last time as b moves from the left to the right. The inequality (2.3) implies that these quantities are finite with probability 1.

Next, let $\hat{\beta}_d$ denote a minimizer of $T_d(b)$. Write $T_d(Y, bx)$ for $T_d(b)$ and observe that

(2.4) $$T_d(Y + ax, bx) = T_d(Y, (b - a)x), \quad -\infty < a, b < \infty.$$

Thus, if $\hat{\beta}_d(Y, x)$ denotes $\hat{\beta}_d$ of (1.4) based on $\{(x_i, Y_i), 1 \leq i \leq n\}$, then (2.4) implies that

(2.5) $$\hat{\beta}_d(Y + ax, x) = \hat{\beta}_d(Y, x) + a \quad \text{for all real } a.$$

Consequently, the distribution of $\hat{\beta}_d - \beta$ does not depend on β.

Another interesting property is that $\hat{\beta}_d(Y, ax) = a^{-1}\hat{\beta}_d(Y, x)$, for all $a \neq 0$. This means that the estimators $\{\hat{\beta}_d\}$ are invariant under the reparameterization of the design, a desirable property.

Next, we mention the symmetry property. If either (i) F and H are symmetric about 0 and H is continuous or (ii) $d_i = -d_{n-i+1}$, $x_i = -x_{n-i+1}$, $1 \leq i \leq n$, then $\hat{\beta}_d$ is symmetrically distributed about β. This follows by observing that under (i) or (ii), $T_d(-Y, bx) = T_d(Y, -bx)$ for every b and hence $\hat{\beta}_d(-Y, x) = -\hat{\beta}_d(Y, x)$. The details are similar to the proof of the symmetry of R estimators of Adichie (1967).

Finally, we would like to point out that if $d_i = n^{-1/2}$ then (2.1a) or (2.1b) is a restriction on the design variables $\{x_i\}$ whereas if $d_i = x_i(\sum x_i^2)^{-1/2}$, $1 \leq i \leq n$, then (2.1a) is a priori satisfied thereby giving $\hat{\beta}_x$ an added advantage.

3. Asymptotic distribution of $\hat{\beta}_d$. To begin with we state our assumptions as follows.

(3.1) $$\max_i x_{ni}^2 \tau_x^{-2} \to 0;$$

(3.2) $$\tau_d^2 = 1 \quad \text{and} \quad \max_i d_{ni}^2 \to 0;$$

(3.3) $$\liminf \tau_x^{-1} |\textstyle\sum d_i x_i| \geq \eta > 0 \quad \text{for some } \eta;$$

(3.4) H is a nondecreasing right continuous real valued function inducing a σ-finite measure on $(\mathcal{R}, \mathcal{B})$, the Borel line;

(3.5) F has a continuous density f with respect to (w.r.t.) the Lebesgue measure λ on $(\mathcal{R}, \mathcal{B})$;

(3.6) (i) $0 < \int f\,dH < \infty$, (ii) $\lim_{s\to 0} \int f(y+s)\,dH(y) = \int f\,dH$;

$$\int f^2\,dH < \infty. \tag{3.7}$$

$$\int F(1-F)\,dH < \infty; \tag{3.8}$$

(3.9) $\lim \int [\bar{J}_{nk}(y,\Delta) - \bar{J}_{nk}(y,0) - \Delta\xi_{nk}(y)]^2\,dH(y) = 0$ for $k = 1, 2$ and for every fixed real number Δ, where

$$\begin{aligned} \bar{J}_{n1}(y,\Delta) &= \sum d_{ni}F(y + \Delta x_{ni}\tau_x^{-1})I(d_{ni}x_{ni} \ge 0), \\ \bar{J}_{n2}(y,\Delta) &= \sum d_{ni}F(y + x_{ni}\tau_x^{-1})I(d_{ni}x_{ni} < 0), \\ \xi_{n1}(y) &= \tau_x^{-1}\sum d_{ni}x_{ni}I(d_{ni}x_{ni} \ge 0)f(y); \end{aligned} \tag{3.10}$$

and

$$\xi_{n2}(y) = \tau_x^{-1}\sum d_{ni}x_{ni}I(d_{ni}x_{ni} < 0)f(y), \quad -\infty < y, \Delta < \infty. \tag{3.11}$$

In what follows $\hat{\beta}_d$ is a solution of (1.4).

THEOREM 3.1. *Let $\{Y_{ni}, 1 \le i \le n\}$ be as in the model* (1.1). *Assume that $\{(x_{ni}, d_{ni}), 1 \le i \le n\}$, F and H satisfy* (2.1a) *or* (2.1b) *and* (3.1) *through* (3.9). *Then*

$$\tau_x(\beta_d - \beta) = -\left(\int f^2\,dH\right)^{-1}\int U_d(y,\beta)f(y)\,dH(y)\cdot r_{xd}^{-1} + o_p(1) \tag{3.12}$$

where

$$r_{xd} = \tau_x^{-1}\sum d_{ni}x_{ni}. \tag{3.13}$$

PROOF. The proof is given in Section 5. □

Using the Lindeberg-Feller central limit theorem, one concludes the following:

COROLLARY 3.1. *Under the assumptions of Theorem* 3.1 *the asymptotic distribution of $\tau_x(\hat{\beta}_d - \beta)$ is Normal with mean* 0 *and variance*

$$v_{xd}(F,H) = r_{xd}^{-2}\left(\int f^2\,dH\right)^{-2}\cdot K(F,H) \tag{3.14}$$

where

$$K(F,H) = \int\int \{F(x\wedge y) - F(x)F(y)\}f(x)f(y)\,dH(x)\,dH(y). \;\square \tag{3.15}$$

Now, the Cauchy-Schwarz inequality yields $r_{xd}^{-2} \ge 1$ with equality if, and only if, $d_i = x_i\tau_x^{-1}$, $1 \le i \le n$. Thus we have the following:

THEOREM 3.2. *Among all estimators $\{\hat{\beta}_d\}$ where $\mathbf{d}_n = (d_{n1}, \cdots, d_{nn})$ satisfy* (2.1a) *or* (2.1b), (3.2), (3.3) *and* (3.9) *for every F and H satisfying* (3.4)–(3.8), *the one that minimizes the asymptotic variance v_{xd} is $\hat{\beta}_x$ — the $\hat{\beta}_d$ when $d_i = x_i\tau_x^{-1}$, $1 \le i \le n$.*

REMARKS 3.1. A consequence of Theorem 3.2 is that $\hat{\beta}_1$—the $\hat{\beta}_d$ with $d_i \equiv n^{-1/2}$—is less efficient than $\hat{\beta}_x$. In view of (2.1a) or (2.1b) this comparison is valid only if all $x_i \geq 0$ or all $x_i \leq 0$.

Another important and desirable property of $\hat{\beta}_x$ is that the asymptotic distribution of $\tau_x(\hat{\beta}_x - \beta)$ is the same as that of a standardized m.d. Cramer-Von Mises type estimators of location parameter (see e.g. Parr and DeWet, 1981). The results obtained there concerning the choice of H for optimality or robustness also apply then to the present situation. Thus, e.g., in order to have an asymptotically efficient estimator $\hat{\beta}_x$ or β, H is given by the relation

$$(3.16) \qquad f(y)\, dH(y) = -I^{-1}\, d(f'(y)/f(y)), \quad I = I(f) = \int (f'/f)^2\, dF < \infty.$$

In many interesting cases this H gives infinite mass to the real line. For example if F is logistic then $dH(y) = (2/3)dy$ and in this case $\tau_x|\hat{\beta}_x - \hat{\beta}_W| = o_p(1)$ as may be seen from the results of Theorem 3.1. Here $\hat{\beta}_W$ is the Wilcoxon rank estimator of β. Note that here H induces a σ-finite measure.

Often when m.d. Cramer-Von Mises type estimators are driven to be asymptotically efficient the optimal H turns out to be a σ-finite measure. This phenomenon seems to be in contrast with the robustness of Millar (1981) where finite H are preferred.

REMARK 3.2. An important H that is covered by the above theory is $dH = \{F(1 - F)\}^{-1}dF$, the so called Anderson-Darling (1952) weights. From the property of the Logistic distribution (viz. $f = F(1 - F)$) and from (3.12) one again has $\tau_x|\hat{\beta}_x - \hat{\beta}_W| = o_p(1)$ at Logistic F. In other words the Anderson-Darling type estimator is also asymptotically efficient at the Logistic errors.

REMARK 3.3. *Connection with other estimators.* If we define $\psi(y) = \int_{-\infty}^{y} f\, dH$ then from (3.12) one again has $|\tau_x(\hat{\beta}_x - \hat{\beta}(\psi))| = o_p(1)$ where $\hat{\beta}(\psi)$ is the M estimator (Huber, 1973) corresponding to the score function ψ. Now, it is well known that if $\psi(y) = y$ then $\hat{\beta}(\psi)$ is the least squares estimator. Thus in order for $\hat{\beta}_x$ to be approximately equivalent to the least squares estimator, H would have to be such that $\int_{-\infty}^{y} f\, dH = y$, but this would imply $\int_{-\infty}^{\infty} f/dH = \infty$, violating (3.6i), rather a crucial condition for our theory to hold. Thus $\hat{\beta}_x$ has no connection with the least squares under the conditions of this paper.

Next, let $\hat{\beta}(\varphi)$ denote Adichie's (op. cit.) rank estimator corresponding to the score function φ. If we choose $\varphi(u) = \int_{-\infty}^{F^{-1}(u)} f(x)\, dH(x)$ then again $\tau_x|\hat{\beta}_x - \hat{\beta}(\varphi)| = o_p(1)$ as follows from (3.12) and the asymptotic properties of R estimators.

REMARK 3.4. When $H = F$, the corresponding $\hat{\beta}_x$ has high asymptotic efficiency relative to some of the well known estimators of β. For example the asymptotic variances of $\tau_x\hat{\beta}_x$ at the Double exponential, Logistic and $N(0, 1)$ distribution are 1.2, 3.0357 and 1.0942, respectively. Compare these with those of the Wilcoxon type estimator which are 1.333, 3 and 1.0472, respectively. For comparison with some other estimators see Koul (1979) or Williamson (1979, 1982).

REMARK 3.5. If Millar (1982) is specialized to the above model (1.1) then his estimator $\tilde{\beta}$ is essentially obtained by minimizing

$$\int \int_0^1 [n^{-1/2} \textstyle\sum_{i=1}^{ns} \{I(Y_i \leq y) - F(y - x_{ni}b)\}]^2\, ds\, dH(y)$$

w.r.t. b. Clearly this is different from $\hat{\beta}_x$. Moreover, if $x_{ni} \equiv 1$, (the one sample location model), then

$$\text{asymptotic var.}(n^{1/2}\tilde{\beta}) = (18/15) \text{ asymptotic var.}(n^{1/2}\hat{\beta}_1).$$

If $x_{ni} = i$ (the first order polynomial), then the asymptotic var. $(\tau_x \tilde{\beta}) = (400/168)$ asymptotic var.$(\tau_x \hat{\beta}_x)$. These results are valid for $\beta = 0$ and for all F and H satisfying the above conditions.

Thus, even though Millar considers a much more general problem, in the above special cases the procedure proposed here possesses some superiority.

REMARK 3.6. We now discuss conditions (3.1)–(3.9). All the conditions except for (3.9), are readily verifiable. It is desirable to have readily verifiable sufficient conditions for (3.9). Consider the following conditions.

(3.17) F has uniformly continuous bounded density f.

(3.18) (i) $d_{ni} = \tau_x^{-1} x_{ni},\ 1 \le i \le n$, (ii) $\lim_{s \to 0} \int f^2(y+s)\, dH(y) = \int f^2\, dH.$

Then the following two statements hold.

(3.19) If (3.1), (3.2), (3.4), (3.6), (3.8) and (3.17) hold then (3.9) holds.

(3.20) If (3.1), (3.4)–(3.8) and (3.18) hold then (3.9) holds.

Proofs of these statements use Fubini's theorem and the usual uniform integrability techniques. Details are left out for the sake of brevity.

Note that if H is absolutely continuous w.r.t. λ then (3.6i) and (3.7) imply (3.6ii) and (3.18ii).

REMARK 3.7. *Extension to multiple linear regression.* Suppose

$$Y_i = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i, \quad 1 \le i \le n$$

where $\mathbf{x}_i$ is a $1 \times p$ vector, the ith row of the design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector and $\{\varepsilon_i\}$ are i.i.d. with known d.f. F. An extension of $\hat{\beta}_x$ to the multiple regression model is as follows. Define, for a $p \times 1$ vector $\mathbf{b}$ and a real number y,

$$U_j(y, \mathbf{b}) = \tau_j^{-1} \textstyle\sum_{i=1}^n x_{ij} \{I(Y_i \le y + \mathbf{x}_i \boldsymbol{b}) - F(y)\}, \quad 1 \le j \le p$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{ip})$ and $\tau_j^2 = \sum_{i=1}^n x_{ij}^2$. Let $\mathbf{U}' = (U_1, \cdots, U_p)$ and form

$$T(\mathbf{b}) = \int \mathbf{U}'(y, \mathbf{b}) \mathbf{U}(y, \mathbf{b})\, dH(y).$$

Then $\hat{\boldsymbol{\beta}}$ is defined as a minimizer of $T(\mathbf{b})$. This is one right extension of $\hat{\beta}_x$. Asymptotic theory of this estimator is somewhat involved and will be reported elsewhere. Results analogous to Theorems 3.1 and 3.2 are expected to hold here also.

REMARK 3.8. *Simple linear regression with unknown scale.* Here we will give m.d. estimators of (α, β, σ) of the model (1.5) and mention as to what kind of results can be obtained. Define

$$V_1(y, a, b, s) = n^{-1/2} \textstyle\sum I(Y_i \le ys + a + bx_i)$$

$$V_x(y, a, b, s) = \tau_x^{-1} \textstyle\sum x_i I(Y_i \le ys + a + bx_i), \quad y, a, b \text{ real}, s \ge 0.$$

Let F now stand for the d.f. of e_i and define

$$T(a, b, s) = \int [\{V_1(y, a, b, s) - n^{1/2} F(y)\}^2 + \{V_x(y, a, b, s) - \tau_x^{-1} \textstyle\sum x_i F(y)\}^2]\, dH(y).$$

One way to define m.d. estimators of (α, β, σ) is by the relation

(3.21) $$\inf_{a, b \text{ real}, s > 0} T(a, b, s) = T(\hat{\alpha}, \hat{\beta}, \hat{\sigma}).$$

Using the techniques of this paper, one can show, under the conditions similar to those given above and under some additional mild conditions (involving $\int yf\,dH$, $\int yf^2\,dH$ and $\int (yf)^2\,dH$), that

$$(3.22)\qquad \begin{aligned}&(n^{1/2}(\hat\alpha-\alpha),\ \tau_x(\hat\beta-\beta),\ n^{1/2}(\hat\sigma-\sigma))\\ &\quad= -B^{-1}\Big(\int (W_1+a_nW_2)f\,dH,\ \int (a_nW_1+W_2)f\,dH,\\ &\qquad\qquad \int \{W_1(y)+a_nW_2(y)\}yf(y)\,dH(y)\Big)+o_p(1).\end{aligned}$$

where

$$a_n = n^{-1/2}\tau_x^{-1}\sum x_i,\quad W_1(y)=n^{-1/2}\sum\{I(e_i\le y)-F(y)\}$$
$$W_2(y)=\tau_x^{-1}\sum x_i\{I(e_i\le y)-F(y)\}$$

and where

$$\sigma B=\begin{bmatrix} t_n\int yf^2(y)\,dH(y) & 2a_n\int yf^2(y)\,dH(y) & t_n\int\{yf(y)\}^2\,dH(y)\\ t_n\int f^2\,dH & 2a_n\int f^2\,dH & t_n\int yf^2(y)\,dH(y)\\ 2a_n\int f^2\,dH & t_n\int f^2\,dH & 2a_n\int yf^2(y)\,dH(y)\end{bmatrix}$$

with $t_n = 1 + a_n^2$.

The joint asymptotic normality can be readily deduced from (3.22). Note that $a_n = O(1)$. Note also that if we specialize (1.5) to the one sample location-scale model by taking $x_i = 0$, then (3.22) reduces to (4.3) of Boos (1981).

4. Goodness-of-fit test for the error distribution. The m.d. Cramer-Von Mises type statistics are well known as goodness-of-fit statistics (see, e.g., Durbin, 1973) in the one sample problem. Relatively little is known about their analogues suitable in the regression model. Consider the model (1.1) with error distribution F. We are interested in testing

$$H_0: F = F_0$$

with F_0 a known continuous d.f.

In order to describe the proposed test, write $T_d(b, F)$ for $T_d(b)$ of (1.3). Then the proposed test rejects H_0 for the large values of $T_x(\hat\beta_x, F_0)$. The asymptotic null distribution of this statistic is deduced from the following.

THEOREM 4.1. *Let $\{Y_{ni}, 1\le i\le n\}$ be as in model* (1.1) *with* $\{\varepsilon_{ni}, 1\le i\le n\}$ i.i.d. F_0. *Assume $\{x_{ni}\}$ satisfy* (3.1), *that* (3.4)–(3.8) *and* (3.18ii) *are satisfied by F_0 and H. Then*

$$(4.1)\qquad T_x(\hat\beta_x, F_0) = \int \{U_x(y,\beta)+\tau_x(\hat\beta_x-\beta)f_0(y)\}^2\,dH(y)+o_p(1).$$

PROOF. Without loss of generality assume that the true $\beta = 0$. Then apply Theorem 5.1 with $d_i = x_i\tau_x^{-1}$, $1\le i\le n$, $F = F_0$, $\Delta = \tau_x\hat\beta_x$ to conclude (4.1). Note that we also need the conclusion of Corollary 3.1 which says that $\tau_x|\hat\beta_x| = O_p(1)$. Recall from Remark 3.6 that (3.9) is implied by the conditions of this theorem. □

From now we shall assume that the true $\beta = 0$ and we shall write $U_x(\cdot)$ for $U_x(\cdot, 0)$.

Now observe that

$$\int U_x f_0 \, dH = -\tau_x^{-1} \sum x_i \{\psi_0(Y_i) - E\psi_0(Y_i)\},$$

where $\psi_0(y) = \int_{-\infty}^{y} f_0 \, dH$. Using the Lindeberg-Feller central limit theorem, one readily concludes, in view of (3.1) and (3.6i) with $f = f_0$, that

$$\int U_x f_0 \, dH \to_D \int (B \circ F_0) f_0 \, dH \tag{4.2}$$

where B is the Brownian Bridge. From this one has $\int U_x f_0 \, dH = O_p(1)$. Using this, (3.12), (4.1), (3.6i) and (3.7) with $f = f_0$, and expanding the quadratic one has

$$T_x(\hat{\beta}_x, F_0) = \int U_x^2 \, dH - \left(\int f_0^2 \, dH\right)^{-1} \left(\int U_x f_0 \, dH\right)^2 + o_p(1). \tag{4.3}$$

Now note that, using (3.8)

$$E \int U_x^2 \, dH = \int F_0(1 - F_0) \, dH = E \int (B \circ F_0)^2 \, dH < \infty. \tag{4.4}$$

This together with an argument given in the proof of Proposition 4.1 of Millar (1981) yields that

$$\int U_x^2 \, dH \to_D \int (B \circ F_0)^2 \, dH. \tag{4.5}$$

Therefore, (4.5), (4.4), (4.3) and (4.2) together yield:

COROLLARY 4.1. *Under the assumptions of Theorem* 4.1 *and under* H_0

$$\begin{aligned} T_x(\hat{\beta}_x, F_0) \to_D \int (B \circ F_0)^2 \, dH \\ - \left(\int f_0^2 \, dH\right)^{-1} \left\{\int (B \circ F_0) f_0 \, dH\right\}^2 = G(B \circ F_0), \quad \textit{say}. \end{aligned} \tag{4.6}$$

REMARK 4.1. The first term in the limiting r.v. $G(B \circ F_0)$ is the limiting r.v. of the test statistic had we known β and the second term comes from estimating β. Note also that the regression constants $\{x_i\}$ do not appear in the limiting r.v. $G(B \circ F_0)$. As a matter of fact, $G(B \circ F_0)$ is the same as the limiting r.v. obtained in the one sample location model. The distribution of this r.v. is available, e.g., see Boos (1981) or Martinov (1975).

For similar conclusions pertaining to statistics $\sup_y | V_d(y, \beta) - \sum d_i F_0(y) |$, see Koul (1980) for the cases $d_i = n^{-1/2}$ and $d_i = \tau_x^{-1} x_i$, $1 \le i \le n$. See also Pierce and Kopecky (1979) regarding the process $\{U_1(y, \beta) - \infty < y < +\infty\}$.

REMARK 4.2. Corollary 4.1 holds for H given by $dH = \{F_0(1 - F_0)\}^{-1} \, dF_0$ at F_0 equal to logistic, normal and double exponential and many others. Other examples of weight functions can be found in DeWet and Venter (1973).

REMARK 4.3. Consider the model (1.5) and the above H_0 where now f_0 is the d.f. of e_i. An analogue of $T_x(\hat{\beta}_x, F_0)$ in this problem is given by $T(\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ of (3.21) with F replaced by F_0. Using the methods of this paper one can show, under the conditions of Theorem 4.1 and under some additional conditions (see Remark 3.8), that under H_0

$$\begin{aligned} T(\hat{\alpha}, \hat{\beta}, \hat{\sigma}) = \int [W_1(y) + \{q + \Delta a_n + uy\} f_0(y) \sigma^{-1}]^2 \, dH(y) \\ + \int [W_2(y) + \{a_n(q + uy) + \Delta\} f_0(y) \sigma^{-1}]^2 \, dH(y) + o_p(1) \end{aligned}$$

where $q = n^{1/2}(\hat{\sigma} - \alpha)$, $\Delta = \tau_x(\hat{\beta} - \beta)$, $u = n^{1/2}(\hat{\sigma} - \sigma)$ with $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}$ as defined in (3.21). Using the approximation (3.22) it is possible to arrive at an analogue of Corollary 4.1.

5. Asymptotic quadraticity of T_d and some proofs. In this section we prove the asymptotic quadraticity of T_d in b. Corollary 5.1 is useful in concluding that $\tau_x(\hat{\beta}_d - \beta) = O_p(1)$. The section is concluded with the proof of Theorem 3.1. We begin with the statement and the proof of:

THEOREM 5.1. *Let $Y_{n1}, \cdots, Y_{nn}$ be* i.i.d. *F. Assume that* (3.1), (3.2), *and* (3.4)–(3.19) *are satisfied by $\{(x_{ni}, d_{ni}), 1 \le i \le n\}$, F and H. Then, for any $0 < B < \infty$,*

$$E\{\sup_{|\Delta|\le B} | T_d(\Delta\tau_x^{-1}) - \hat{T}_d(\Delta\tau_x^{-1}) |\} \to 0 \tag{5.1}$$

where

$$\hat{T}_d(b) = \int \{U_d(y, 0) + \tau_x b \cdot r_{xd} f(y)\}^2 \, dH(y) \tag{5.2}$$

with r_{xd} as in (3.13).

PROOF. Recall the definitions of $\bar{J}_k$, ξ_k, $k = 1, 2$ from (3.10) and (3.11). Let

$$\bar{J} = \bar{J}_1 + \bar{J}_2, \quad \xi = \xi_1 + \xi_2. \tag{5.3}$$

Define

$$W(y, \Delta) = \sum d_i\{I(Y_i \le y + \Delta c_i) - F(y + \Delta c_i)\}, \quad -\infty < y, \quad \Delta < \infty, \tag{5.4}$$

with $c_i = x_i\tau_x^{-1}$, $1 \le i \le n$. Note that

$$(\sum c_i d_i)^2 \le 1. \tag{5.5}$$

Observe that

$$T_d(\Delta\tau_x^{-1}) = \int [\{W(y, \Delta) - W(y, 0)\} + \{W(y, 0) + \Delta\xi(y)\} + \{\bar{J}(y, \Delta) - \bar{J}(y, 0) - \Delta\xi(y)\}]^2 \, dH(y).$$

Expanding the quadratic and using the Cauchy-Schwarz inequality on the product terms yield (see Introduction for notation)

$$\begin{aligned} |M(\Delta) - \hat{M}(\Delta)| \le{}& | W_\Delta - W_0|_H^2 + |\bar{J}_\Delta - \bar{J}_0 - \Delta\xi|_H^2 \\ &+ 2| W_0 + \Delta\xi|_H |\bar{J}_\Delta - \bar{J}_0 - \Delta\xi|_H \\ &+ 2| W_\Delta - W_0|_H(| W_0 + \Delta\xi|_H + |\bar{J}_\Delta - \bar{J}_0 - \Delta\xi|_H), \end{aligned} \tag{5.6}$$

where $M(\Delta) = T_d(\Delta\tau_x^{-1})$, $\hat{M}(\Delta) = \hat{T}_d(\Delta\tau_x^{-1})$.

From (5.6), (5.1) will follow if we show (i) $\lim\sup E\{\sup| W_0 + \Delta\xi|_H^2\} < \infty$, (ii) $\sup|\bar{J}_\Delta - \bar{J}_0 - \Delta\xi|_H^2 \to 0$, and (iii) $E\{\sup| W_\Delta - W_0|_H^2\} \to 0$. Here and elsewhere the supremum is being taken over Δ in $[-B, B]$, unless mentioned otherwise.

PROOF OF (i). Using $(a + b)^2 \le 2a^2 + 2b^2$, Fubini and $\tau_d^2 = 1$, one gets

$$E\{\sup| W_0 + \Delta\xi|_H^2\} \le 2\int F(1 - F)\, dH + 2B^2(\sum d_i c_i)^2 \int f^2 \, dH.$$

Therefore (i) follows from (5.5), (3.7) and (3.8).

PROOF OF (ii). Note that $\bar{J}_1(\bar{J}_2)$ is a nondecreasing (non-increasing) function of Δ for each fixed y. Let $-B = \Delta_0 < \cdots < \Delta_r = B$ be a decomposition of $[-B, B]$ such that

$$\max_{1\le j\le r}(\Delta_j - \Delta_{j-1})^2 \to 0 \quad \text{as } r \to \infty. \tag{5.7}$$

Using the monotonicity of $\bar{J}_k$, $k = 1, 2$, and the elementary inequality $a \le b \le c \Rightarrow b^2 \le a^2 + c^2$ and the inequality $(a + b)^2 \le 2a^2 + 2b^2$ one gets that for $\Delta_{j-1} \le \Delta \le \Delta_j$

$$(5.8)\quad \begin{aligned} |\bar{J}_{k\Delta} - \bar{J}_{k0} - \Delta\xi_k|_H^2 &\le 2(|\bar{J}_{k\Delta_j} - \bar{J}_{k0} - \Delta_j\xi_k|_H^2 + |\bar{J}_{k\Delta_{j-1}} - \bar{J}_{k0} - \Delta_{j-1}\xi_k|_H^2) \\ &\quad + 4(\Delta_j - \Delta_{j-1})^2|\xi_k|_H^2, \quad k = 1, 2. \end{aligned}$$

Therefore (5.3) yields

$$(5.9)\quad \begin{aligned} \sup|\bar{J}_\Delta - \bar{J}_0 - \Delta\xi|_H^2 &\le 4\textstyle\sum_{j=1}^r (|\bar{J}_{1\Delta_j} - \bar{J}_{10} - \Delta_j\xi_1|_H^2 + |\bar{J}_{2\Delta_j} - \bar{J}_{20} - \Delta_j\xi_2|_H^2) \\ &\quad + 8\max_{1\le j\le r}(\Delta_j - \Delta_{j-1})^2(|\xi_1|_H^2 + |\xi_2|_H^2). \end{aligned}$$

Hence (ii) follows from (5.5), (3.9), (3.7), (5.7) and (5.9) by letting first $n \to \infty$ and then $r \to \infty$ in (5.9).

PROOF OF (iii). Write $W = W_1 + W_2$ where $W_1(W_2)$ consists of those summands in W for which $x_i/d_i \ge 0$ $(x_i/d_i < 0)$. Direct calculations and Fubini yield

$$\begin{aligned} E|W_{k\Delta} - W_{k0}|_H^2 &\le \sum d_i^2 \int |F(y + \Delta c_i) - F(y)|\, dH(y) \\ &\le \int_{-Bm}^{Bm} \left\{\int f(y + s)\, dH(y)\right\} ds, \qquad \text{for} \quad k = 1, 2. \end{aligned}$$

Here $m = \max|c_i|$ and we used $\tau_d^2 = 1$. Thus for every fixed $|\Delta| \le B$, by (3.6ii),

$$(5.10)\qquad E|W_{k\Delta} - W_{k0}|_H^2 \to 0, \qquad k = 1, 2.$$

Next, exploit the monotonic structure that is inherent in these processes to get, just like (5.9),

$$(5.11)\quad \begin{aligned} \sup|W_\Delta - W_0|_H^2 &\le 4\textstyle\sum_{j=1}^r (|W_{1\Delta_j} - W_{10}|_H^2 + |W_{2\Delta_j} - W_{20}|_H^2) \\ &\quad + 8\max_{1\le j\le r}(|\bar{J}_{1\Delta_j} - \bar{J}_{1\Delta_{j-1}}|_H^2 + |\bar{J}_{2\Delta_j} - \bar{J}_{2\Delta_{j-1}}|_H^2). \end{aligned}$$

By (5.5), (3.9), (5.10) and (5.11) for every fixed r,

$$(5.12)\qquad \limsup E(\sup|W_\Delta - W_0|_H^2) \le 32\max_{1\le j\le r}(\Delta_j - \Delta_{j-1})^2 \int f^2\, dH.$$

Therefore (iii) follows by letting $r \to \infty$ in (5.12). This also completes the proof of the Theorem. □

COROLLARY 5.1. *In addition to the conditions of Theorem* 5.1 *assume that* (2.1a) *or* (2.1b) *and* (3.3) *hold. Then for any* $\varepsilon > 0$, $0 < M < \infty$ *there exists an* N_ε *and* $0 < g < \infty$ *(depending on* ε *and* M*) such that*

$$(5.13)\qquad P\{\inf_{|\Delta|>g} T_d(\Delta\tau_x^{-1}) \ge M\} \ge 1 - \varepsilon \quad \textit{for} \quad n \ge N_\varepsilon.$$

PROOF. From the inequality (2.3),

$$(5.14)\qquad P\{\inf_{|\Delta|>g} T_d(\Delta\tau_x^{-1}) \ge M\} \ge P\{\inf_{|\Delta|>g} L_d^2(\Delta\tau_x^{-1}) \ge Mq\},$$

where $q = \int f\, dH$. Now, define

$$(5.15)\qquad \hat{L}_d(b) = \int \{U_d(y, 0) + \tau_x b \cdot r_{xd} f(y)\} f^{1/2}(y)\, dH(y).$$

Observe that

$$|L_d(\Delta\tau_x^{-1}) - \hat{L}_d(\Delta\tau_x^{-1})| \le q^{1/2}(|W_\Delta - W_0|_H + |\bar{J}_\Delta - \bar{J}_0 - \Delta r_{xd} f|_H). \tag{5.16}$$

From (5.16), (ii) and (iii) contained in the proof of Theorem 5.1 we have that for any $0 < B < \infty$,

$$E\{\sup_{|\Delta|\le B}|L_d(\Delta\tau_x^{-1}) - \hat{L}_d(\Delta\tau_x^{-1})|\} \to 0. \tag{5.17}$$

Next, let $S = \int U_d(y, 0)f^{1/2}(y)dH(y)$, $v = \int f^{3/2}\, dH$. Note that

$$ES = 0;\ ES^2 \le q \int F(1-F)\, dH.$$

Thus, for any $\varepsilon > 0$ there exists $N_{1\varepsilon}$ and K_ε such that

$$P(|S| \le K_\varepsilon) \ge 1 - \varepsilon/2, \qquad n \ge N_{1\varepsilon}. \tag{5.18}$$

Let g satisfy

$$g > \{K_\varepsilon + (Mq)^{1/2}\}(\eta v)^{-1} \quad \text{where} \quad \eta \text{ is as in (3.3).} \tag{5.19}$$

Then we have the following inequalities.

$$\begin{aligned} P\{\inf_{|\Delta|=g}\hat{L}_d^2(\Delta\tau_x^{-1}) \ge Mq\} &\ge P\{|S| \le -(Mq)^{1/2} + g|r_{xd}|v\} \\ &\ge P(|S| \le K_\varepsilon) \ge 1 - \varepsilon/2, \qquad n > N_{1\varepsilon}. \end{aligned} \tag{5.20}$$

By (5.17), for every $\varepsilon > 0$ there exists $N_{2\varepsilon}$ such that for all $n \ge N_{2\varepsilon}$

$$P\{\inf_{|\Delta|=g}L_d^2(\Delta\tau_x^{-1}) \ge Mq\} \ge P\{\inf_{|\Delta|=g}\hat{L}_d^2(\Delta\tau_x^{-1}) \ge Mq\} - \varepsilon/2. \tag{5.21}$$

Thus, choose $N_\varepsilon = \max(N_{1\varepsilon}, N_{2\varepsilon})$ and use the monotoneity of L_d in Δ together with (5.21) and (5.20) to conclude (5.13) for g given by (5.19). □

PROOF OF THEOREM 3.1. Because of (2.5), without loss of generality, we will assume $\beta = 0$. Then $Y_1, \cdots, Y_n$ are i.i.d. F and the above results are applicable. From (5.13) one concludes that $\tau_x|\hat{\beta}_d| = O_p(1)$. Details are similar to those in Millar (1981, 1982) or Williamson (1982). This together with (5.1) and the quadratic nature of $\hat{T}_d$ of (5.2) yields (3.12). □

REFERENCES

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

ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

BERAN, R. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BERAN, RUDOLF (1978). An efficient and robust adaptive estimator of location. *Ann. Statist.* **6** 292–313.

BOOS, DENNIS, D. (1981). Minimum distance estimators for location and goodness of fit. *J. Amer. Statist. Assoc.* **76** 663–670.

DURBIN, J. (1973). Distribution theory for tests based on the sample distribution functions. CBM-NSF Regional Conference series in Applied Math. No. 9. SIAM.

HUBER, P. J. (1973). Robust regression; asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

KOUL, H. L. (1977). Behavior of robust estimators in the regression model with dependent errors. *Ann. Statist.* **5** 681–699.

KOUL, H. L. (1979). Weighted empirical process and the regression model. *J. Indian Statist. Assoc.* **17** 83–91.

KOUL, H. L. (1980). Some weighted empirical inferential procedures for a simple regression model. To appear in Colloq. Nonparametric Inference, held in June 1980, Budapest.

MARTINOV, S. V. (1975). Computation of distribution functions of quadratic forms of normally distributed random variables. *Theor. Probab. Appl.* **20** 782–793.

MILLAR, P. W. (1981). Robust estimation via minimum distance methods. *Z. Wahrsch. verw. Gebiete* **55** 73–89.

MILLAR, P. W. (1982). Optimal estimation of a general regression function. *Ann. Statist.* **10** 717–740.

PARR, W. C. (1981). Minimum distance estimation: A bibliography. *Comm. Statist. Theor. Math.* **A10(12)** 1205–1224.

PARR, W. C. and SCHUCANY, W. R. (1980). Minimum distance and robust estimation. *J. Amer. Statist. Assoc.* **75** 616–624.

PARR, W. C. and DEWET, T. (1981). On minimum Cramer-Von Mises non parameter estimation. *Commun. Statist. Theor. Math.* **A10(12)** 1149–1166.

PIERCE, D. A. and KOPECKY, K. J. (1979). Testing goodness of fit for the distribution of errors in regression models. *Biometrika* **66** 1–5.

WILLIAMSON, M. A. (1979). Weighted empirical type estimation of the regression parameter. Ph.D. Thesis, Michigan State University.

WILLIAMSON, M. A. (1982). Cramer-Von Mises type estimation of the regression parameter: The rank analogue. *J. Multivariate Anal.* **12** 248–255.

DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824

INSTITUTE FOR MARITIME TECHNOLOGY
PO BOX 181
SIMONSTOWN, 7995, SOUTH AFRICA