The Annals of Statistics
1988, Vol. 16, No. 4, 1369–1389

RANK REGRESSION

BY JACK CUZICK

Imperial Cancer Research Fund Labs

An estimation procedure for (b, g) is developed for the transformation model $g(Y) = bz + \text{error}$, where g is an unspecified strictly increasing function. The estimator for b can be viewed as a hybrid between an M-estimator and an R-estimator. It differs from an M-estimator in that the dependent variable is replaced by a score based on ranks and from an R-estimator in that the ranks of dependent variable itself are used, not the ranks of the residuals. This provides robustness against the scale on which the variables are thought to be linearly related, as opposed to robustness against misspecification of the error distribution. Existence, uniqueness, consistency and asymptotic normality are studied.

1. Introduction. Consider the model

$$g(Y_i) = b_0 z_i + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1}$$

where g is some strictly increasing function, ε_i are i.i.d. and $z_1, \dots, z_n$ are nonrandom covariate values. This model states that, after some order preserving transformation, the dependent variable is related to z in a simple linear fashion except for errors of observation. Parametric forms for g have been studied extensively in the literature, notably the Box–Cox (1964) transform,

$$g(y, \alpha) = \begin{cases} (y^\alpha - 1)/\alpha, & \alpha > 0, \\ \log y, & \alpha = 0. \end{cases}$$

See also Tukey (1957).

In this paper we consider the case in which g is completely unspecified except for the fact that it is strictly increasing. This sort of model has been used extensively in medical applications and includes as special cases the proportional hazards model [Cox (1972)], the proportional odds model [Bennett (1983)] and the Pareto family of models [Clayton and Cuzick (1985, 1986)]. See also Prentice (1978), Kalbfleisch (1978), Kalbfleish and Prentice (1980), Pettitt (1982, 1984, 1987), Bickel (1986) and Doksum (1987). Applications exist in many other fields [cf. Heckman and Singer (1984) and Lancaster and Nickell (1980)].

Assuming that the errors ε_i are i.i.d. and independent of the $\{z_i\}$, which are assumed to be fixed, we seek a method for estimating b_0 from the pairs (y_i, z_i). (If the z_i are assumed random but independent of the ε_i, we may argue conditionally on their values.) Invariance arguments suggest basing inference on the vector of ranks $R = (R_1, \dots, R_n)$ of the $\{y_i\}$ via the marginal likelihood of ranks. If we assume ε has density f_0, let $t_i = g(y_i)$ and write $\{t \in R\}$ for the set

Received May 1986; revised February 1988.

AMS 1980 *subject classifications*. Primary 62G05, 62J05; secondary 62G30.

Key words and phrases. Transformation models, rank regression, semiparametric models, robust estimation, linear models.

of $\{t_1, \ldots, t_n\}$ which satisfy the ordering determined by R, then the marginal likelihood takes the form

$$L(b; R, z) = \int_{\{t \in R\}} \prod_{i=1}^{n} f_0(t_i - bz_i)\, dt_i$$

and, assuming $\phi(t) = -f_0'(t)/f_0(t)$ exists, the estimating equation takes the form

$$\frac{\partial \log L}{\partial b} = \sum_{i=1}^{n} z_i E_R \phi(t_i - bz_i) = 0, \tag{2}$$

where

$$E_R h(t_i) = \frac{\int_{\{t \in R\}} h(t_i) \prod_{j=1}^{n} f_0(t_j - bz_j)\, dt_j}{\int_{\{t \in R\}} \prod_{j=1}^{n} f_0(t_j - bz_j)\, dt_j}$$

is the conditional expectation given the ranks R and that the regression parameter equals b. Under weak conditions it can be shown that asymptotically equivalent estimators are obtained if (2) is replaced by

$$\sum_{i=1}^{n} z_i \phi(\hat{t}_i^b - bz_i) = 0, \tag{3}$$

where $\hat{t}_i^b = E_R(t_i)$. Note that $\hat{t}_i^b$ depends on b. This approach has been developed by Clayton and Cuzick (1985, 1986), but computation of $\hat{t}_i^b$ and the variance of the MLE is difficult, and the rigorous establishment of asymptotic properties of the estimator is not currently available.

In this paper an alternative approach, which is mathematically easier, is developed. It is based on replacing $\hat{t}_i^b$ by simpler scores, denoted $\bar{t}_i^b$, in (3) where $\bar{t}_i^b$ is defined as follows. Let F_0 be the distribution function for ε, let $F_b(t)$ be the distribution function of a randomly chosen t_i, i.e.,

$$F_b(t) = \frac{1}{n} \sum_{i=1}^{n} F_0(t - bz_i), \tag{4}$$

and define the adjusted empirical distribution function of the t_i as

$$\hat{F}(t) = \frac{1}{n+1} \sum_{i=1}^{n} I_{\{t_i \leq t\}};$$

then $\bar{t}_i^b = F_b^{-1}(\hat{F}(t_i))$. We shall often omit reference to the subscript on F_b when b equals b_0 so that $E\hat{F}(t) = n/(n+1)F(t)$. Other approximate marginal likelihood procedures have been considered by Pettitt (1982, 1984, 1987) and Doksum (1987).

Asymptotic normality of this estimator is established and a consistent estimate of the asymptotic variance is given. The special case $\phi(t) = t$ is considered in detail and it is shown that in this case (3) (with $\bar{t}_i^b$ in place of $\hat{t}_i^b$) always has a unique solution. Other simplifications are also discussed in this case. Limited simulations confirm the usefulness of the estimator and its estimated variance and compare $\hat{b}$ to the least squares estimator with g known. Comparisons with Cox's estimator for the proportional hazards model are also made.

Model (1) is clearly location and scale invariant, so that any constant is subsumed in the g function, and the regression parameter b_0 must be scaled in relation to the standard deviation of the error variable. The absolute values of these quantities require knowledge of the transformation g. This can be estimated by

$$\hat{g}(y_i) = \bar{t}_i^{\hat{b}}(y_i)$$

at the data points and interpolated in any reasonable way between them. When g is smooth, weak convergence of this estimator and joint normality of $(\hat{g}, \hat{b})$ are established.

These methods differ from R-estimators or "aligned-rank" estimators [Hodges and Lehmann (1963) and Sen (1963)] in that they are concerned with accounting for the transformation g and thus are based on the ranks of the y_i, whereas R-estimators aim at robustness against misspecification of the error distribution and use ranks of the residuals.

2. Main results. We first set out the basic notation and assumptions. Directions in which they can be weakened are considered later. All sums will be from 1 to n unless otherwise stated. Assume that:

(A1) The errors ε_i are i.i.d. with distribution function F_0 and density f_0 which is uniformly continuous and positive on $(-\infty, \infty)$. Uniform continuity of f_0 implies that it is bounded.

(A2) Let $G_n(z) = (1/n)\Sigma I_{\{z_i \leqq z\}}$. Assume G_n is nondegenerate, $G_n \to G$ as $n \to \infty$, where G is some nondegenerate distribution function and if Z_n has distribution G_n, then $|Z_n|^p$ is uniformly integrable for some p to be specified below (we must have $p > 4$). We also assume $\Sigma z_i = 0$, but this can always be achieved since estimation is invariant under shifts of the $\{z_i\}$.

The requirement that $G_n \to G$ can be weakened, since the uniform integrability implies that the $\{G_n\}$ are tight and so a convergent subsequence can always be taken, and the results stated below can be rephrased to accommodate this.

(A3) For $F_b(t)$ defined at (4), there exists $K < \infty$ such that

$$F_b^{-1}(u) + u(1-u)\big(F_b^{-1}(u)\big)' \leqq K\{u(1-u)\}^{-\alpha}, \qquad 0 < u < 1, \tag{5}$$

for all n and $b \in B$, some neighborhood of b_0. We also require that α satisfies $\alpha + p^{-1} < \frac{1}{2}$.

This is essentially a moment condition on $T = bZ_n + \varepsilon$, although it also implies further smoothness of the distribution. Since F_b is the distribution function for T, (A3) implies that

$$E|T|^{\gamma} < \infty \quad \text{for } \gamma < 1/\alpha. \tag{6}$$

The requirement $\alpha + p^{-1} < \frac{1}{2}$ plus (6), which implies $E|Z_n|^p < \infty$ for $p < \alpha^{-1}$, shows that we must have $p > 4$.

If $f_b(t) = (d/dt)F_b(t)$, the inequality for the second term on the left-hand side of (5) is equivalent to

$$f_b(t) \geqq K^{-1}\{F_b(t)(1 - F_b(t))\}^{1+\alpha}, \qquad -\infty < t < \infty.$$

The inequality for the first term on the left-hand side of (5) alone is implied by this inequality. The existence of a derivative for F_b and its inverse is implied by (A1).

(A4) The function ϕ is continuously differentiable and ϕ' is bounded on $(-\infty, \infty)$. Also $\phi'(t) \geqq 0$ for all t and is strictly positive on a set of positive measure. Since we assume $\Sigma z_i = 0$, by adding a constant if necessary we can always assume $E\phi(\varepsilon_i) = 0$.

(A5) The functions $\bar{Z}(bZ + \varepsilon)$ and $\phi'(U_b)$ are L_2 continuous as $b \to b_0$. These functions are defined at (10) and (19) below.

(A6) The expression $\sigma_2^2 = E(Z\{Z - \bar{Z}(b_0 Z + \varepsilon)\}\phi'(\varepsilon))$ is positive.

THEOREM 1. *Under assumptions* (A1)–(A6), *except on a set of probability tending to zero as* $n \to \infty$, *there exists a solution* $\hat{b}$ *to the equation*

$$l(b) \equiv \sum z_i \phi(\bar{t}_i^b - bz_i) = 0, \tag{7}$$

such that as $n \to \infty$,

$$n^{1/2}(\hat{b} - b_0) \to \mathcal{N}(0, \sigma^2),$$

where b_0 *is the true value of* b *given in* (1). *When* $\phi(t) = t$, *this solution exists and is unique for all* $n \geqq 2$.

The asymptotic variance is given by $\sigma^2 = \sigma_1^2/\sigma_2^4$, where

$$\begin{aligned}\sigma_1^2 &= \lim_{n\to\infty} \frac{1}{n} \sum \operatorname{Var}\left(z_i\phi(\varepsilon_i) - \int_0^{b_0 z_i + \varepsilon_i} h(s)\,ds \right)\\ &= E(Z^2)E(\phi^2(\varepsilon)) - 2E\left(Z\phi(\varepsilon)\int_0^{b_0 Z + \varepsilon} h(s)\,ds\right) \qquad (8)\\ &\quad + E\left(\int_0^{b_0 Z+\varepsilon} h(s)\,ds\right)^2 - E\left(\int_0^{b_0 Z+\varepsilon_1} h(s)\,ds\right)\left(\int_0^{b_0 Z+\varepsilon_2} h(s)\,ds\right)\end{aligned}$$

and

$$\begin{aligned}\sigma_2^2 &= \lim_{n\to\infty} \frac{1}{n} \sum z_i\left(z_i - \bar{z}_i^{b_0}\right)\phi'(\varepsilon_i) \qquad (9)\\ &= E\left(Z\{Z - \bar{Z}(b_0 Z + \varepsilon)\}\phi'(\varepsilon)\right).\end{aligned}$$

Here

$$h(s) = E(Z\phi'(\varepsilon)|b_0 Z + \varepsilon = s),$$

$$\bar{z}_i \equiv \frac{d\bar{t}_i^b}{db} = E(Z_n|bZ_n + \varepsilon = \bar{t}_i^b)$$

$$= \frac{\sum_{j=1}^n z_j f_0(\bar{t}_i^b - bz_j)}{\sum_{j=1}^n f_0(\bar{t}_i^b - bz_j)},$$

$$\bar{Z}(s) = E(Z|b_0 Z + \varepsilon = s), \tag{10}$$

where $\varepsilon, \varepsilon_1, \varepsilon_2, \ldots,$ are i.i.d. with distribution function F_0, independent of Z which has distribution function G. The variance σ^2 can be estimated consistently by replacing b_0 by $\hat{b}$ and Z by a variable with distribution G_n in these expressions.

REMARK 1. It is disappointing not to be able to give more comprehensive results about existence, uniqueness and consistency. One would expect that the results for $\phi = t$ would carry over to a very general set of (ϕ, ε, Z_n). However counterexamples can occur in small samples and pathological outcomes. For example let $Z = (-2, -1, 3)$, $R = (2, 1, 3)$ and assume $\phi(-\infty) = -\infty$, $\phi(+\infty) = C < \infty$. Then

$$l(b) = -2\phi(\bar{t}_{(2)}^b + 2b) - \phi(\bar{t}_{(1)}^b + b) + 3\phi(\bar{t}_{(3)}^b - 3b)$$

and as $b \to \infty$,

$$(t_{(1)}, t_{(2)}, t_{(3)}) \cong (-2b + \text{const.}, -b + \text{const.}, 3b + \text{const.}),$$

so that

$$l(b) \cong -2\phi(b) - \phi(-b) + 3\phi(\text{const.}) \to +\infty, \quad \text{as } b \to +\infty.$$

If $b \to -\infty$, then $(t_{(1)}, t_{(2)}, t_{(3)}) \cong (3b, -b, -2b)$, so

$$l(b) \cong -2\phi(b) - \phi(4b) + 3\phi(-5b) \to +\infty \quad \text{as } b \to -\infty.$$

Thus $l(b) = 0$ has either no solution or an even number of solutions, depending on the detailed specification of F_0 and ϕ. This behaviour is illustrated in Figure 1, where the pathological nature of the extra solution is apparent.

REMARK 2. The assumption (A6) that $\sigma_2^2 > 0$ would appear to hold quite generally, the case $\phi(t) = t$ being demonstrated in Lemma 5. It would be nice to have more general conditions under which it is true. The following can no doubt be extended:

PROPOSITION. *Assume φ', Z, ε are all symmetric*, $\mathrm{Var}(Z)\mathrm{Var}(\varepsilon) > 0$ *and for $t > 0$, $0 \le \bar{Z}(t) \le b_0^{-1}t$, and $\bar{Z}(t)$ is (weakly) concave. Then* $\sigma_2^2 > 0$.

PROOF. Without loss of generality take $b_0 = 1$. Condition on $|\varepsilon|$ to obtain

$$E(Z\bar{Z}(Z + \varepsilon)\varphi'(\varepsilon)) = \tfrac{1}{2}E(Z\{\bar{Z}(Z + \varepsilon) + \bar{Z}(Z - \varepsilon)\}\varphi'(\varepsilon)).$$

Now concavity and the fact that $\bar{Z}(t)$ is an odd function allows one to conclude that

$$\tfrac{1}{2}Z\{\bar{Z}(Z + \varepsilon) + \bar{Z}(Z - \varepsilon)\} \le Z\bar{Z}(Z) \le Z^2$$

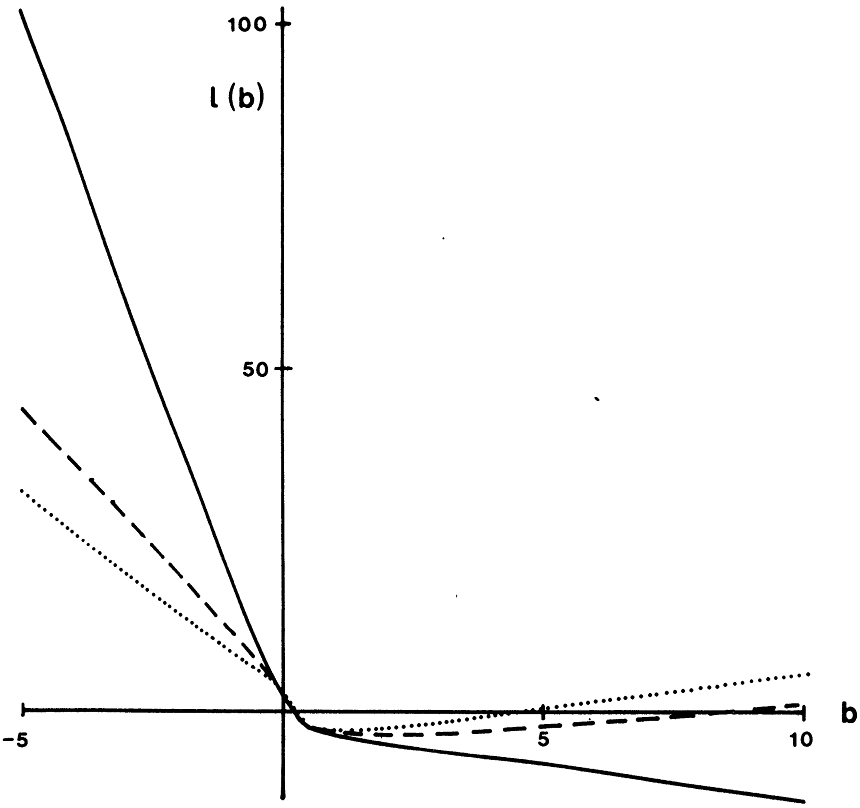

FIG. 1. *Plot of the estimating equation* (7) *for* $n = 3, \mathbf{Z} = (-2, -1, 3)', \mathbf{R} = (2, 1, 3)', (f, F) \sim N(0,1)$ *and three choices of* φ: (i) $\varphi(t) = t$ (—), (ii) $\varphi(t) = t,\ t \le 0;\ \varphi(t) = t^{1/2},\ t \ge 0$ (– –) *and* (iii) $\varphi(t) = t,\ t \le 1;\ \varphi(t) = 1,\ t \ge 1$ ($\cdots$).

so that $\sigma_2^2 \ge 0$. Equality can hold only if $\operatorname{Var}(Z) = 0$ or if $\bar{Z}(t) = t$ for all $t > 0$ and this requires $\operatorname{Var}(\varepsilon) = 0$. □

THEOREM 2. *If we assume* (A1)–(A6) *and define* $\hat{g}(y) = \bar{t}^{\delta}(y_{(i)})$ *for* $y \in [y_{(i)}, y_{(i+1)})$, *then* $\hat{g}(y) \to g(y)$ *as* $n \to \infty$ *for all continuity points* y *of* g. *If* g *is continuously differentiable, then*

$$n^{1/2}(\hat{g}(y) - g(y)) \to \gamma(y)$$

weakly in the Skorohod space D *on every bounded set, where* $\gamma(\cdot)$ *is a mean zero Gaussian process with covariance function*

$$E(\gamma(y_1)\gamma(y_2)) = \sigma^2\bar{Z}(t_1)\bar{Z}(t_2) + \frac{\rho(t_1)}{f(t_1)} + \frac{\rho(t_2)}{f(t_2)} + \frac{1}{f(t_1)}\frac{1}{f(t_2)}\left[F(t_1) - E\{F_0(t_1 - b_0Z)F_0(t_2 - b_0Z)\}\right], \qquad y_1 \leqq y_2,$$

where σ^2 and $\bar{Z}(t)$ are as in Theorem 1, $t_i = g(y_i)$, $i = 1, 2$, (f, F) *are the density and distribution function of* $b_0 Z + \varepsilon$, *respectively, and*

$$\rho(t) = \lim_{n\to\infty} n \operatorname{Cov}(\hat{b}, \hat{F}(t))$$
$$= E\big(Z\phi(\varepsilon) I_{\{b_0 Z + \varepsilon \leq t\}}\big) + E\bigg(F_0(t - b_0 Z)\int_t^\infty h(s)(1 - F_0(s - b_0 Z))\,ds\bigg).$$

Also $\hat{b}$ *and* $\hat{g}(y)$ *are asymptotically jointly normal with limiting covariance given by*

$$n \operatorname{Cov}(\hat{b}, \hat{g}(y)) \to \bar{Z}(t)\sigma^2 + \frac{\rho(t)}{f(t)},$$

where $t = g(y)$.

REMARK 3. When Z takes only a fixed finite set of values $z_1, \dots, z_k$, Doksum (1987) has proposed an alternative estimate for g. His estimator would appear to be less stable at values of y, where $\min_{j=1,\dots,k} f_0(g(y) - b_0 z_j)$ is small, but may perform better at other values of y.

The proofs of these results rely on a Taylor expansion of the estimating equation and use results of Neuhaus (1975) and van Zuijlen (1978) on empirical distributions of independent but not identically distributed random variables. The remainder terms consist of complicated expressions involving the joint empirical distribution of (t, z) and do not appear to be covered by general empirical process theory.

3. A special case. The results and expressions in the above theorems can be simplified in the important special case when $\phi(t) = t$. Most importantly a unique solution to the estimating equation always exists. Uniqueness follows from the monotonicity of $l(b)$ in this case which is established in Lemma 5 below.

We now show a solution always exists. For all b, $l(b)$ is increased if R_i and R_j are interchanged if $z_i > z_j$, but $R_i < R_j$. Thus $l(b)$ and $\hat{b}$ are maximized when z_i and R_i have perfect rank correlation and they are minimized when they have perfect inverse rank correlation. Thus it is enough to show $\lim_{b\uparrow\infty} l(b) < 0$ when z_i and R_i have perfect rank correlation and $\lim_{b\downarrow -\infty} l(b) > 0$ when they are inversely ordered. We will only establish the former statement. If all z_i are distinct, then, since $\hat{F}(t_i) = R_i/(n+1)$, it follows that as $b \to +\infty$,

$$F_b^{-1}(R_i/(n+1)) \cong bz_i + F_0^{-1}(\{i/(n+1) - (i-1)/n\}n)$$

so that $\bar{t}_i^b - bz_i \to F_0^{-1}(1 - i/(n+1))$. It follows that for b large enough z_i and $\bar{t}_i^b - bz_i$ will be perfectly negatively rank correlated. Since $\Sigma z_i = 0$, the fact that $\lim_{b\uparrow\infty} l(b) < 0$ for $n \geqq 2$ in this case is easily proved by induction on n. When some z_i are equal, $l(b)$ is unchanged if the associated $\phi(\bar{t}_i^b - bz_i)$ are averaged and the result still holds so long as the z_i are not all equal.

In this special case the estimating equation (7) can be rewritten as

$$\hat{b} = \sum z_i \bar{t}_i^{\hat{b}} / \sum z_i^2,$$

which suggests a simple iterative method of solution. Also we obtain (Lemma 5)

$$\sigma_2^2 = E\left(Z - \bar{Z}(b_0 Z + \varepsilon)\right)^2 \quad \text{and} \quad h(t) = \bar{Z}(t).$$

If ε is $\mathcal{N}(0,1)$ then

$$\bar{Z}(t) = \frac{d}{dt}\log\int_{-\infty}^{\infty}\exp\left(btz - \frac{b^2 z^2}{2}\right) dG(z).$$

If additionally Z is normal, say $\mathcal{N}(0, \sigma_z^2)$, then

$$\bar{t}_i^b \cong \left(1 + (b\sigma_z)^2\right)^{1/2} \Phi^{-1}\left(\frac{R_i}{n+1}\right),$$

where Φ is the standard normal distribution function,

$$\bar{Z}(t) = \frac{t}{b}\frac{(b\sigma_z)^2}{1 + (b\sigma_z)^2},$$

$$\sigma_1^2 = \sigma_z^2\left(1 - \frac{(b_0\sigma_z)^2}{2} + \frac{(b_0\sigma_z)^4}{2}\right)\left(1 + (b_0\sigma_z)^2\right)^{-1},$$

$$\sigma_2^2 = \sigma_z^2\left(1 + (b_0\sigma_z)^2\right)^{-1},$$

so that

$$\sigma^2 = \sigma_z^{-2}\left(1 + \tfrac{1}{2}(b_0\sigma_z)^2 + \tfrac{1}{2}(b_0\sigma_z)^6\right),$$

which can be compared to the value σ_z^{-2} obtained in parametric maximum likelihood estimation. Thus $\hat{b}$ is nearly fully efficient when either b_0 or σ_z^2 is close to zero, relative to the error variance.

4. Numerical results. Some simulation results are presented in Table 1. We have taken $n = 50$, the $\{z_i\}$ have been sampled from a uniform $[0,1]$ distribution and $b = 0.5, 1{,}0, 2.0, 4.0$. The errors were either standard normal or (minus) standard extreme value (s.e.v.) corresponding to a proportional hazards model. Each configuration was simulated 200 times and analyzed in four ways: methods (i) and (iii) used (7) with $\varphi(t) = t$ or $\varphi(t) = e^t - 1$ corresponding to normal errors and extreme value errors, respectively. Method (ii) was least squares which is optimal for the parametric model with normal errors and method (iv) was Cox's (1972) approach to analyzing the proportional hazards model which is known to have optimality properties amongst rank methods in this case. The same sequence of random numbers was used for different values of

TABLE 1

Mean, mean square root, skewness and kurtosis of the estimate $\hat{b}$ based on 200 simulations, $n = 50$, and $\{z_i\}$ drawn from a uniform [0, 1] distribution.

Method	b	Error distribution	$\bar{\hat{b}}$	mse($\hat{b}$)	γ_1	γ_2
(i) $\varphi(t) = t$			0.496	0.175	0.129	−0.105
(ii) L.S.			0.511	0.180	0.162	0.048
(iii) $\varphi(t) = e^t - 1$	0.5	$N(0,1)$	0.467	0.182	0.252	0.350
(iv) Cox P. H.			0.504	0.215	0.218	0.273
(i)			0.971	0.178	0.110	−0.362
(ii)			1.011	0.180	0.162	0.048
(iii)	1.0	$N(0,1)$	0.932	0.204	0.360	0.099
(iv)			1.002	0.225	0.413	0.026
(i)			1.908	0.201	0.200	−0.311
(ii)			2.011	0.180	0.162	0.048
(iii)	2.0	$N(0,1)$	1.928	0.291	0.392	0.012
(iv)			2.034	0.310	0.375	−0.076
(i)			3.631	0.409	0.133	−0.525
(ii)			4.011	0.180	0.162	0.048
(iii)	4.0	$N(0,1)$	4.058	0.852	−0.283	3.778[a]
(iv)			4.259	0.763	0.281	−0.364
(i)			0.452	0.157	0.091	0.331
(ii)			0.549	0.281	0.183	0.350
(iii)	0.5	s.e.v.	0.470	0.159	−0.024	−0.046
(iv)			0.513	0.191	−0.043	−0.066
(i)			0.877	0.176	0.221	0.186
(ii)			1.049	0.281	0.183	0.350
(iii)	1.0	s.e.v.	0.954	0.175	0.187	−0.072
(iv)			1.044	0.206	0.094	−0.196
(i)			1.668	0.306	0.374	−0.235
(ii)			2.049	0.281	0.183	0.350
(iii)	2.0	s.e.v.	1.900	0.243	0.504	0.042
(iv)			2.093	0.287	0.427	−0.177
(i)			3.080	1.146	0.415	−0.484
(ii)			4.049	0.281	0.183	0.350
(iii)	4.0	s.e.v.	3.726	0.525	0.790	0.681
(iv)			4.160	0.513	0.610	0.147

[a] Contains two outliers (0.75, 7.42). If these are omitted parameters are (4.066, 0.680, 0.328, −0.324).

b. The results are very encouraging. For $b = 0.5$ or $b = 1.0$ method (i) is indistinguishable from least squares for normal errors. When $b = 2.0$ method (i) has about 10% greater mean squared error than least squares and only when $b = 4.0$ corresponding to $b/\text{s.e.}(\hat{b}) \cong 15$ does method (i) show appreciable degradement below the optimal parametric method. This is due to both negative bias and increased variance for large b. This procedure appears to exhibit less bias and smaller increases in variance for large b_0 than those of Pettitt (1987)

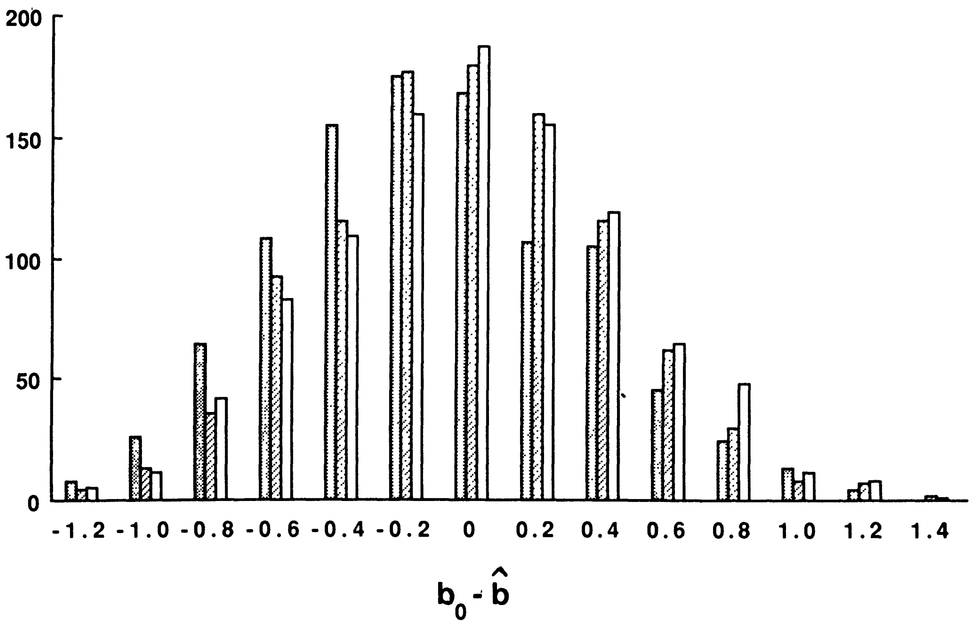

FIG. 2. *Histogram of* $\hat{\beta} - \beta$ *for standard normal errors,* $n = 50$, *and* $\varphi(t) = t$ *based on* 1000 *simulations for* $\beta = 0.5$, $\beta = 2.0$, *and for least squares (where the distribution does not depend on* β*). The* $\{z_i\}$ *are a fixed sample from a uniform* $[0, 1]$ *distribution and* $\Sigma_{i=1}^{50}(z_i - \bar{z})^2 = 5.249$. *Ordinate values refer to the center of an interval, i.e.,* 0.2 *refers to the interval* (0.1, 0.3).

and Doksum (1987). However all these procedures have very similar performance when b_0 is near zero. Similar results are found when comparing method (iii) with Cox's method for s.e.v. errors, although since Cox's method is known to lose efficiency compared to the best parametric method when b is large, the difference between method (iii) and Cox's method (iv) is smaller as b increases.

The table also illustrates the effects of analyzing extreme value errors as if they were normal and vice versa. Very little bias appears to result in analyzing normal errors by methods appropriate for proportional hazards data [methods (iii) and (iv)]. This is somewhat surprising since the regression coefficient must always be normalized to a standard error variable with rank methods and the variance of an extreme value variate is $\pi^2/6$ so one might expect the estimate to be inflated by a factor $\pi/\sqrt{6} \cong 1.28$. However a downward bias of about this factor is observed when s.e.v. data is analyzed by method (i). Thus, an assessment of the bias introduced when the shape of the error distribution is misspecified appears to be complicated.

The table suggests that the distribution of $\hat{b}$ is approximately normal and this is confirmed in Figure 2 where the histograms of $\hat{b}$ are plotted for method (i) and least squares for $b = 0.5$ and $b = 2.0$ and standard normal errors. These plots are based on 1000 simulations.

The estimator for the variance of $\hat{b}$ has been implemented as follows. Let $\hat{\sigma}^2(\hat{b}) = n^{-1}[\hat{\sigma}_1^2/\hat{\sigma}_2^4]$, define $\hat{\varepsilon}_i = \bar{t}_i^{\hat{b}} - \hat{b}z_i$ and compute $\hat{\sigma}_2^2$ from the first expression in (9) ignoring the limit and replacing b_0 with $\hat{b}$ and ε_i with $\hat{\varepsilon}_i$. Now from the proof of Lemma 4 one can write

$$\sigma_1^2 \cong \frac{1}{n}\operatorname{Var}\left[\sum_{i=1}^{n} z_i\left\{\varphi(\varepsilon_i) + \frac{\varphi'(\varepsilon_i)}{f_{b_0}(t_i)}\left(\hat{F}(t_i) - \frac{n}{n+1}F(t_i)\right)\right\}\right]$$
$$= A + B + C,$$

say, where these terms are approximated as follows:

$$A \cong \left[\frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2\right]\left[\frac{1}{n-1}\sum_{i=1}^{n}(\varphi(\hat{\varepsilon}_i) - \bar{\varphi})^2\right],$$
$$\bar{\varphi} = n^{-1}\sum_{i=1}^{n}\varphi(\hat{\varepsilon}_i),$$

which corresponds to the parametric variance,

$$B = \frac{2}{n}\sum_i\sum_j z_i z_j \operatorname{Cov}\left[\varphi(\varepsilon_i), \frac{\varphi'(\varepsilon_j)}{f_{b_0}(t_j)}\left(\hat{F}(t_j) - \frac{n}{n+1}F(t_j)\right)\right]$$
$$\cong \frac{2}{n(n+1)}\sum\sum_{i\neq j} z_i z_j \varphi(\hat{\varepsilon}_i)\frac{\varphi'(\hat{\varepsilon}_j)}{f_b(\bar{t}_j^b)} I_{\{\bar{t}_j^b \geq \bar{t}_i^b\}}\Bigg|_{b=\hat{b}}$$

and

$$C = \frac{1}{n}\operatorname{Var}\left\{\sum_{i=1}^{n} z_i \frac{\varphi'(\varepsilon_i)}{f_{b_0}(t_i)}\left(\hat{F}(t_i) - \frac{n}{n+1}F(t_i)\right)\right\}$$
$$\cong \frac{1}{n}\left(\frac{1}{n+1}\right)^2\sum_i\sum_j z_i z_j \frac{\varphi'(\hat{\varepsilon}_i)}{f_b(\bar{t}_i^b)}\frac{\varphi'(\hat{\varepsilon}_j)}{f_b(\bar{t}_j^b)}$$
$$\times\left\{2\left[I_{\{\bar{t}_j^b \leq \bar{t}_i^b\}} - F_0(\bar{t}_i^b - bz_j)\right]\left[I_{\{\bar{t}_i^b \leq \bar{t}_j^b\}} - F_0(\bar{t}_j^b - bz_i)\right]\right.$$
$$\left. + \sum_{\substack{k\neq i\\ k\neq j}} F_0\left[\min(\bar{t}_i^b, \bar{t}_j^b) - bz_k\right] - F_0(\bar{t}_i^b - bz_k)F_0(\bar{t}_j^b - bz_k)\right\}\Bigg|_{b=\hat{b}}.$$

The performance of this estimator would appear to be quite satisfactory, based on the evidence provided in Table 2.

TABLE 2
Comparison of the simulated and approximated variance for $n = 50$ based on 200 simulations. The simulation is based on the same data as for Table 1.

Method	b	Error distribution	Simulated variance	$\bar{\hat{\sigma}}^2$	$\mathrm{se}(\hat{\sigma}^2)$
(i) $\varphi(t) = t$	0.5	$N(0,1)$	0.174	0.178	0.010
(ii) $\varphi(t) = e^t - 1$			0.181	0.162	0.045
(i)	2.0	$N(0,1)$	0.192	0.215	0.028
(ii)			0.285	0.293	0.216
(i)	0.5	s.e.v.	0.155	0.178	0.009
(ii)			0.158	0.159	0.044
(i)	2.0	s.e.v.	0.195	0.210	0.030
(ii)			0.233	0.235	0.115

5. Proofs.

The following lemma is due to van Zuijlen (1978).

LEMMA 1 (van Zuijlen). *Let Y_i be independent random variables with distribution functions F_i, $i = 1, \dots, n$, and define*

$$F(t) = n^{-1} \sum_{i=1}^{n} F_i(t), \qquad \hat{F}(t) = (n+1)^{-1} \sum I_{\{Y_i \leq t\}}.$$

Then for any $\varepsilon > 0$, $\delta > 0$, there exist constants $0 < K_1 < K_2 < \infty$ depending only on ε and δ such that for all t and any fixed n,

$$K_1 F(t) \leq \hat{F}(t) \leq K_2 F(t) \quad \text{on } \hat{F}(t) > 0, \tag{11}$$

$$K_1(1 - F(t)) \leq 1 - \hat{F}(t) \leq K_2(1 - F(t)) \quad \text{on } \hat{F}(t) < n/(n+1), \tag{12}$$

$$n^{1/2}|\hat{F}(t) - F(t)| \leq K_2\{F(t)(1 - F(t))\}^{(1/2) - \varepsilon}, \tag{13}$$

with probability greater than or equal to $1 - \delta$.

The following two lemmas bound remainder terms in the expansion of (7) about b_0.

LEMMA 2. *Assume* (A1)–(A4), *let $u_i = F(t_i)$ and $\hat{u}_i = \theta_i F(t_i) + (1 - \theta_i)\hat{F}(t_i)$ for any random $\theta_i \in [0, 1]$. Then, as $n \to \infty$,*

$$\begin{aligned} &n^{-1/2} \sum z_i \left[\phi'\left(F_b^{-1}(u_i) - bz_i\right)\left(F_b^{-1}(u_i)\right)' - \phi'\left(F_b^{-1}(\hat{u}_i) - bz_i\right)\left(F_b^{-1}(\hat{u}_i)\right)'\right] \\ &\quad \times \left(F(t_i) - \hat{F}(t_i)\right) \to_P 0. \end{aligned} \tag{14}$$

PROOF. From Lemma 1 we know that given any $\delta > 0$, there exist $0 < K_1 < K_2 < \infty$ such that when $0 < \hat{F}(t_i) < n/(n+1)$,

$$K_1 u_i(1 - u_i) \leq \hat{u}_i(1 - \hat{u}_i) \leq K_2 u_i(1 - u_i)$$

with probability at least $1 - \delta$. Thus, using (A3) and (A4), except on a set of small probability, we can bound (14) on the set $\{F(t_i) < \varepsilon_0\}$ by a constant times

$$n^{-1/2} \sum |z_i| \frac{|F(t_i) - \hat{F}(t_i)|}{F^{1+\alpha}(t_i)} I_{\{F(t_i) < \varepsilon_0\}}.$$

Using the bound (13) and letting $\hat{F}_z(t) = n^{-1} \sum |z_i| I_{\{t_i \leq t\}}$, this can be bounded by a constant times

$$\int_{\{F(t) < \varepsilon_0\}} F^{-(1/2+\alpha+\varepsilon)} \, d\hat{F}_z(t)$$

except on a set of vanishing probability. After an integration by parts, this equals

$$(15) \qquad F^{-(1/2+\alpha+\varepsilon)}(t) \hat{F}_z(t) \Big|_{F(t)=0}^{F(t)=\varepsilon_0} - \int_{\{F(t) < \varepsilon_0\}} \hat{F}_z(t) \, dF^{-(1/2+\alpha+\varepsilon)}(t).$$

Now for $p \geqq 1$,

$$(16) \qquad \hat{F}_z(t) \leqq \left(\frac{1}{n} \sum |z_i|^p \right)^{p^{-1}} \left(\hat{F}(t)\right)^{1-p^{-1}},$$

which by (A4) and (11) is less than a constant times $F(t)^{1-p^{-1}}$ with probability exceeding $1 - \delta$. By choosing ε so that $\alpha + \varepsilon + p^{-1} < \frac{1}{2}$, the expression (15) can be bounded by a constant times

$$\left(F(\varepsilon_0)\right)^{1/2-(\alpha+\varepsilon+p^{-1})} \to 0 \quad \text{as } \varepsilon_0 \to -\infty.$$

Similarly we can bound (14) on the set where $\{1 - F(t_i) < \varepsilon_0\}$. After deleting these sets it is clear than we can choose M so large that the contribution from terms with $|z_i| > M$ can also be ignored. On the remaining set use the bound (13), write the summation as an integral as above and invoke the bounded convergence theorem and the continuity of ϕ', F^{-1} and $(F^{-1})'$ to obtain the result. Continuity of F^{-1} and $(F^{-1})'$ follows from the assumption of continuity of f_0 in (A1). □

LEMMA 3. *Let* $\tau(t) = F_b^{-1}(F(t))$. *Under assumptions* (A1)–(A4),

$$(17) \qquad n^{1/2} \int z \frac{\left(\hat{F}(t) - F(t)\right)}{f_b(\tau(t))} \phi'(\tau(t) - bz) \, d\left(\hat{F}(t, z) - F(t, z)\right)$$

tends to zero in probability as $n \to \infty$. *Here*

$$\hat{F}(t, z) = n^{-1} \sum I_{\{t_i \leq t, \, z_i \leq z\}}$$

and

$$F(t, z) = E\hat{F}(t, z) = \int_{\substack{y \leq z \\ s \leq t}} dF_0(s - by) \, dG_n(y).$$

PROOF. From (A3) there exists a K such that

$$(f_b(\tau(t)))^{-1} \leqq K(F(t))^{-(1+\alpha)}.$$

Given any $\delta > 0$, bounding $d(\hat{F}(t, z) - F(t, z))$ by $d\hat{F}(t, z) + dF(t, z)$ and estimating each term separately it can be shown, as in Lemma 2, that the integral over $\{F(t) < \varepsilon_0\}$ is less than a constant times

$$\int_{\{F(t) < \varepsilon_0\}} F^{-(1/2+\alpha+\varepsilon)}(t)\, dF_z(t)$$

with probability $1 - \delta$. An integration by parts and use of the bound (16) shows that when $\alpha + p^{-1} + \varepsilon < \frac{1}{2}$, this tends to zero in probability as $\varepsilon_0 \downarrow 0$ uniformly in n. A similar argument can be used to bound the set on which $1 - F(t) \leqq \varepsilon_0$. Also this argument can be adapted to omit the set $\{|z| > M\}$ from the domain of integration for M large enough. Thus it is enough to show the expected square of (17) restricted to the set $\{\varepsilon_0 \leqq F(t) \leqq 1 - \varepsilon_0, |z| < M\}$ tends to zero as $n \to \infty$. Writing $\hat{F}(t)$, $F(t)$, $\hat{F}(t, z)$, $F(t, z)$ in terms of their defining sums gives the expression

$$\begin{aligned}
&\frac{1}{n(n+1)^2} \iint_{\substack{\varepsilon_0 \leqq F(s_1) \leqq 1-\varepsilon_0 \\ \varepsilon_0 \leqq F(s_2) \leqq 1-\varepsilon_0}} \sum_{\substack{i=1 \\ |z_i|<M}}^{n} \sum_{j=1}^{n} \sum_{\substack{k=1 \\ |z_j|<M}}^{n} \sum_{l=1}^{n} z_i z_j \frac{\phi'(\tau(s_1) - bz_i)\phi'(\tau(s_2) - bz_j)}{f_b(\tau(s_1)) f_b(\tau(s_2))} \\
(18)\qquad & \times E\Big[\big(I_{\{t_k \leqq s_1\}} - F_0(s_1 - bz_k)\big)\big(\big(I_{\{t_l \leqq s_2\}} - F_0(s_2 - bz_l)\big) \\
&\times d\big\{\big(I_{\{t_i \leqq s_1\}} - F_0(s_1 - bz_i)\big)\big(I_{\{t_j \leqq s_2\}} - F_0(s_2 - bz_j)\big)\big\}\Big].
\end{aligned}$$

The integrals can be viewed as Riemann–Stieljes integrals and written as limits of approximating Riemann sums. Dominated convergence allows one to interchange the expectation and limit operations. It then follows from independence that all terms are zero unless

$$\{i = j,\ k = l,\ i \neq k\}, \qquad \{i = k,\ j = l,\ i \neq j\},$$
$$\{i = l,\ j = k,\ i \neq j\} \quad \text{or} \quad \{i = j = k = l\}.$$

There are less than $3n^2$ such terms and the contribution from each of these sets will be shown to go to zero. Without loss of generality we can assume $s_1 \leqq s_2$. Consider, first the terms where $\{i = j,\ k = l,\ i \neq k\}$. The expectation equals

$$\begin{aligned}
&\{F_0(s_1 - bz_k)(1 - F_0(s_2 - bz_k))\} \\
&\quad \times \big\{I_{\{s_1 = s_2\}}\, dF_0(s_1 - bz_i) - dF_0(s_1 - bz_i)\, dF_0(s_2 - bz_i)\big\}.
\end{aligned}$$

Now all the terms in the integrand are bounded and the integral of the expression above is bounded in absolute value by

$$\int dF_0(s_1 - bz_i) + \iint dF_0(s_1 - bz_i)\, dF(s_2 - bz_i) = 2,$$

so that these terms give a contribution which is $O(n^{-1})$. On the set $\{i = k,\ j = l,\ i \neq j\}$, the expectation in (18) equals

$$(1 - F_0(s_1 - bz_i))(1 - F_0(s_2 - bz_j))\, dF_0(s_1 - bz_i)\, dF_0(s_2 - bz_j),$$

whereas on the set $\{i = l,\ j = k,\ i \neq j\}$ it equals

$$(1 - F_0(s_2 - bz_i))(-F_0(s_1 - bz_j))\, dF_0(s_1 - bz_i)\, dF_0(s_2 - bz_j).$$

Both these sets again yield terms which when integrated and summed are $O(n^{-1})$. Finally, when $\{i = j = k = l\}$, by considering the sets $\{t_i = s_1 = s_2\}$, $\{t_i < s_1\}$, $\{t_i = s_1 < s_2\}$, $\{s_1 < t_i < s_2\}$, $\{s_1 < t_i = s_2\}$ and $\{t_i > s_2\}$ separately, the expectation becomes

$$\begin{aligned}
I_{\{s_1 = s_2\}}&(1 - F_0(s_1 - bz_i))^2\, dF_0(s_1 - bz_i)\\
&+ \{F_0(s_1 - bz_i)(1 - F_0(s_1 - bz_i))(1 - F_0(s_2 - bz_i))\\
&\quad - (1 - F_0(s_1 - bz_i))(1 - F_0(s_2 - bz_i))\\
&\quad + (F_0(s_2 - bz_i) - F_0(s_1 - bz_i))(-F_0(s_1 - bz_i))(1 - F_0(s_2 - bz_i))\\
&\quad - (-F_0(s_1 - bz_i))(1 - F_0(s_2 - bz_i))\\
&\quad + (1 - F_0(s_2 - bz_i))(-F_0(s_1 - bz_i))(-F_0(s_2 - bz_i))\}\\
&\quad \times dF_0(s_1 - bz_i)\, dF_0(s_2 - bz_i).
\end{aligned}$$

When integrated and summed these terms will give a contribution of order $O(n^{-2})$. □

The bulk of the proofs of Theorems 1 and 2 is carried out in the following two lemmas which are of some independent interest.

LEMMA 4. *Assume* (A1)–(A4). *Then for* $b \in B$ *as defined in* (A3) *and* $n \to \infty$,

$$n^{1/2}\left(\frac{1}{n}\sum z_i \phi(\bar{t}_i^b - bz_i) - \mu_n\right) \to \mathcal{N}(0, \sigma_1^2(b)),$$

where

$$\begin{aligned}
&\mu_n \equiv E(Z_n \phi(U_{b,n})) \to \mu = E(Z\phi(U_b)),\\
(19)\qquad &U_b = F_b^{-1} \circ F_{b_0}(bZ + \varepsilon) - bZ
\end{aligned}$$

and $U_{b,n}$ *is defined similarly except* Z_n *replaces* Z.

The variance satisfies $0 < \sigma_1^2(b) < \infty$ *and is given by*

$$\begin{aligned}
\sigma_1^2(b) &= E(Z^2 \mathrm{Var}(\phi(U_b)|Z)) - 2E(Z\phi(U_b) - \mu)\left(\int_0^{b_0 Z + \varepsilon} h_b(s)\, ds\right)\\
(20)\qquad &+ E\left(\int_0^{b_0 Z + \varepsilon} h_b(s)\, ds\right)^2 - E\left(\int_0^{b_0 Z + \varepsilon_1} h_b(s)\, ds\right)\left(\int_0^{b_0 Z + \varepsilon_2} h_b(s)\, ds\right),
\end{aligned}$$

where

$$h_b(s) = \frac{f_{b_0}(s)}{f_b(s)} E(Z\phi'(U_b)|b_0 Z + \varepsilon = s). \tag{21}$$

When $b = b_0$, U_{b_0} *and* Z *are independent and* $\mu_n = \mu = 0$.

The random variables Z_n, Z, ε, ε_1, ε_2 *are all assumed independent*; Z_n *has df* G_n, Z *has df* G *and* ε, ε_1, ε_2 *have df* F_0.

PROOF. Let $\tau_i = F_b^{-1} \circ F(t_i)$ and expand about $F(t_i)$ to get

$$\begin{aligned} n^{-1/2} \sum z_i \phi(\bar{t}_i^b - bz_i) &= n^{-1/2} \sum z_i \phi(\tau_i - bz_i) \\ &\quad + n^{-1/2} \sum z_i \frac{(\hat{F}(t_i) - F(t_i))}{f_b(\tau_i)} \phi'(\tau_i - bz_i) \\ &\quad + n^{-1/2} \sum z_i (\hat{F}(t_i) - F(t_i)) \left(\frac{\phi'(\tilde{\tau}_i - bz_i)}{f_b(\tilde{\tau}_i)} - \frac{\phi'(\tau_i - bz_i)}{f_b(\tau_i)} \right), \end{aligned} \tag{22}$$

where $\tilde{\tau}_i \in [\tau_i, F_b^{-1} \circ \hat{F}(t_i)]$. By Lemma 2 the last term can be ignored. The second term can be written as

$$n^{1/2} \left[\int z \frac{\hat{F}(t) - F(t)}{f_b(\tau(t))} \phi'(\tau(t) - bz)\, dF(t, z) + \int z \frac{\hat{F}(t) - F(t)}{f_b(\tau(t))} \phi'(\tau(t) - bz)\, d(\hat{F}(t, z) - F(t, z)) \right],$$

where $\tau(t)$, $\hat{F}(t, z)$ and $F(t, z)$ are defined in Lemma 3.

Now Lemma 3 permits one to neglect the last term in this expression, so that (22) is asymptotically equivalent to

$$n^{-1/2} \sum \left[z_i \phi(\tau_i - bz_i) + \int z \left(I_{\{t_i \le t\}} - F_0(t - bz_i) \right) \times \frac{\phi'(\tau(t) - bz)}{f_b(\tau(t))}\, dF_0(t - bz)\, dG_n(z) \right], \tag{23}$$

where $F(t)$ has been replaced by $n(n+1)^{-1}F(t)$ and $\hat{F}(t) - n(n+1)^{-1}F(t)$ has been written as a sum. This is a sum of independent random variables and has mean

$$n^{-1/2} \sum z_i E\phi(\tau_i - bz_i) = n^{1/2} E(Z_n \phi(\tau(\varepsilon + b_0 Z_n) - bZ_n)).$$

Asymptotic normality is easily established from the Liapounov theorem and boundedness of the variance can be established by bounding the variance of the sum of each of the two terms in (23) using the methods of Lemmas 2 and 3. The form given for the asymptotic variance can easily be derived by noting that

$$\int z\left\{I_{\{t_i \leqq t\}} - F_0(t - bz_i)\right\} \frac{\phi'(\tau(t) - bz)}{f_b(\tau(t))} \, dF_0(t - bz)\, dG_n(z)$$

$$= \int \left\{I_{\{t_i \leqq t\}} - F_0(t - bz_i)\right\} h_b(t)\, dt$$

$$= -\int_0^{t_i} h_b(t)\, dt + \text{const.} \qquad \square$$

Lemma 5. *Under* (A1) *and* (A4)

$$-\frac{d}{db}\sum z_i \phi\left(\bar{t}_i^b - bz_i\right) = \sum z_i(z_i - \bar{z}_i)\phi'\left(\bar{t}_i^b - bz_i\right).$$

If (A2), (A3) *and* (A5) *are also assumed and* $b \to b_0$ *as* $n \to \infty$, *then*

$$\frac{1}{n}\sum z_i(z_i - \bar{z}_i)\phi'\left(\bar{t}_i^b - bz_i\right) \to \sigma_2^2, \tag{24}$$

where σ_2^2 *is given by* (9). *When* $\phi(t) = t$, *the left-hand side of* (24) *is positive for all* b *and all* $n \geqq 2$ *and* $\sigma_2^2 = E\{Z - \bar{Z}(b_0 Z + \varepsilon)\}^2$.

Proof. The first equality follows from

$$\frac{dF_b^{-1}(u)}{db} = -\left.\frac{dF_b(y)}{db}\middle/\frac{dF_b(y)}{dy}\right|_{y = F_b^{-1}(u)},$$

which can be obtained from letting $y = F_b^{-1}(u)$, so that $u = F_b(y)$ and

$$0 = \frac{du}{db} = \frac{\partial F_b(y)}{\partial y}\frac{dy}{db} + \frac{\partial F_b(y)}{\partial b}.$$

Now the left-hand side of (24) can be written as

$$E\left(Z_n\{Z_n - \bar{Z}(\bar{t}^b)\}\phi'\left(\bar{t}^b - bZ_n\right)\right), \tag{25}$$

where $\bar{t}^b = F_b^{-1} \circ \hat{F}(b_0 Z_n + \varepsilon)$ and Z_n is defined in (A2).

Since $EZ_n = 0$ and $\bar{Z}$ is a conditional expectation,

$$EZ_n^2 = E\left(Z_n - \bar{Z}\right)^2 + E(\bar{Z}^2),$$

where the argument of $\bar{Z}$ has been suppressed. Therefore if M is any set and I_M denotes the indicator of this set

$$\begin{aligned}\left|E\left(Z_n(\bar{Z} - Z_n)\phi'\left(\bar{t}^b - bZ_n\right)I_M\right)\right| &\leqq (\text{const.})E^{1/2}\left(Z_n^2 I_M\right)E^{1/2}\left(Z_n - \bar{Z}\right)^2 \\ &\leqq (\text{const.})E\left(Z_n^2\right)E^{1/2}\left(Z_n^2 I_M\right),\end{aligned}$$

since $|\phi'|$ is bounded, and this will tend to zero if $EI_M \to 0$. Thus by taking M large enough, we can restrict (25) to the set on which $|Z_n| < M$ and see that (25) converges to

$$E\left(Z\{Z - \bar{Z}(b_0 Z + \varepsilon)\}\phi'(\varepsilon)\right) \equiv \sigma_2^2. \tag{26}$$

Now $\bar{z}_i$ is a conditional expectation so that $\bar{Z}_n^b \perp Z_n - \bar{Z}_n^b$ for all n and b. Thus when $\varphi(t) = t$ so that $\varphi'(t) = 1$, the left-hand side of (24) can be written as

$$\frac{1}{n}\sum (z_i - \bar{z}_i)^2 = E(Z_n - \bar{Z}_n)^2$$

which is positive provided $\operatorname{Var}(Z_n) > 0$. In particular $\sigma_2^2 = E\{Z - \bar{Z}(b_0 Z + \varepsilon)\}^2$. □

PROOF OF THEOREM 1. We first show that except for a set of probability tending to zero (which may vary with n) that a solution to (7) exists which tends to b_0. Consider the mean function $\mu(b) = E(Z\phi(U_b))$. We know $\mu(b_0) = 0$ and $-d\mu/db = E(Z(Z - \bar{Z})\phi'(U_b))$ is positive at b_0 and by (A5) is continuous there. Thus there exist $b_1 < b_0 < b_2$ such that $\mu(b)$ is strictly decreasing on (b_1, b_2) and so given any $\varepsilon > 0$, there exist

$$b_0 - \varepsilon < b_3 < b_0 < b_4 < b_0 + \varepsilon$$

such that $\mu(b_3) > 0$, $\mu(b_4) < 0$. It follows from the asymptotic normality of $l(b)$ for any $b \in B$ (Lemma 4) that given any $\delta > 0$, there exists an N, such that for $n \geqq N$, $l(b_3) > 0$ and $l(b_4) < 0$ except on a set of probability less than δ. Since $l(b)$ is continuous, this establishes the existence of a $\hat{b}$ such that $l(\hat{b}) = 0$ and $\hat{b} \to b_0$.

Now expand $l(\hat{b})$ around b_0 to obtain

$$0 = l(b_0) + (\hat{b} - b_0)l'(\tilde{b}),$$

where $\tilde{b} \in [b_0, b]$, so that

$$n^{1/2}(\hat{b} - b_0) = -\frac{n^{-1/2}l(b_0)}{n^{-1}l'(\tilde{b})} \to N(0, \sigma^2)$$

with $\sigma^2 < \infty$ by Lemmas 4 and 5 and (A6). The fact that $\sigma^2 > 0$ and that σ^2 can be estimated consistently from the data are easily checked. □

PROOF OF THEOREM 2. Weak convergence of $n^{1/2}\{\hat{F}(t) - F(t)\}$ to a mean zero Gaussian process follows from the results of Neuhaus (1975). Continuity of g then ensures that $\hat{g}(y) \to g(y)$, a.s. at continuity points, and differentiability ensures weak convergence of $n^{1/2}(\hat{g}(y) - g(y))$ to a mean zero Gaussian process on every bounded set. The covariance function of this process is easily computed, as is the asymptotic joint normally of $\hat{b}$ and $\hat{g}(t)$ and their covariance. The methods are extremely similar to those in the proof of Lemma 4. □

6. Extensions. The theorems can be made more general at the expense of added mathematical complexity. An indication of the possibilities is given below.

(i) The assumption that ϕ' is continuous can be dropped completely since the existence of ϕ' implies that

$$\liminf_{\varepsilon \downarrow 0} (2\varepsilon)^{-1} \int_{t-\varepsilon}^{t+\varepsilon} \phi'(s)\, ds = \phi'(t) \quad \text{a.e.}$$

and that will suffice to prove Lemma 3. Thus truncated versions of ϕ, i.e., $\phi^M(t) = \min(M, \max(-M, \phi(t)))$, can be considered. Discontinuities in ϕ itself, such as $\phi(t) = -1 + 2I_{\{t \geq 0\}}$, appear to be more difficult to treat asymptotically, in particular $l(b)$ is no longer continuous.

(ii) The assumption that $|\phi'(t)|$ is bounded can be weakened if stronger conditions are placed on the $\{z_i\}$. The proportional hazards model corresponds to $\phi'(t) = e^t$ and in this case the results can be established if we alter (A1) to $E\phi^2(\varepsilon_i) < \infty$, strengthen (A2) to $\exp(bZ_n)$ is uniformly integrable in n for all b and replace (A3) by the requirement that $|d/du\phi(F_b^{-1}(u))| \leq K\{u(1-u)\}^{-(1+\alpha)}$ for all $b \in B$ and $\alpha < \frac{1}{2}$.

(iii) Correction to Cuzick (1985). This paper addressed similar questions related to nonparametric testing when censoring is present. Unfortunately the condition "$|z_i|$ are bounded" was omitted from the statement of Theorem 1. The conditions used in this paper can be weakened by using current techniques.

(iv) Although nonparametric in the transformation, the current analysis retains a parametric assumption about the error distribution. Incorrect specification of f_0 will lead to biased estimators, although it is not clear how serious this will be in practice. A more general theory would allow an adaptive estimate of f_0 also. Clayton and Cuzick (1985, 1986) consider a parametric family of error distributions, but a more general theory would be of interest.

(v) Extensions to multivariate predictors are straightforward. $\mathbf{b} = (b_1, \ldots, b_k)$ and $\mathbf{z}_i = (z_i^1, \ldots, z_i^k)'$ now become vectors and a solution is sought to the vector equation

$$(27) \qquad \mathbf{0} = \sum_{i=1}^{n} \mathbf{z}_i \varphi\left(\bar{t}_i^{\mathbf{b}} - \mathbf{b}'\mathbf{z}_i\right).$$

Under technical assumptions similar to (A1)–(A6), Theorems 1 and 2 can be extended in a straightforward manner and the special properties associated with $\varphi(t) = t$ can also be established. In particular uniqueness of a solution to (27) follows by taking a derivative with respect to $\mathbf{b}$ of the right-hand side and yields the matrix

$$\sum_{i=1}^{n} \mathbf{z}_i(\bar{\mathbf{z}}_i - \mathbf{z}_i)' = -\sum_{i=1}^{n} (\mathbf{z}_i - \bar{\mathbf{z}}_i)(\mathbf{z}_i - \bar{\mathbf{z}}_i)',$$

which can be shown to be symmetric and negative definite in a manner similar to Lemma 5. Here

$$\bar{\mathbf{z}}_i = \left(\bar{z}_i^1, \ldots, \bar{z}_i^k\right)' \quad \text{and} \quad \bar{z}_i^j = \frac{d\bar{t}_i^{\mathbf{b}}}{db_j}.$$

The existence of a solution follows by considering an iterative solution to (27). We obtain

$$(28) \qquad \mathbf{b}_n = \left(\sum z_i z_i'\right)^{-1}\left(\sum z_i \bar{t}_i^{\mathbf{b}_{n-1}}\right)$$

so that

$$\mathbf{b}_{n+1} - \mathbf{b}_n = \left(\sum \mathbf{z}_i \mathbf{z}_i'\right)^{-1} \sum \mathbf{z}_i\left(\bar{t}_i^{\mathbf{b}_n} - \bar{t}_i^{\mathbf{b}_{n-1}}\right)$$
$$= \left(\sum \mathbf{z}_i \mathbf{z}_i'\right)^{-1}\left(\sum \mathbf{z}_i \bar{\mathbf{z}}_i'\right)(\mathbf{b}_n - \mathbf{b}_{n-1}),$$

where $\bar{\mathbf{z}}_i$ is evaluated between $\mathbf{b}_{n-1}$ and $\mathbf{b}_n$. Now $\Sigma \bar{\mathbf{z}}_i(\mathbf{z}_i - \bar{\mathbf{z}}_i)' = 0$, so that

$$\sum \mathbf{z}_i \bar{\mathbf{z}}_i' = \sum \bar{\mathbf{z}}_i \bar{\mathbf{z}}_i',$$
$$\sum \mathbf{z}_i \mathbf{z}_i' = \sum (\mathbf{z}_i - \bar{\mathbf{z}}_i)(\mathbf{z}_i - \bar{\mathbf{z}}_i)' + \sum \bar{\mathbf{z}}_i \bar{\mathbf{z}}_i'$$

and $(\Sigma \mathbf{z}_i \mathbf{z}_i')^{-1}(\Sigma \bar{\mathbf{z}}_i \bar{\mathbf{z}}_i')$ has all eigenvalues less than unity in absolute value implying that the mapping (28) is a contraction and must have a fixed point.

Numerical work is needed in this multivariate setting.

(vi) The method can also be extended to deal with censored data. The quantity $R_i/(n+1)$ must be replaced by a Kaplan–Meier (or equivalent) estimator in the definition of the $\bar{t}_i^b$ and the score function φ must be extended to deal with censored observations as in Prentice (1978). More work is needed to establish the properties of b in this setup.

Acknowledgment. I thank Robert Edwards doing the programming associated with the numerical results in Section 4.

REFERENCES

BENNETT, S. (1983). Analysis of survival data by the proportional odds model. In *Statistics in Medicine* **2** 273–277. Wiley, New York.

BICKEL, P. J. (1986). Efficient testing in a class of transformation models. In *Papers on Semiparametric Models at the ISI Centenary Session* (R. D. Gill and M. N. Voors, eds.) 63–81. Report MS-R8614, Centre for Mathematics and Computer Science, Amsterdam.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26** 211–252.

CLAYTON, D. and CUZICK, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *J. Roy. Statist. Soc. Ser. A* **148** 82–117.

CLAYTON, D. and CUZICK, J. (1986). The semi-parametric Pareto model for regression analysis of survival times. In *Papers on Semiparametric Models at the ISI Centenary Session* (R. D. Gill and M. N. Voors, eds.) 19–30. Report MS-R8169, Centre for Mathematics and Computer Science, Amsterdam.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

CUZICK, J. (1985). Asymptotic properties of censored linear rank tests. *Ann. Statist.* **13** 133–141.

DOKSUM, K. A. (1987). An extension of partial likelihood methods for proportional hazards models to general transformation models. *Ann. Statist.* **15** 325–345.

HECKMAN, J. and SINGER, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* **52** 271–319.

HODGES, J. L., JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math., J. Statist.* **34** 598–611.

KALBFLEISCH, J. D. (1978). Likelihood methods and non-parametric tests. *J. Amer. Statist. Assoc.* **73** 167–170.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

LANCASTER, T. and NICKELL, S. (1980). The analysis of re-employment probabilities for the unemployed (with discussion). *J. Roy. Statist. Soc. Ser. A* **143** 141–165.

NEUHAUS, G. (1975). Convergence of the reduced empirical process for non-i.i.d. random vectors. *Ann. Statist.* **3** 528–531.

PETTITT, A. N. (1982). Inference for the linear model using a likelihood based on ranks. *J. Roy. Statist. Soc. Ser. B* **44** 234–243.

PETTITT, A. N. (1984). Proportional odds models for survival data and estimates using ranks. *Appl. Statist.* **33** 169–175.

PETTITT, A. N. (1987). Estimates for regression parameters using ranks. *J. Roy. Statist. Soc. Ser. B* **49** 58–67.

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65** 167–179.

SEN, P. K. (1963). On the estimation of relative potency in dilution (-direct) assays by distribution-free methods. *Biometrics* **19** 532–552.

TUKEY, J. W. (1957). On the comparative anatomy of transformations. *Ann. Math. Statist.* **28** 602–632.

VAN ZUIJLEN, M. C. A. (1978). Properties of the empirical distribution function for independent non-identically distributed random variables. *Ann. Probab.* **6** 250–266.

DEPARTMENT OF MATHEMATICS, STATISTICS,
AND EPIDEMIOLOGY
IMPERIAL CANCER RESEARCH FUND LABS
P.O. BOX 123
LINCOLN'S INN FIELDS
LONDON WC2A 3PX
UNITED KINGDOM