The Annals of Statistics
1984, Vol. 12, No. 2, 590–600

CONSISTENCY RESULTS FOR LINEAR REGRESSION WITH CENSORED DATA

BY I. R. JAMES AND P. J. SMITH

The University of Western Australia

Buckley and James (1979) proposed an estimation method for linear regression models with unspecified residual distribution and right-censored response variables. In this paper we consider weak consistency of the regression parameter estimators under regularity conditions which avoid restrictions on the censoring patterns.

1. Introduction. For a vector of n responses $\mathbf{y}$, a matrix of explanatory variables X and a parameter vector β, assume that

$$\mathbb{E}(\mathbf{y}) = X\beta,$$

and that the residuals

$$\varepsilon_i = y_i - \mathbf{x}_i^\top \beta, \quad i = 1, 2, \cdots, n,$$

are iid with an unspecified distribution and finite variance σ^2, where $\mathbf{x}_i^\top$ is the ith row of X. However, for each i, instead of observing $(y_i, \mathbf{x}_i)$, one observes $(z_i, \delta_i, \mathbf{x}_i)$, where $z_i = \min(y_i, t_i)$ for some censor variables t_i, and δ_i is an indicator variable taking the value 1 if $y_i \leq t_i$ (uncensored) and 0 if $y_i > t_i$ (censored). The censor variables t_i need not be random in general.

Three estimation methods for the above model have been proposed in the literature: by Miller (1976), Buckley and James (1979) and Koul, Susarla and Van Ryzin (1981). Of these, only the last has been studied comprehensively at a theoretical level, but it has been found to sometimes perform unsatisfactorily in practice (Miller and Halpern, 1982) since it is sensitive to the requirement that the censor variables t_i have a distribution which does not depend on the explanatory variables $\mathbf{x}_i$. Heuristic arguments by Miller (1976) indicate that his method requires the variables $t_i - \mathbf{x}_i^\top \beta$ to be identically distributed for the estimators to be consistent. However Mauro (1983) gives a counter-example showing that this assumption is not sufficient. In simulations, Buckley and James (1979) found that their method gave approximately unbiased estimates for the slope parameter in the simple linear regression model for a wide range of censoring patterns, some depending on the explanatory variables and others not.

This paper investigates consistency properties of the Buckley-James estimators. The simple linear regression model with a single explanatory variable is considered in detail in Section 2, where conditions are given which ensure the existence of a (possibly nonunique) consistent "solution" to the estimating equations. These conditions avoid assumptions about censoring patterns, thus

Received December 1982; revised December 1983.
AMS 1980 *subject classifications*. Primary 62J05, 62F12.
Key words and phrases. Censored regression, consistency, Kaplan-Meier.

supporting the empirical observation of approximate unbiasedness of the slope estimator mentioned above. An extension of the results to multiple linear regression is considered briefly in Section 3.

2. Simple linear regression.

2.1 *Buckley-James estimators.* Suppose we have a single explanatory variable and consider the model

$$(1) \qquad y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \cdots, n,$$

where the independent partial residuals $r_i = y_i - \beta x_i$ have distribution function F and survival function $S = 1 - F$. Note that F has mean α and variance σ^2. The method of Buckley and James (1979) is motivated by the expectation identity

$$\mathbb{E}[y_i\delta_i + \mathbb{E}(y_i \mid y_i > t_i)(1 - \delta_i)] = \alpha + \beta x_i,$$

and replaces the censored observations in the usual least-squares normal equations by their estimated conditional expectations in the following way. Let

$$e_i(b) = z_i - bx_i, \quad i = 1, \cdots, n,$$

and let

$$\hat{F}_b(u) = 1 - \prod\nolimits_{i:e_{(i)}(b)\le u} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}$$

denote the Kaplan-Meier product limit estimator (Kaplan and Meier, 1958) calculated from the $e_i(b)$. In this formula $e_{(i)}(b)$ is the ith ordered observed residual and $\delta_{(i)}$ its associated indicator. To overcome problems of definition if $e_{(n)}(b)$ is censored, the product limit estimator is modified here by always defining $\hat{F}_b(e_{(n)}(b)) = 1$ (Meier, 1975; Miller, 1976).

Putting $\hat{S}_b = 1 - \hat{F}_b$, define for each $i = 1, \cdots, n$,

$$\hat{\mathbb{E}}_b(y_i \mid y_i > t_i) = \begin{cases} t_i + \int_{t_i - bx_i}^{\infty} [\hat{S}_b(u)\, du/\hat{S}_b(t_i - bx_i)], & \text{if } t_i - bx_i < e_{(n)}(b) \\ 0, & \text{if } t_i - bx_i \ge e_{(n)}(b) \end{cases}$$

and let

$$\hat{y}_i(b) = y_i\delta_i + \hat{\mathbb{E}}_b(y_i \mid y_i > t_i)(1 - \delta_i).$$

Thus $\hat{y}_i(b)$ is the observed response y_i if uncensored, or an estimate of it, based on the $e_i(b)$, if censored. One then attempts to find estimators $\hat{\alpha}$, $\hat{\beta}$ of α, β, which satisfy the equations

$$(2) \qquad \sum\nolimits_{i=1}^{n} (\hat{y}_i(\hat{\beta}) - \hat{\alpha} - \hat{\beta}x_i) = 0$$

$$(3) \qquad \sum\nolimits_{i=1}^{n} (x_i - \bar{x})(\hat{y}_i(\hat{\beta}) - \hat{\beta}x_i) = 0.$$

Note that (3) depends only on $\hat{\beta}$, and once $\hat{\beta}$ is obtained, $\hat{\alpha}$ is found explicitly from (2).

If we define

$$\gamma_n(b) = (\textstyle\sum_{i=1}^n (x_i - \bar{x})\hat{y}_i(b))/\sum_{i=1}^n (x_i - \bar{x})^2,$$

equation (3) is equivalent to $\gamma_n(\hat{\beta}) = \hat{\beta}$. A natural iterative scheme for attempting to solve this starts with an initial estimate of $\hat{\beta}$ and successively updates it by $\gamma_n(\hat{\beta})$. However, these iterations need not converge, and indeed, since $\gamma_n(b)$ is discontinuous and piecewise linear in b, an exact solution of (3) need not exist. This situation suggests alternatives for the definition of a "solution" to (3): either

(A) a point $\hat{\beta}$ at which the left-hand side of (3) changes sign,

or

(B) a point at which the left-hand side of (3) is closest to zero.

We have found definition (A) convenient to work with in theoretical calculations in Section 2.3, although for a model with multiple covariates the analogue of (A) is difficult to formulate while (B) generalizes in a straightforward way. Note that neither definition implies uniqueness of the "solution".

2.2 *Consistency of* $\gamma_n(\beta)$ *for* β. In what follows we assume that F is absolutely continuous with density f and has support bounded above. The latter assumption avoids the considerable theoretical difficulties encountered in estimating F with censored data over the whole line (see for example Gill, 1983; Susarla and Van Ryzin, 1980), and will be reasonable for most practical applications. On the other hand we do not require that the support of F be bounded below. Often one will be working with positive measurements and the linearity will pertain to the log scale, in which case arbitrarily large negative values may be observed. Let $U = \sup\{u; F(u) < 1\} < \infty$.

Since we wish to regard the censor variables t_i as fixed, our proofs rely on the product limit estimator results of Meier (1975) rather than work which regards the censor times as random variables (for example, Breslow and Crowley, 1974; Gill, 1983; Yang, 1977). Following Meier, let $\mathscr{N}(u)$ denote the expected number of censored and uncensored values $e_i(\beta)$ exceeding u. Define $c_i = t_i - \beta x_i$ and let $\rightarrow_P$ denote convergence in probability.

THEOREM 1. *Suppose that*

(a) $\mathscr{N}(u) \rightarrow \infty$ *as* $n \rightarrow \infty$ *for all* $u < U$

(b) $\sum_{i=1}^n (x_i - \bar{x})^2 \rightarrow \infty$ *as* $n \rightarrow \infty$

(c) $\limsup_{n\rightarrow\infty}\left\{\dfrac{\sum_{i=1}^n S(c_i)\,|\,x_i - \bar{x}\,|}{\sum_{i=1}^n (x_i - \bar{x})^2}\right\} < \infty.$

Then $\gamma_n(\beta) \rightarrow_P \beta$.

REMARK. A sufficient condition for (c) is

$$\liminf_{n\rightarrow\infty}(1/N_c) \textstyle\sum_{i=1}^n (x_i - \bar{x})^2 > 0,$$

where N_c is the expected number of censored observations, since by the Cauchy-

Schwartz inequality

$$\sum_{i=1}^n S(c_i)\,|\,x_i - \bar{x}\,| \le \left(\sum_{i=1}^n S^2(c_i)\right)^{1/2}\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^{1/2}$$
$$\le N_c^{1/2}\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^{1/2}. \qquad \square$$

In order to prove Theorem 1, we make use of the following lemma, the proof of which is deferred to Section 2.4.

LEMMA 1. *If $\mathcal{N}(u) \to \infty$ as $n \to \infty$ for all $u < U$, then*

$$\sup_{c\in(-\infty,U)} |\,\hat{\mathbb{E}}_\beta(r_i\,|\,r_i > c) - \mathbb{E}(r_i\,|\,r_i > c)\,| \to_P 0,$$

where $\hat{\mathbb{E}}_\beta(r_i\,|\,r_i > c)$ is the estimator of $\mathbb{E}(r_i\,|\,r_i > c)$ based on the product limit estimator $\hat{F}_\beta$.

PROOF OF THEOREM 1. Let $\mu_i = \mathbb{E}(r_i\,|\,r_i > c_i)$ and let $\hat{\mu}_i = \hat{\mathbb{E}}_\beta(r_i\,|\,r_i > c_i)$. Then

$$\gamma_n(\beta) - \beta = \frac{\sum_{i=1}^n (x_i - \bar{x})r_i^*}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})(1 - \delta_i)(\hat{\mu}_i - \mu_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}$$
$$= \mathrm{I} + \mathrm{II}, \quad \text{say},$$

where $r_i^* = r_i\delta_i + \mu_i(1 - \delta_i)$. Since $\mathbb{E}\,(r_i^*) = \alpha$ and the r_i^* are independent, we have

$$\mathbb{E}(\mathrm{I}) = 0, \quad \mathrm{var}(\mathrm{I}) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2\sigma_i^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2},$$

where $\sigma_i^2 = \mathrm{var}(r_i^*) = \sigma^2 - \mathrm{var}(r_i\,|\,r_i > c_i)S(c_i)$. By (b) and Chebychev's inequality it follows that $\mathrm{I} \to_P 0$. Next,

$$|\,\mathrm{II}\,| \le \sup_i |\,\hat{\mu}_i - \mu_i\,| \left\{\frac{\sum_{i=1}^n |\,x_i - \bar{x}\,|\,(1 - \delta_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}\right\},$$

and again by Chebychev's inequality the term in braces converges in probability to $\left(\sum_{i=1}^n |\,x_i - \bar{x}\,|\,S(c_i)\right)/\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)$. Condition (c) and Lemma 1 now show that $\mathrm{II} \to_P 0$ and hence $(\gamma_n(\beta) - \beta) \to_P 0$. $\square$

Theorem 1 shows that under appropriate restrictions, if n is large the function $\gamma_n(b) - b$ will be close to zero at $b = \beta$ with high probability. Indeed, if we define a *δ-solution* as any value b for which $|\,\gamma_n(b) - b\,| < \delta$, then the following corollary is a direct consequence of Theorem 1.

COROLLARY 1. *If conditions* (a)–(c) *of Theorem* 1 *are satisfied, then for* ***any*** *$\delta > 0$, β is a δ-solution of $\gamma_n(b) = b$ with probability tending to* 1 *as $n \to \infty$.* $\square$

The above is not sufficient to imply that all or any "solutions" of $\gamma_n(b) = b$, in the sense of (A) or (B) in Section 2.1, are consistent. To examine properties of these "solutions" we need to consider the behaviour of $\gamma_n(b)$ as $n \to \infty$ for $b \ne \beta$, and this requires assumptions about the explanatory variables. We consider this aspect further in the next section.

2.3 *Properties of* $\gamma_n(b) - b$. To investigate the behaviour of $\gamma_n(b) - b$, suppose that $x_1, \cdots, x_n$ behave as if they were realizations of a random variable with nondegenerate distribution function G, in the sense that the proportion of the x_i's not exceeding x tends to $G(x)$ for each x as $n \to \infty$. We also assume that G has bounded support, so that the partial residuals $r_i(b) = y_i - bx_i$ remain bounded above. Since we can write $r_i(b) = r_i(\beta) + (\beta - b)x_i$, the variables $r_i(b)$ arise from a convolution distribution

$$F_b(u) = \int F(u - (\beta - b)x)\, dG(x), \tag{4}$$

with corresponding survival function $S_b(u)$. Note that $F_\beta(u) = F(u)$ and $r_i(\beta) = r_i$ in the previous notation.

Under these assumptions the product limit estimator $\hat{F}_b$ estimates F_b. Let

$$\mu_i(b) = \mathbb{E}(r_i(b) \mid r_i(b) > t_i - bx_i),$$

where the conditional mean is calculated from the distribution F_b, and denote by $\hat{\mu}_i(b)$ the estimate of $\mu_i(b)$ based of $\hat{F}_b$. Define $U(b) = \sup\{u; F_b(u) < 1\} < \infty$ and let $\mathcal{N}_b(u)$ be the expected number of $e_i(b)$ exceeding u. If we now put

$$\psi_n(b) = \frac{\sum_{i=1}^n (x_i - \bar{x})S(c_i)(\theta_i(b) - \theta_i(\beta))}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

with

$$\theta_i(b) = \int_{t_i - bx_i}^{U(b)} \frac{S_b(u)\, du}{S_b(t_i - bx_i)},$$

we have the following result.

LEMMA 2. *If* $\mathcal{N}_b(u) \to \infty$ *as* $n \to \infty$ *for all* $u < U(b)$, *then*

$$| \gamma_n(b) - \beta - \psi_n(b) | \to_P 0. \tag{5}$$

PROOF. Note firstly that our assumptions about G ensure conditions (b) and (c) of Theorem 1 hold, and by Lemma 1, $\sup_i | \hat{\mu}_i(b) - \mu_i(b) | \to_P 0$. Now letting $r_i^* = r_i\delta_i + \mu_i(1 - \delta_i)$ as before, we may write

$$\begin{aligned}
&\gamma_n(b) - \beta - \psi_n(b) \\
&\quad = \frac{\sum_{i=1}^n (x_i - \bar{x})r_i^*}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})(\hat{\mu}_i(b) - \mu_i(b))(1 - \delta_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&\qquad + \frac{\sum_{i=1}^n (x_i - \bar{x})(\theta_i(b) - \theta_i(\beta))(1 - \delta_i - S(c_i))}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&\quad = \mathrm{I} + \mathrm{II} + \mathrm{III}, \quad \text{say.}
\end{aligned}$$

Terms I and II tend in probability to zero by arguments similar to those in the proof of Theorem 1. Further, since for each i, $\alpha + (\beta - b)e_x + bx_i \le t_i + \theta_i(b) \le U(b) + bx_i$, it follows that there exists a function $\rho(b) < \infty$ such that $|\theta_i(b) - \theta_i(\beta)| \le \rho(b)$ for all i. By Chebychev's inequality term III also tends in probability to zero. □

Lemma 2 shows that, with high probability, $\gamma_n(b) - b$ is approximated by $\psi_n(b) - b + \beta$ for n large, and the limiting behaviour of $\gamma_n(b)$ is determined by that of $\psi_n(b)$. Note that $\psi_n(\beta) = 0$. In order to investigate properties of $\psi_n(b)$ in a neighbourhood of β in Theorem 2 we first need the following lemma concerning the derivative of $\theta_i(b)$. For the remainder of this section we assume that F is differentiable.

LEMMA 3. *Suppose the density function f of the partial residuals r_i is bounded on $(-\infty, U)$. Then for each $i = 1, 2, \cdots, n$,*

$$\theta_i'(\beta) = (x_i - e_x)\left(1 - \frac{f(c_i)}{S^2(c_i)} \int_{c_i}^{U} S(u)\, du\right),$$

where e_x is the mean of G.

PROOF. It is straightforward to show that if f is bounded, we may interchange the order of integration and differentiation, giving

$$\frac{\partial}{\partial b} S_b(u) \bigg|_{b=\beta} = -f(u)e_x.$$

Further,

$$\theta_i'(b) = x_i + \int_{t_i - bx_i}^{U(b)} \left[\frac{\partial}{\partial b}\left(\frac{S_b(u)}{S_b(t_i - bx_i)}\right)\right] du$$

from which the result follows. □

THEOREM 2. *Let $\delta > 0$ be **any** value for which the conditions of Lemma 2 hold for all $b \in (\beta - \delta, \beta + \delta)$, and suppose f is bounded on $(-\infty, U)$. Suppose further that*

(a) $$\zeta = \limsup_{n\to\infty} \left\{\frac{|\sum_{i=1}^n (x_i - \bar{x})^2 S(c_i)|}{\sum_{i=1}^n (x_i - \bar{x})^2}\right\} < 1$$

(b) $$S(c_i)\,|\theta_i'(b) - \theta_i'(\beta)| \to 0 \quad \text{as} \quad b \to \beta \quad \text{uniformly in } i.$$

Then with probability tending to 1 as $n \to \infty$, there exists a point b in $(\beta - \delta, \beta + \delta)$ at which $\gamma_n(b) - b$ changes sign.

PROOF. From Lemma 3 and condition (a) we obtain

$$\limsup_{n\to\infty} \psi_n'(\beta) \le \zeta < 1.$$

By condition (b), for any $\varepsilon > 0$ we can find a $\delta^* \le \delta$ small enough so that if $|b - \beta| \le \delta^*$,

$$|\psi_n'(b) - \psi_n'(\beta)| \le \varepsilon \limsup_{n\to\infty} \frac{\sum_{i=1}^n |x_i - \bar{x}|}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Since the lim sup term is bounded, we deduce that $\limsup_{n\to\infty}\psi_n'(b) < 1$ for all b in some neighbourhood of β. It follows that there exist values $b_1 < \beta$, $b_2 > \beta$ and $\eta > 0$ such that $\liminf_{n\to\infty}(\psi_n(b_1) - b_1 + \beta) > \eta$ and $\limsup_{n\to\infty}(\psi_n(b_2) - b_2 + \beta) < -\eta$. Lemma 2 then gives the result. □

REMARK (i). Note that the term in braces in condition (a) of Theorem 2 is the expected proportion of the sum of squared deviations of explanatory variables which correspond to censored observations.

REMARK (ii). Condition (b) is clearly met if f is continuous and the t_i and x_i take only a finite number of different values. Moreover, since S is assumed to have a bounded derivative and G has bounded support, the following hold:

(iia) $S_b(t_i - bx_i) \to S(c_i)$ as $b \to \beta$ uniformly in i.

(iib) Letting $h_b(u) = \partial S_b(u)/\partial b$, $f_b(u) = -\partial S_b(u)/\partial u$ and assuming f is *uniformly* continuous, we have

$$h_b(t_i - bx_i) \to -e_x f(c_i) \quad \text{as} \quad b \to \beta$$

and

$$f_b(t_i - bx_i) \to f(c_i) \quad \text{as} \quad b \to \beta,$$

both uniformly in i.

From (iia) and (iib) it follows by expanding out the expression for $\theta_i'(b)$ that condition (b) of Theorem 2 will be satisfied if f is uniformly continuous and for all $c_i < U$, $S(c_i)/S_b(t_i - bx_i) \to 1$ uniformly in i as $b \to \beta$. In particular, this will be true if there exists an $m > 0$ such that there are no c_i in $(U - m, U)$, for then $S(c_i)$ is bounded away from 0 for $c_i < U$.

REMARK (iii). Theorem 2 states nothing about asymptotic uniqueness of a "solution" to $\gamma_n(b) = b$. Our experience with the method suggests, however, that such uniqueness holds under a wide range of conditions.

2.4 *Proof of Lemma* 1. Note that

$$|\hat{\mathbb{E}}_\beta(r_i \mid r_i > c) - \mathbb{E}(r_i \mid r_i > c)| = \left| \int_c^U \left(\frac{\hat{S}(u)}{\hat{S}(c)} - \frac{S(u)}{S(c)}\right) du \right| = |\phi_n(c)|, \quad \text{say.}$$

For any $\varepsilon > 0$, since the mean of the residuals is finite, we can choose $U^* < 0$ sufficiently small that $\int_{-\infty}^{U^*} (1 - S(u))\, du < \varepsilon/4$. Then for $c \le U^*$,

$$|\phi_n(c)| \le |A_c| + |B_c|$$

where

$$A_c = \int_c^{U^*} \left(1 - \frac{\hat{S}(u)}{\hat{S}(c)}\right) du - \int_c^{U^*} \left(1 - \frac{S(u)}{S(c)}\right) du$$

and

$$B_c = \int_{U^*}^{U} \left(\frac{\hat{S}(u)}{\hat{S}(c)} - \frac{S(u)}{S(c)}\right) du.$$

Now

$$\begin{aligned} |A_c| &\le \int_c^{U^*} \left(1 - \frac{\hat{S}(u)}{\hat{S}(c)}\right) du + \int_c^{U^*} \left(1 - \frac{S(u)}{S(c)}\right) du \\ &\le \int_c^{U^*} (1 - \hat{S}(u))\, du + \int_c^{U^*} (1 \dot{-} S(u))\, du \\ &\le \int_{-\infty}^{U^*} (1 - \hat{S}(u))\, du + \int_{-\infty}^{U^*} (1 - S(u))\, du. \end{aligned}$$

From Meier (1975),

$$\int_{-\infty}^{U^*} (1 - \hat{S}(u))\, du \to_P \int_{-\infty}^{U^*} (1 - S(u))\, du < \frac{\varepsilon}{4},$$

so $P(\sup_{c \le U^*} |A_c| < \varepsilon/2) \to 1$ as $n \to \infty$. Now write

$$\begin{aligned} |B_c| &= \left| \frac{1}{S(c)} \int_{U^*}^{U} (\hat{S}(u) - S(u))\, du + \left(\int_{U^*}^{U} \hat{S}(u)\, du\right)\left(\frac{1}{\hat{S}(c)} - \frac{1}{S(c)}\right) \right| \\ &\le \frac{1}{S(U^*)} \int_{U^*}^{U} |\hat{S}(u) - S(u)|\, du + \left(\int_{U^*}^{U} \hat{S}(u)\, du\right) \frac{|\hat{S}(c) - S(c)|}{\hat{S}(U^*)S(U^*)} \\ &\le \frac{\sup_{u<U} |\hat{S}(u) - S(u)|\,(U - U^*)}{S(U^*)} + \frac{\sup_{u \le U^*} |\hat{S}(u) - S(u)|\,(U - U^*)}{\hat{S}(U^*)S(U^*)}. \end{aligned}$$

Again from Meier (1975), $\sup_{u<U} |\hat{S}(u) - S(u)| \to_P 0$ so $\sup_{c \le U^*} |B_c| \to_P 0$. Since $\sup_{c \le U^*} |\phi_n(c)| \le \sup_{c \le U^*} |A_c| + \sup_{c \le U^*} |B_c|$, it follows that

$$P(\sup_{c \le U^*} |\phi_n(c)| < \varepsilon) \to 1 \quad \text{as} \quad n \to \infty. \tag{6}$$

Next, for any $\delta > 0$, use a similar argument to that for B_c to obtain

$$P(\sup_{U^* < c \le U - \delta} |\phi_n(c)| < \varepsilon) \to 1 \quad \text{as} \quad n \to \infty. \tag{7}$$

Finally,

$$\begin{aligned} \sup_{U-\delta<c<U} |\phi_n(c)| &\le \sup_{U-\delta<c<U} \left| \int_c^U \frac{\hat{S}(u)}{\hat{S}(c)}\, du + \int_c^U \frac{S(u)}{S(c)}\, du \right| \\ &\le 2\delta \quad \text{for all} \quad n. \end{aligned} \tag{8}$$

Lemma 1 now follows immediately from (6), (7) and (8) by noting ε is arbitrary and by choosing $\delta < \varepsilon/2$. □

2.5 *Estimation of* α. Buckley and James (1979) note that $\hat{\alpha}$ is the mean of the estimated distribution $\hat{F}_{\hat{\beta}}$. If assumptions regarding G are as in the previous section, then continuity of F_b in b ensures that $\hat{\alpha}$ will be consistent if $\hat{\beta}$ is. In practice the estimator of α tends to be biased downwards, due to the censoring and to the policy of assuming the largest observation is always uncensored.

3. Multiple linear regression. In the general multiple linear regression case with design matrix X the equations corresponding to (2) and (3) are

$$X^T(\hat{\mathbf{y}}(\hat{\beta}) - X\hat{\beta}) = \mathbf{0}$$

with $\hat{\mathbf{y}}(\mathbf{b})$ the vector of estimated responses $\hat{y}_i$ obtained by the obvious analogue of the simple regression case in Section 2.1. Standard linear regression programs are easily modified to implement the iterative scheme where at each step $\hat{\mathbf{y}}(\hat{\beta})$ is obtained from the previous estimate $\hat{\beta}$, then regarded as the response vector in the next iteration. As for the simple regression case the iterations may not converge but they provide a satisfactory way of obtaining approximate solutions. If $X^\top X$ has full rank, one attempts to solve

$$\hat{\beta} = (X^\top X)^{-1}X^\top\hat{\mathbf{y}}(\hat{\beta}) = \gamma_n(\hat{\beta}) \quad \text{say.}$$

Using the same methods as in the proof of Theorem 1, we then have

THEOREM 3. *Suppose that*

(a) $\mathscr{N}(u) \to \infty$ *as* $n \to \infty$ *for all* $u < U$,
(b) $\text{trace}(X^\top X)^{-1} \to 0$ *as* $n \to \infty$,
(c) $\limsup_{n\to\infty} X_+\mathbf{S} < \infty$, *where* X_+ *is the matrix of absolute values of* $(X^\top X)^{-1}X^\top$ *and* $\mathbf{S}^\top = (S(c_1), \cdots, S(c_n))$.

Then $\gamma_n(\beta) \to_P \beta$. □

4. Discussion. It is important that the conditions we have assumed in proving our consistency results place few restrictions on censoring patterns, since in practice a number of different random and nonrandom mechanisms produce the censor values t_i, and one could rarely assume a simple model for their distribution. For this reason we consider the t_i's as fixed and rely heavily on the product limit estimator results of Meier (1975). Földes and Rejtő (1981) also give asymptotic results for the product limit estimator which are relevant to our work, including rate of convergence results, but assuming that the censor variables are independent with continuous distribution functions.

The product limit estimator of F used in the Buckley-James method may be replaced by other distribution-free estimators such as those of Susarla and Van Ryzin (1976, 1978, 1980), Ferguson and Phadia (1979), Phadia (1980) or Rai, Susarla and Van Ryzin (1980). These have properties similar to the product limit estimator and one would expect the regression estimators derived from them to have the same asymptotic properties.

Extensive simulation studies we have carried out suggest that the conditions used to prove the consistency results in this paper are much more stringent than necessary. In particular, violation of the assumptions regarding explanatory variables in Section 2.3 appears not to be critical, but theoretical study of $\gamma_n(b) - b$ is difficult without some such simplifying assumptions. Further, suppose that the partial residuals r_i have support in $(-\infty, U)$, with U possibly infinite, and $\sup_i(t_i - \beta x_i) = t < U$. Then $\mathscr{N}(u) \nrightarrow \infty$ as $n \to \infty$ for $u > t$, but provided $\mathscr{N}(u) \to \infty$ for all $u < t$, the results of Section 2.2 remain valid with U replaced by t and F truncated at t. In this case the regression line being estimated is $y = \alpha^* + \beta x$ with $\alpha^* < \alpha$, so the intercept estimator will be biased downwards. However, this bias should not affect the slope estimator. Conditions for Theorem 2 are violated by the truncated distribution function, and the consistency results in this case will be considered elsewhere.

Finally, it is worth noting that the estimation method of Buckley and James adapts easily in principle to weighted and nonlinear regressions, although the technical problems then encountered have yet to be studied.

Acknowledgements. We thank the referees for comments and suggestions leading to substantial improvements in the presentation of this work.

REFERENCES

BRESLOW, N. and CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–453.

BUCKLEY, J. and JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.

FERGUSON, T. S. and PHADIA, E. G. (1979). Bayesian non-parametric estimation based on censored data. *Ann. Statist.* **7** 163–186.

FÖLDES, A. and REJTŐ, L. (1981). Asymptotic properties of the nonparametric survival curve estimators under variable censoring. In *The First Pannonian Symposium on Mathematical Statistics* (Eds. P. Révész, L. Schmetterer and V. M. Zolotarev). *Springer Lecture Notes in Math* **8** 52–54. New York.

GILL, R. D. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

KOUL, H., SUSARLA, V. and VAN RYZIN, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9** 1276–1288.

MAURO, D. W. (1983). A note on the consistency of Kaplan-Meier least squares estimators. *Biometrika* **70** 534–535.

MEIER, P. (1975). Estimation of a distribution function from incomplete observations. In *Perspectives in Probability and Statistics. Papers in Honour of M. S. Bartlett* 67–82. (Ed. J. Gani). Academic, New York.

MILLER, R. G. (1976). Least squares regression with censored data. *Biometrika* **63** 449–464.

MILLER, R. G. and HALPERN, J. (1982). Regression with censored data. *Biometrika* **69** 521–533.

PHADIA, E. G. (1980). A note on empirical Bayes estimation of a distribution function based on censored data. *Ann. Statist.* **8** 226–229.

RAI, K., SUSARLA, V. and Van Ryzin, J. (1980). Shrinkage estimation in non-parametric Bayesian analysis: a simulation study. *Commun. Statist.-Simul. Comp* **B9** 271–298.

SUSARLA, V. and VAN RYZIN, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Statist. Assoc.* **71** 897–902.

SUSARLA, V. and VAN RYZIN, J. (1978). Empirical Bayes estimation of a distribution (survival) function from right censored observations. *Ann. Statist.* **6** 740–754.

SUSARLA, V. and VAN RYZIN, J. (1980). Large sample theory for an estimator of the mean survival time from censored samples. *Ann. Statist.* **8** 1002–1016.

YANG, G. (1977). Life expectancy under random censorship. *Stoch. Process. Appl.* **6** 33–39.

DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF WESTERN AUSTRALIA
NEDLANDS, WESTERN AUSTRALIA 6009
AUSTRALIA