Biometrika (1987), **74**, 2, *pp.* 301-9
Printed in Great Britain

Linear models, random censoring and synthetic data

BY SUE LEURGANS
Department of Statistics, Ohio State University, Columbus, Ohio 43210-1247, U.S.A.

SUMMARY

Estimators for the linear model in the presence of censoring are available. A new extension of the least-squares estimator to censored data is equivalent to applying the ordinary least-squares estimator to synthetic times, time constructed by magnifying the gaps between successive order statistics. Under suitable regularity conditions, the synthetic data estimator is Fisher consistent and asymptotically normal. Examples facilitate comparison of the synthetic data estimator with estimators proposed by Buckley & James (1979) and by Koul, Susarla & Van Ryzin (1981).

Some key words: Censored data; Least-squares estimation; Linear model; Two-sample estimator.

1. INTRODUCTION

The entire lifetime of a person, a machine, a plant or an animal is not always observable. Some lifetimes may be censored, in that only a lower bound for the lifetime is recorded. Statisticians are often interested in modelling the distribution of the true lifetimes as a function of covariates. When censoring is present, a proportional hazards model is often used. When there is no censoring, linear models are often used for ad hoc modelling of dependent variables. Miller & Halpern (1982) survey the methods that have been proposed for the linear model in the presence of censoring. This paper describes a closed-form method which is consistent and asymptotically normal. This method incorporates censoring in a natural way.

Section 2 formulates classical least-squares estimation in a manner that generalizes to accommodate random right censoring. Section 3 gives some theoretical results for the simplest nontrivial linear model, the two sample case. Section 4 contains examples and the final section discusses the role of this estimator.

The covariate vector of the ith person will be denoted by x_i, and usually includes one as the first component. The true lifetime Y_i follows a linear model if its conditional distribution given x_i is

$$F_i(y) = F(y - \beta_0^{\mathrm{T}} x_i),$$

for some fixed distribution function and some parameter vector β_0. If β_0 includes an intercept term, F must have a known mean, which can be taken to be zero. Since the linear model may be applied to some function of time, such as the logarithm, it is not appropriate to assume that the times are nonnegative. The censoring will be modelled as a random variable C_i with the distribution H_i and independent of Y_i. The observed lifetime is $T_i = \min(C_i, Y_i)$. Censoring and death are distinguishable, so that every observed time can be classified as either a censoring time or a death time. The variance σ^2 of F is finite. The notation $(x_{(i)}, T_{(i)})$ refers to the ith order statistic of the T's and the corresponding vector x.

2. LEAST-SQUARES ESTIMATION

Following Miller (1976), least-squares estimation can be formulated in terms of the mean squared error of G_β, the average distribution of the residuals $Y-\beta^{\mathrm{T}}x$, defined by

$$G_\beta(y)=\sum_{i=1}^{n} n^{-1}F_i(y+\beta^{\mathrm{T}}x_i)=\sum_{i=1}^{n} n^{-1}F\{y+(\beta-\beta_0)^{\mathrm{T}}x_i\}.$$

The mean squared error of this distribution is

$$Q(\beta)=\sigma^2+(\beta-\beta_0)^{\mathrm{T}}\Big(\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}\Big)(\beta-\beta_0),$$

a function which has a single minimum at β_0 whenever the matrix sum is positive-definite. Classical least-squares estimation can be thought of as estimating G_β by

$$\hat{G}_\beta(y)=\sum_{i=1}^{n} I(Y_i\leqslant y+\beta^{\mathrm{T}}x_i)n^{-1},$$

where $I(A)$ denotes the indicator of the event A, and then finding that β which minimizes $\hat{Q}(\beta)$, mean squared error of $\hat{G}_\beta(y)$. The function $\hat{Q}(\beta)$ reduces to

$$\hat{Q}(\beta)=\sum_{i=1}^{n} (Y_i-\beta^{\mathrm{T}}x_i)^2 n^{-1},$$

which is a constant multiple of the residual sum of squares.

In the presence of random censoring, it is still possible to estimate G_β. If the censoring time is independent of the lifetime,

$$1-G_\beta(t)=1-\sum_{i=1}^{n} n^{-1}F_i(t+\beta^{\mathrm{T}}x_i)=\sum_{i=1}^{n}\frac{\mathrm{pr}\,(T_{(i)}\geqslant t+\beta^{\mathrm{T}}x_{(i)})}{n\{1-H_i(t+\beta^{\mathrm{T}}x_{(i)})\}}.$$

This equation can be used for estimation whenever there are plausible assumptions about the way in which the censoring distributions H_i depend on x. In this subsection, the H_i are assumed not to depend on x, so that $H_i=H$, say. An extension to other censoring schemes is illustrated for the heart transplant data in § 4. When all H_i are the same,

$$1-G_\beta(t)=\sum_{i=1}^{n} g_i(t+\beta^{\mathrm{T}}x_{(i)})n^{-1}, \tag{2·1}$$

where $g_i(v)=\mathrm{pr}\,(T_{(i)}\geqslant v)/\{1-H(v)\}$. The fact that G_β is the average of translates of functions, where only the sizes of the translations and not the functions g_i depend on β, is at the root of the closed-form expression.

A natural estimator for g_i is

$$\hat{g}_i(v)=I(T_{(i)}\geqslant v)\{1-\hat{H}(v)\}^{-1},$$

where $\hat{H}(v)$ is an estimator of H. This estimator is a naive estimator of $1-F_{(i)}(v)$, because it increases from 1 on $(-\infty, T_i]$, is zero on (T_i,∞), and is slightly biased in general. If $1-\hat{H}$ is a left-continuous step function which is constant between event times, the corresponding step functions $\hat{g}_i$ can be written in terms of w_j, the sizes of the jumps up of $1/(1-\hat{H})$, and W_i, the size of the jump down:

$$\hat{g}_i(v)=1+\sum_{j=1}^{i-1} w_j I(T_{(j)}\leqslant v)-W_i I(T_{(i)}<v),$$

where $w_j = 1/\{1-\hat{H}(T_{(j)}+)\} - 1/\{1-\hat{H}(T_{(j)})\}$ and $W_i = 1/\{1-\hat{H}(T_{(i)})\}$. Substitution in (2·1) gives

$$1-\hat{G}_\beta(t) = \sum_{i=1}^{n} n^{-1}\{W_i I(T_{(i)} \geqslant t+\beta^{\mathrm{T}}x_{(i)}) - \sum_{j=1}^{i-1} w_j I(T_{(j)} > t+\beta^{\mathrm{T}}x_{(i)})\}.$$

The corresponding mean squared error is

$$\hat{Q}(\beta) = \sum_{i=1}^{n}\left\{(T_{(i)}-\beta^{\mathrm{T}}x_{(i)})^2 W_i - \sum_{j=1}^{i-1} w_j (T_{(j)}-\beta^{\mathrm{T}}x_{(i)})^2\right\} n^{-1}.$$

With the convention that $T_{(0)}=0$, substitution for w_j in terms of W and algebraic manipulation show that $n\hat{Q}(\beta)$ can be rewritten as

$$\sum_{i=1}^{n}\left\{\sum_{j=1}^{i} W_j (T_{(j)}^2 - T_{(j-1)}^2) - 2\beta^{\mathrm{T}}x_{(i)} \sum_{j=1}^{i} (T_{(j)}-T_{(j-1)}) W_j + (\beta^{\mathrm{T}}x_{(i)})^2\right\}.$$

Letting $Z_{(i)}$ denote half the linear coefficient of $\beta^{\mathrm{T}}x_{(i)}$, and completing the square, one sees that

$$\hat{Q}(\beta) = \hat{Q}_z(\beta) + \hat{Q}_0, \tag{2·2}$$

where

$$\begin{gathered}\hat{Q}_z(\beta) = \sum_{i=1}^{n} n^{-1}(Z_i - \beta^{\mathrm{T}}x_i)^2, \quad Z_{(i)} = \sum_{j=1}^{i} (T_{(j)}-T_{(j-1)}) W_j, \\ \hat{Q}_0 = \sum_{j=1}^{n} \{(n-j+1) W_j (T_{(j)}^2 - T_{(j-1)}^2) - Z_{(j)}^2\} n^{-1}.\end{gathered} \tag{2·3}$$

Since $\hat{Q}_0$ is free of $\hat{\beta}$, the decomposition (2·2) implies that $\hat{Q}$ is minimized by the β which minimizes $\hat{Q}_z$. Thus $\hat{\beta}$ can be obtained by replacing the observed times T_i by the synthesized times Z_i and then applying ordinary least-squares, without regard to censoring.

The synthetic times have a simple relation to the observed times. From the definition (2·3), the difference between two consecutive synthetic times is the product of the difference of the two corresponding observed times and a weight W_i. Given the ordered event times and the censoring times, synthesis is a linear operation and synthesis is therefore equivariant under increasing linear transformation of the time variable. Since ordinary least-squares estimation is also a linear operation, the synthetic data estimator transforms suitably when the times are replaced by their images under an increasing linear transformation. Decreasing linear transformations should not be used, unless the original data were left-censored.

Since the weight is the reciprocal of the estimated survival function of the censoring times, the weight is always greater than or equal to 1. The weight W_i will be large if the corresponding time $T_{(i)}$ exceeds many censored times. Therefore synthesis requires magnifying the gaps between consecutive observed times and then reaccumulating to form the consecutive synthetic times. Since the step function $1/\{1-\hat{H}(t)\}$ is equal to W_i on the interval $(T_{(i-1)}, T_{(i)}]$, the second formula (2·3) implies

$$\begin{aligned} Z_{(i)} &= T_{(1)} + \sum_{j=2}^{i} W_j (T_{(j)} - T_{(j-1)}) = T_{(1)} + \int_{T_{(1)}}^{T_{(i)}} \frac{1}{1-\hat{H}(s)}\, ds \\ &= \int \left\{\frac{I(s<T_{(i)})}{1-\hat{H}(s)} - I(s<0)\right\} ds = Z_n(T_{(i)}). \end{aligned} \tag{2·4}$$

If no times are censored, all the weights W_i are 1, the synthetic times equal the original times, and the synthetic data estimator reduces to the usual least-squares estimator.

3. THE DIFFERENCE BETWEEN TWO MEANS

3·1. *Notation and formulae*

The covariate x for group 1 will be $(1, 0)^T$ and that for group 2 will be $(0, 1)^T$. Thus the coordinates of β are the two group means. The numbers in each group are n_1 and n_2. Summation over the individuals of sample j will be denoted by Σ_j $(j = 1, 2)$. The numbers of deaths and losses in each sample up to time t are counted by four empirical subdistribution functions:

$$F_{jD}(t) = \sum_j \frac{I(T_i \leq k) I(T_i = Y_i)}{n_j}, \quad F_{jC}(t) = \sum_j \frac{I(T_i \leq t) I(T_i = C_i)}{n_j} \quad (j = 1, 2).$$

The empirical distribution of observed times in sample j is

$$F_{j+}(t) = F_{jD}(t) + F_{jC}(t) \quad (j = 1, 2).$$

The least-squares estimator of a population mean is the average of the corresponding sample. Therefore the synthetic estimator of a population mean is the average of the corresponding synthetic sample:

$$\hat{\beta}_j = \sum_j Z_i / n_j = \int Z_n(t)\, dF_{j+}(t) = \int_{T_{(1)}}^{\infty} \left(\int_{T_{(1)}}^{t} \frac{ds}{1 - \hat{H}(s)} + T_{(1)} \right) dF_{j+}(t),$$

where the last equation follows from (2·4). Interchanging the integrals implies that

$$\hat{\beta}_j = \int \left\{ \frac{1 - F_{j+}(s)}{1 - \hat{H}(s)} - I(s \leq 0) \right\} ds. \tag{3·1}$$

Although $\hat{H}$ may be undefined on $(T_{(n)}, \infty)$, no ambiguity arises because F_{j+} is one on that interval. If the empirical distribution functions above are replaced by their theoretical limits, the resulting expression reduces to (3·2) whenever $F_j(s) < 1$ implies $H(s) < 1$:

$$\int \{1 - F_j(s) - I(s \leq 0)\}\, ds = \int s\, dF_j(s) = \beta_j. \tag{3·2}$$

The first expression in (3·2) is the familiar expression for the expected value of a random variable in terms of its distribution function. The second equation holds whether or not the second distribution function is a translate of the first, so that the synthetic estimators are Fisher consistent whenever censoring is not certain for long lifetimes.

An estimator proposed by Koul et al. (1981) can also be written as a functional of the same subdistribution functions. Ignoring their truncations M_n, their estimator reduces to

$$\tilde{\beta}_j = \int t\{1 - \hat{H}(t)\}^{-1}\, dF_{jD}(t);$$

clearly $\tilde{\beta}_j$ is also Fisher consistent under the same conditions. The example of § 4·1 shows that $\hat{\beta}$ and $\tilde{\beta}$ are different functionals of the data. If pseudo-data are defined by $P_i = I(T_i = Y_i) W_i T_i$, then $\tilde{\beta}_j = \Sigma_j (P_i / n_j)$. Koul et al. (1981) use variants of the product-limit estimator to determine the weights W_i. The pseudo-data estimator therefore differs slightly from their estimators.

3·2. *Theoretical properties*

In this section, conditions are stated for the asymptotic normality of the synthetic data estimator in the two-sample case. The consistency of the estimator is then automatic. A variance estimator is indicated. The regularity conditions are stated in terms of the behaviour of $1-\hat{F}_P$, the product-limit estimator applied to the pooled sample, and the corresponding limiting survival function $1-F_P$.

THEOREM. *If* $\lim(n_j/n)=p_j$, *if* F_j *and* H *are continuous, and if*

$$n^{\frac{1}{2}}\left[\int\{1-\hat{F}_P(s)-I(s<0)\}\,ds-\int\{1-F_P(s)-I(s<0)\}\,ds\right]$$

converges weakly to a normal distribution with mean zero and finite variance, then $n^{\frac{1}{2}}(\hat{\beta}-\beta)$ *converges weakly to*

$$\begin{bmatrix}\int U_1(s)\{1-F_1(s)\}\,ds\\ \int U_2(s)\{1-F_2(s)\}\,ds\end{bmatrix},$$

where U_1 *and* U_2 *are time-changed Wiener processes with covariance structure given by*

$$E\{U_j(s)\,U_j(t)\}=\frac{1-\{1-F_j(s)\}\{1-H(s)\}}{p_j\{1-F_j(s)\}\{1-H(s)\}}-\tau_P(s),$$

$$E\{U_1(s)\,U_2(t)\}=E\{U_1(t)\,U_2(s)\}=-\tau_P(s)\quad(s\leqslant t),$$

where τ_P *is the limiting covariance of the estimated cumulative hazard of censoring in the presence of death distribution* F_P *and is therefore given by*

$$\tau_P(t)=\int^t\frac{h(s)}{\{1-F_P(s)\}\{1-H(s)\}^2}\,ds.$$

The proof of the theorem uses the integral expression (2·4) to write the estimators as stochastic integrals with respect to martingales obtained from counting processes. The techniques of Gill (1980, Ch. 2) then apply. Some details are indicated in an unpublished technical report available from the author.

The variance of the limiting distribution of $n^{\frac{1}{2}}\{(\hat{\beta}_1-\hat{\beta}_2)-(\beta_1-\beta_2)\}$ follows from the theorem. The general expression is

$$V=\sum_{j=1}^{2}-\frac{1}{p_j}\int\left[\frac{I_j(t)}{\{1-F_j(t)\}\{1-H(t)\}}\right]^2\{1-H(t)\}\,dF_j(t)$$
$$+\int\left(I_1(t)\left[\frac{p_2\{1-F_2(t)\}}{p_1\{1-F_1(t)\}}\right]^{\frac{1}{2}}+I_2(t)\left[\frac{p_1\{1-F_1(t)\}}{p_2\{1-F_2(t)\}}\right]^{\frac{1}{2}}\right)^2 d\tau_P(t),\qquad(3\cdot3)$$

where

$$I_J(t)=\int_t^\infty\{1-F_j(s)\}\,ds\quad(j=1,2).$$

This variance cannot be factorized into a term depending on the design, that is, on p_1, and the factor σ^2, the underlying conditional variance of the true times Y, unless censoring is absent or unless $F_1=F_2$. When $F_1=F_2=F$, V reduces to $v(1/p_1+1/p_2)$, where

$$v=\int\left[\frac{I(t)}{\{1-F(t)\}\{1-H(t)\}}\right]^2-d[\{1-F(t)\}\{1-H(t)\}].$$

The integrals in the summation in (3·3) are the asymptotic variance of the mean of a one-sample product-limit estimator (Reid, 1981, p. 87). Since the other term is nonnegative, V is greater than or equal to the asymptotic variance of the difference of two separate sample estimators. Presumably this inequality is due to the ad hoc nature of this estimator.

Empirical estimation of V and v is clearly possible. The empirical estimator

$$\hat{v} = \sum_{i=1}^{n} \left\{ \frac{1-\hat{F}(T_{(i)})}{(n-i+1)/n} \right\}^2 n^{-1}$$

can be used to provide tests of equality of F_1 and F_2 with approximately the desired level in large samples.

4. EXAMPLES

Synthetic data estimates for two sets of data are described in this section. The two-sample data of Freireich, Gehan & Frei (1963) and the Stanford heart transplant data have been used to illustrate many censored data procedures. The two-sample data is small enough to compare the synthetic times and the pseudo-times. The heart transplant data applies synthesis when censoring may vary with covariates

Example 1: Two small samples. Freireich et al. (1963) gave the survival times in weeks of 21 terminal cancer patients treated with a drug regimen as 6, 6, 6, 6+, 7, 9+, 10, 10+, 11+, 13, 16, 17+, 19+, 20+, 22, 23, 25+, 32+, 32+, 34+ and 35+, where 6+ denotes a patient censored at 6 weeks. Survival times of 21 untreated patients were 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22 and 23. The assumption of equal random censoring is unreasonable. Miller & Halpern (1982) point out that the pseudo-data method is sensitive to different censoring patterns. This example will demonstrate the effects of unequal censoring.

Most of the estimates reported in Table 1 suggest that treated people live roughly 3 times as long as controls. However, the estimator of Koul et al. (1981) implies that untreated patients tend to live longer than treated patients, a conclusion not supported by the data. In Fig. 1, five values are connected for each individual on the drug regimen: an estimated log time for the Buckley & James (1979) algorithm, the observed log time, the synthetic log time, and the pseudo log time. The Buckley & James (1979) algorithm oscillates between two close values because the relative order of the residual from log (5) in the control group and from log (19+) in the drug group oscillates. The large number of pseudo-times equal to zero lowers the estimated average log time for the drug regimen pseudo-times. The iterated Buckley & James (1979) estimated log times are close to the synthetic times.

Table 1. *Estimated location parameters for two groups, data from Freireich et al.* (1963)

Method	Control	Treatment	Difference
Synthetic data (raw times)	9·22	21·232	12·01
Synthetic data (log times)	1·866	2·855	0·989
Koul et. al.	2·124	1·233	−0·891
Buckley & James 1-step from (0, 0)	1·825	2·909	1·084
	1·825	3·159	1·334
Buckley & James stable points	1·825	3·161	1·337

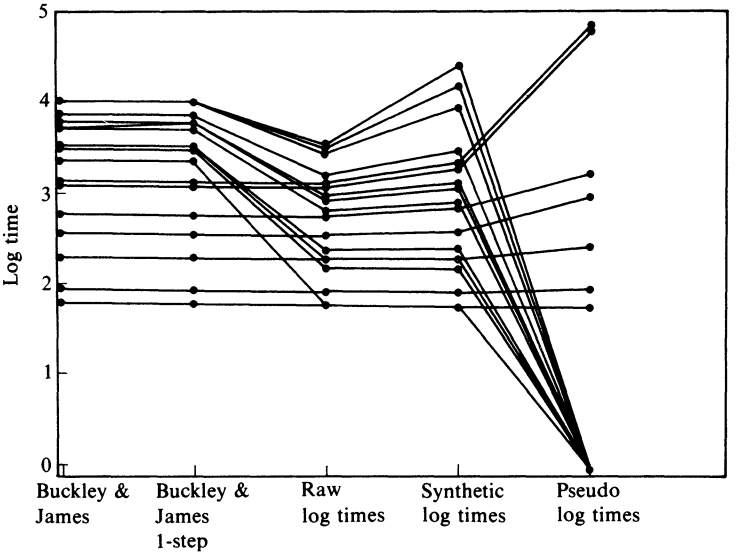

Fig. 1. Artificial log times for Freireich et al. (1963) data for the treated group. For each subject, lines connect Buckley & James (1979) estimated log times, one-step Buckley & James (1979) log times, raw log times, synthetic log times, and pseudo log times.

Example 2: Stanford heart transplant. Data from the Stanford heart transplant program has also been used to illustrate many regression methods for censored data. Miller & Halpern (1982) give survival times of the first 184 patients, censored at February, 1980. They analyse the dependence of the base 10 logarithm on age at transplant and on a mismatch score. Table 2 compares coefficients reported by Miller & Halpern (1982) with two synthetic data fits.

Two methods of synthesis were used. Pooled synthesis refers to the method described above. Common random censorship may not be plausible, because the selection criteria have gradually changed, and now favour younger subjects. If random censoring is assumed, the censoring distribution should vary with age. One crude way to permit different censoring distributions for different ages is to synthesize separately within age groups. Group synthesis will refer to log times synthesized separately for four age groups: less than 30, 30–39, 40–49, and 50 or older. The groups contain 30, 23, 66 and 38 people, respectively. The survival time of one person was recoded from 0 to 1.

The leftmost three columns of Table 2 report estimates for the model with covariates age and mismatch scores, including fits from Miller & Halpern (1982). Since 27 mismatch scores are missing, Table 2 is based on 157 cases. Only these cases were used in the synthesis. The coefficients in the rightmost three columns of Table 2 are for fits with covariates age and age^2. Following Miller & Halpern (1982), the 5 patients with survival time less than 10 days are omitted in the age analyses.

The only coefficient sensitive to the synthesis method is the intercept in the quadratic model. All the other synthesis estimates are close to the Buckley & James estimates. The intercept in the regression of synthetic time is sensitive to a few large synthesized times. The two largest pooled synthetic times are 409 and 3130 years, far exceeding documented human lifespans. For grouped synthesis, the corresponding synthetic times are 181 and 555 years. While still large, the anomaly is less pronounced, in part because the magnification factors $1/(1-\hat{H})$ are bounded above by the sample size, which is necessarily smaller for the subgroups.

Table 2. *Estimated coefficients for prediction of* $\log_{10}$ *(lifetime) of Stanford heart transplant recipients. Proportional hazards, Buckley & James, Miller and Koul et al. coefficients reported from Miller & Halpern* (1982, *Tables* 2, 3)

Method	Intercept	Age	Mismatch score	Intercept	Age	Age^2
Proportional hazards	0·30	0·167		−1·46	0·0023	
Buckley & James	3·23	−0·015	−0·003	1·35	0·107	−0·0017
Miller (b)	2·54	0	0·040			
Koul, Susarla & Van Ryzin	0·72	0·024	0·251			
Pooled synthesis	3·03	−0·008	−0·091	2·981	0·103	−0·0015
Grouped synthesis	3·08	−0·10	−0·072	1·494	0·089	−0·0014

5. DISCUSSION

Two numerical properties distinguish the synthetic data estimators from the consistent estimators of Koul et al. (1981) and of Buckley & James (1979). The Buckley & James (1979) algorithm estimates a true lifetime for each censored lifetime, then estimates the parameters, and repeats these steps until convergence. Unlike the estimated lifetimes used by the iterative algorithm, the pseudo-times and synthetic times are constructed independently of the covariates. Consequently, synthesis need not be repeated even if the covariates are changed. The synthetic data estimator is the only estimator discussed whose artificial lifetimes are in the same order as the observed lifetimes.

The synthetic data estimator is based on the assumption that the censoring mechanism can be modelled by random times, but is less sensitive to the assumptions of the linear model. When all the covariates are group indicators, the synthetic data estimator will be consistent even if the shape of the error distribution differs across group. The Buckley & James (1979) algorithm has complementary properties: the censoring mechanism does not affect Fisher consistency in the two-sample case, but the distributions within group must all be translates of each other. Presumably the relative validity of the censoring and the translation assumptions will vary in applications. Neither method appears to be efficient relative to parametric maximum likelihood methods (Kalbfleisch & Prentice, 1980, § 3.6).

Since the largest times will be increased the most, the impact of positive residuals will be magnified, exaggerating the lack of robustness of the linear model. The approach outlined here can be applied to functionals of $\hat{G}_\beta$ other than Q; M-estimators for censored data will be described elsewhere.

Several questions should be answered before synthetic data estimators enter routine use. Variance expressions and estimators need further study. Bootstrap methods could be used (Efron, 1982). Formulae for the influence of individual points in the presence of censoring would be useful.

The plausibility of the statistical model does not guarantee the applicability of the linear model to lifetimes. Linear models may be most successful if lifetimes are determined by a static characteristic, rather than by dynamic response to sudden shocks.

ACKNOWLEDGEMENTS

This research was supported in part by National Institute of Health grants at the University of Wisconsin and the University of Washington. Computations were performed on a VAX 11/750 at the University of Washington.

References

BUCKLEY, J. & JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-36.
EFRON, B. (1982). *The Jackknife, the Bootstrap, and other Resampling Plans.* Philadelphia: S.I.A.M.
FREIREICH, E. O., GEHAN, E. & FREI, E. (1963). The effect of 6-mercaptopmine on the duration of steroid induced remission in acute leukemia. *Blood* **21**, 699-716.
GILL, R. D. (1980). *Censoring and Stochastic Integrals*, Mathematics Center Tract 124. Amsterdam: Mathematische Centrum.
KALBFLEISCH, J. D. & PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.
KOUL, H., SUSARLA, V. & VAN RYZIN, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9**, 1276-88.
MILLER, R. G. (1976). Least squares regression with censored data. *Biometrika* **63**, 449-64.
MILLER, R. G. & HALPERN, J. (1982). Regression with censored data. *Biometrika* **69**, 521-31.
REID, N. (1981). Influence functions for censored data. *Ann. Statist.* **9**, 78-92.

[*Received August* 1984. *Revised June* 1986]