The Annals of Statistics
1982, Vol. 10, No. 1, 101–113

PIECEWISE EXPONENTIAL MODELS FOR SURVIVAL DATA WITH COVARIATES[1]

BY MICHAEL FRIEDMAN

Rutgers University

A general class of models for analysis of censored survival data with covariates is considered. If n individuals are observed over a time period divided into $I(n)$ intervals, it is assumed that $\lambda_j(t)$, the hazard rate function of the time to failure of the individual j, is constant and equal to $\lambda_{ij} > 0$ on the ith interval, and that the vector $\ell = \{\log \lambda_{ij} : j = 1, \cdots, n; i = 1, \cdots, I(n)\}$ lies in a linear subspace. The maximum likelihood estimate $\hat{\ell}$ of ℓ provides a simultaneous estimate of the underlying hazard rate function, and of the effects of the covariates. Maximum likelihood equations and conditions for existence of $\hat{\ell}$ are given.

The asymptotic properties of linear functionals of $\hat{\ell}$ are studied in the general case where the true hazard rate function $\lambda_0(t)$ is not a step function, and $I(n)$ increases without bound as the maximum interval length decreases. In comparison with recent work on regression analysis of survival data, the asymptotic results are obtained under more relaxed conditions on the regression variables.

1. Introduction. In recent years, much attention has been devoted to medical survival studies in which the data on the jth individual include the observed survival time t_j, which may be terminated either by a failure or by censoring, and a vector of covariates $\mathbf{x}_j$.

For a study with one covariate, Feigl and Zelen (1965) proposed an exponential survival model in which the time to failure of the jth individual has the density

$$f_j(t) = \lambda_j \exp(-\lambda_j t), \qquad \lambda_j^{-1} = \alpha \exp(\beta x_j), \tag{1.1}$$

where α and β are unknown parameters.

In a groundbreaking paper, Cox (1972) offered a survival model in which the hazard function of the time to failure of the jth individual is expressed as

$$\lambda_j(t) = \lambda_0(t) \exp\{(\boldsymbol{\beta}, \mathbf{x}_j)\}, \tag{1.2}$$

where $\lambda_0(t)$ is an underlying hazard rate function not restricted by any assumptions. The partial likelihood function proposed for the analysis of the model (1.2), see Cox (1972) and Cox (1975), does not involve the function $\lambda_0(t)$, and allows maximum likelihood estimation of the regression coefficients $\boldsymbol{\beta}$. The model (1.2) has been further elaborated by Breslow (1974) and Kalbfleisch and Prentice (1973). An extensive discussion and bibliography appear in Kalbfleisch and Prentice (1979) and Prentice and Kalbfleisch (1979).

In (1.2), a log-linear model is in effect used to describe the effects of the covariates upon the individual hazard rates. In the piecewise exponential approach, a log-linear model is used to model both the effects of the covariates and the underlying hazard rate function, which is approximated by a step function. Maximum likelihood estimates of the underlying hazard rate function are then obtained simultaneously with the estimates of the regression parameters expressing the effect of the covariates. This approach was first studied by

Received July 1979; revised August 1981.

[1] Support for this research was provided in part by NSF Grant No. SOC76-80389.

AMS subject classifications. Primary 62E20; secondary 62F10.

Key words and phrases. Asymptotic theory, censored data, log-linear model, maximum likelihood estimation, piecewise exponential model, survival data.

Holford (1976), and is also the subject of work by Holford (1980) and Laird and Olivier (1981).

The class of piecewise exponential models is defined in Section 2, and conditions for the existence of maximum likelihood estimates (MLE's) are explored. As pointed out by Friedman (1978), there is a similarity between the likelihood function for the piecewise exponential model and the likelihood function for a log-linear model for frequency data. As a result of this similarity, known results concerning the existence of MLE's for log-linear models in contingency table analysis can be applied to the present problem. This similarity has also been the focus of recent work by Holford (1980) and Laird and Olivier (1981), who have shown that in certain cases Iterative Proportional Fitting may be used to obtain the MLE's of the piecewise exponential model, thus further bringing the analysis of survival data within the scope of the familiar techniques used to analyze frequency data.

Section 3 presents results on the asymptotic convergence and normality of maximum likelihood estimates. These results are obtained under very general conditions, namely, when the number of unknown parameters in the piecewise exponential model increases without bound as n goes to infinity, as a result of the shrinking of the intervals on which the step function approximating $\lambda_0(t)$ is defined. The fixed-point methods developed by Haberman (1977) in his study of exponential response models are adopted to obtain these results. There are basically three kinds of conditions that need to be imposed in order to prove the desired asymptotic results: conditions on the distribution of the covariates, conditions on the size and number of the intervals, and conditions on the shape of the underlying hazard function. In contrast to other recent work on the asymptotic behavior of regression parameter estimators in survival studies (for a summary see Kalbfleisch and Prentice, 1979), the conditions imposed on the covariates are relaxed; it is not required that the values of the covariates be bounded.

A brief numerical example is discussed in Section 4.

2. The piecewise exponential model: basic properties and maximum likelihood estimation. Suppose that the survival times $\{t_j : j \in \bar{n}\}$, where $\bar{n}$ is the set of integers from 1 to n, are observed. The observed survival times may be terminated either by failure or by censoring (withdrawal). It is assumed that conditionally on **x** the times to failure are independent of the times to withdrawal. Let the time scale be divided into $I(n)$ intervals $(0, T_1]$, $(T_1, T_2]$, $\cdots$, $(T_{I(n)-1}, T_{I(n)}]$. In many applications, the last interval may be considered infinite in length. However, in proving the asymptotic results of Section 3, a fixed finite value independent of n is assigned to $T_{I(n)}$.

2.1. *Basic properties.* Under the piecewise exponential model, the times to failure satisfy the following two assumptions:

(1) The hazard rate function of each individual is constant over any given interval. The hazard rate of the jth individual in the ith interval is denoted by λ_{ij}, and it is assumed that $\lambda_{ij} > 0$ for each (i, j).

(2) *If* ℓ is the vector with components $\ell_{ij} = \log(\lambda_{ij})$, then $\ell \in \Omega(n)$ for a given linear subspace $\Omega(n)$ of $R^{nI(n)}$.

The likelihood function can be expressed in terms of the statistics $\{t_{ij}, I_{ij} : j \in \bar{n}, i \in \overline{I(n)}\}$ defined as follows:

$$I_{ij} = 1 \quad \text{if the } j\text{th individual fails during the } i\text{th interval} \qquad (2.1)$$
$$= 0 \quad \text{otherwise,}$$

and

$$t_{ij} = \max\{0, \min(T_i - T_{i-1}, t_j - T_{i-1})\}. \qquad (2.2)$$

The log likelihood function is then

$$L(\ell) = \sum_{i,j} \ell_{ij} I_{ij} - \sum_{i,j} t_{ij} \exp(\ell_{ij}). \tag{2.3}$$

EXAMPLE 2.1. Consider a survival study in which each individual is associated with a K-dimensional covariate vector

$$\mathbf{x}_j = (x_j^1, \cdots, x_j^K),$$

and $\Omega(n) = \text{span}\{\mathbf{u}_m : m = 1, \cdots, I^*(n) + K\}$, where $I^*(n) \leq I(n)$, and

$$\begin{aligned} u_{mij} &= u_{mi}, & m &\leq I^*(n), \\ u_{mij} &= x_j^k, & m &= I^*(n) + k. \end{aligned} \tag{2.4}$$

Each vector ℓ in $\Omega(n)$ then has coordinates

$$\ell_{ij} = \sum_{m=1}^{I^*(n)} \alpha_m u_{mi} + \sum_{k=1}^{K} \beta_k x_j^k. \tag{2.5}$$

The situation $I^*(n) < I(n)$ arises when the values of the underlying hazard rate function on different intervals are constrained to satisfy certain relationships. For instance, when $I^*(n) = 1$, and $u_{1ij} = 1$, then

$$\ell_{ij} = \alpha + \sum_k \beta_k x_j^k. \tag{2.6}$$

Except for differences in notation, this model is equivalent to the exponential model (1.1), generalized to multiple covariates.

EXAMPLE 2.2. When $I^*(n) = I(n)$, the first $I(n)$ basis vectors of $\Omega(n)$ may be defined by

$$\begin{aligned} u_{mij} &= 1 & &\text{if} \quad i = m, \\ &= 0 & &\text{otherwise.} \end{aligned} \tag{2.7}$$

In this case, for every ℓ in $\Omega(n)$,

$$\ell_{ij} = \alpha_i + \sum_k \beta_k x_j^k. \tag{2.8}$$

This very important piecewise exponential model was first discussed by Holford (1976).

2.2. *Existence of Maximum Likelihood Estimates.* If a unique vector $\hat{\ell}$ which maximizes the function L exists, it can be readily found by an iterative procedure. If all the summands in (2.3) were positive, the existence of a unique maximum likelihood estimate would be easy to prove. A similar problem has been studied for Poisson models in contingency table analysis (Haberman, 1973), where a log likelihood function of the form

$$L(\boldsymbol{\mu}) = c + \sum_i \mu_i n_i - \sum_i \exp(\mu_i)$$

appears, with n_i's the observed frequencies and c a term not involving the parameters μ_i.

Let the set of indices (i, j) be partitioned into $A = \{(i, j) : t_{ij} > 0\}$ and $B = \{(i, j) : t_{ij} = 0\}$. Clearly, the value of $L(\ell)$ is not changed when the summation is only over A. We will henceforth assume the following condition, which insures that if the likelihood function has a maximum, then a unique MLE exists.

CONDITION 1. For any pair of vectors $\mathbf{x}, \mathbf{y} \in \Omega(n)$, if $x_{ij} = y_{ij}$ for every (i, j) in A, then $\mathbf{x} = \mathbf{y}$.

Let L' and L'' be the first and second differential, respectively, of L at ℓ. Then for any $\mathbf{x}$ and $\mathbf{y}$ in $\Omega(n)$,

$$L'_\ell(\mathbf{x}) = \Sigma x_{ij} I_{ij} - \Sigma x_{ij} t_{ij} \exp(\ell_{ij}),$$

and

$$L''_\ell(\mathbf{x}, \mathbf{y}) = \Sigma x_{ij} y_{ij} t_{ij} \exp(\ell_{ij}).$$

Given Condition 1, it is then easy to see that $L(\ell)$ is a strictly concave function. Let $\mathbf{I}$ be the vector in $R^{nI(n)}$ whose components are I_{ij}, and $\mathbf{t}^*(\ell)$ the vector whose components are $t_{ij}\exp(\ell_{ij})$. Then we have

THEOREM 2.1. *Assume that Condition* 1 *is satisfied. If a maximum likelihood estimate* $\hat{\ell}$ *exists, it is unique, and satisfies the equation*

$$P_{\Omega(n)}\mathbf{I} = P_{\Omega(n)}\mathbf{t}^*(\ell), \tag{2.9}$$

where $P_{\Omega(n)}$ *is the orthogonal projection onto* $\Omega(n)$. *Conversely, a solution* $\hat{\ell}$ *of* (2.9) *is the* MLE.

PROOF. The proof is based on the observation that if $L'_\ell(x) = 0$ for every $\mathbf{x}$ in $\Omega(n)$, then $\mathbf{I} - \mathbf{t}^*(\hat{\ell})$ is orthogonal to $\Omega(n)$. We omit the details.

THEOREM 2.2. *Assume that Condition* 1 *holds. In order that the maximum likelihood estimate exists, it is necessary and sufficient that there exists a vector* $\mathbf{d}$ *in* $R^{nI(n)}$ *such that*

$$\begin{aligned} &\text{(a)}\ \ \mathbf{d} \in \Omega(n)^\perp, \quad \textit{the orthogonal complement of } \Omega(n);\\ &\text{(b)}\ \ d_{ij} + I_{ij} > 0 \quad \textit{for each} \quad (i,j) \in A, \quad \textit{and}\\ &\text{(c)}\ \ d_{ij} = 0 \quad \textit{for each} \quad (i,j) \in B. \end{aligned} \tag{2.10}$$

PROOF. When $\hat{\ell}$ exists, then the vector $\mathbf{d} = \mathbf{t}^*(\hat{\ell}) - \mathbf{I}$ satisfies (2.10). On the other hand, if a vector satisfying (2.10) exists, then

$$L(\ell) = \textstyle\sum_A \ell_{ij}(I_{ij} + d_{ij}) - \sum_A t_{ij}\exp(\ell_{ij}).$$

Since all the quantities $(I_{ij} + d_{ij})$ and t_{ij} in the last expression are positive, the rest of the proof follows as in Haberman (1973).

Let S be any set of ordered pairs (i, j) and let M be any subspace of $R^{nI(n)}$. Define the function p_S from $R^{nI(n)}$ to $R^{|S|}$, where $|S|$ denotes the number of elements in a subset S of the index set, by $p_S(\mathbf{x}) = \{x_{ij}: (i,j) \in S\}$, and let $p_S(M) = \{p_S(\mathbf{x}): \mathbf{x} \in M\}$. Then we have the following corollary.

COROLLARY 2.1. *Let* $\mathbf{I}' = p_A(I)$. *The* MLE *exists if and only if there exists a vector* $\mathbf{d}$ *such that if* $\mathbf{d}' = p_A(\mathbf{d})$, *then*

$$\text{(a)}\ \ \mathbf{d}' = [p_A(\Omega(n))]^\perp; \qquad \textit{and} \quad \text{(b)}\ \ d'_{ij} + I'_{ij} > 0 \quad \textit{for every} \quad (i,j). \tag{2.11}$$

Because of the similarities between (2.11) and the condition in Theorem 3.2 of Haberman (1973), corollaries such as 3.3 and 3.4 of the latter paper also apply in the present situation.

3. Asymptotic properties. When the true underlying hazard rate function is not a step function and $I(n)$ is constant, the maximum likelihood estimates $\hat{\beta}$ of the regression parameters, which can be expressed as linear functionals of $\hat{\ell}$, are, as expected, not consistent. This fact was first pointed out by Holford (1976).

The focus of this section will be on the asymptotic properties of the MLEs in the situation where the true underlying hazard rate is not a step function, the interval lengths decrease, and the dimension of $\Omega(n)$ increases without bound as n increases, but $T_{I(n)}$ is fixed. The proofs of the asymptotic properties will be obtained by adapting the fixed-point methods first used by Haberman (1977) in a study of exponential response models.

3.1. *Fixed-point theorems.* The asymptotic methods used in this section rely on the fixed point theorems of Kantorovich and Akilov (1964). (A fixed point of a function $f(x)$ satisfies the equation $x = f(x)$.) Consider the equation $\mathbf{x} = S(\mathbf{x})$, where S has a continuous differential on the closed sphere $\{\mathbf{x}: |\mathbf{x} - \mathbf{x}_0| \leq f\}$ of a complete normed vector space V.

Let there also be a real equation $t = g(t)$, with g defined and differentiable on an interval $[t_0, t']$, where $t' = t_0 + f$.

THEOREM 3.1. *Let the functions S and g, and the points $\mathbf{x}_0$ and t_0, be defined as above, and suppose that they satisfy*

$$\text{(3.1)} \qquad \begin{aligned} &\text{(a)} \ |S(\mathbf{x}_0) - \mathbf{x}_0| \le g(t_0) - t_0; \\ &\text{(b)} \ \| S'_x \| \le g'_t, \quad \textit{whenever} \quad |\mathbf{x} - \mathbf{x}_0| \le t - t_0. \end{aligned}$$

Define the two sequences $\{t_n\}$ and $\{\mathbf{x}_n\}$, $n = 0, 1, \cdots$, by $t_{n+1} = g(t_n)$ and $\mathbf{x}_{n+1} = S(\mathbf{x}_n)$. Then if the equation $t = g(t)$ has a root in $[t_0, t']$, the sequence $\{\mathbf{x}_n\}$ converges to a fixed point $\mathbf{x}^*$ of $S(\mathbf{x})$. Moreover, $\mathbf{x}^*$ satisfies the inequality

$$|\mathbf{x}^* - \mathbf{x}_n| \le t^* - t_n, \qquad n = 0, 1, \cdots$$

where t^* is the smallest root of $t = g(t)$ in $[t_0, t']$.

PROOF. See Kantorovich and Akilov (1964, pages 697–700).

The following theorem is implicitly used in Haberman (1977). It provides sufficient conditions for the convergence of a sequence $\{\mathbf{x}_n\}$ defined as in Theorem 3.1 without any explicit reference to a real function $g(t)$.

THEOREM 3.2. *Let S, $\{\mathbf{x}_n\}$, and f be defined as above. If b is a real number, and the following three conditions are satisfied*

$$\text{(3.2)} \qquad \begin{aligned} &\text{(a)} \ fb \le 1/2, \\ &\text{(b)} \ |\mathbf{x}_1 - \mathbf{x}_0| \le f/2 \\ &\text{(c)} \ \| S'_x \| \le b\,|\mathbf{x} - \mathbf{x}_0| \quad \textit{whenever} \quad |\mathbf{x} - \mathbf{x}_0| \le f, \end{aligned}$$

then the sequence $\{\mathbf{x}_n\}$ converges to a fixed point.

PROOF. The theorem can be proven by working with the function $g(t) = |\mathbf{x}_1 - \mathbf{x}_0| + \tfrac{1}{2}bt^2$, and a sequence $\{t_n\}$, with $t_0 = 0$, which converges to a fixed point t^*. It can also be shown that

$$\text{(3.3)} \qquad |\mathbf{x}^* - \mathbf{x}_1| \le t^* - t_1 \le b\,|\mathbf{x}_1 - \mathbf{x}_0|^2.$$

3.2 *Preliminaries.* Three different inner products need to be defined on the vector spaces $\Omega(n)$. Let $\{\cdot, \cdot\}$ denote, for every $\mathbf{x}, \mathbf{y} \in R^{nI(n)}$, the function

$$\{\mathbf{x}, \mathbf{y}\} = \textstyle\sum_{i,j} x_{ij} y_{ij} P_{ij},$$

where $P_{ij} = E(I_{ij})$. If the unique maximum likelihood estimate $\hat{\ell}_n$ exists, let

$$(\mathbf{x}, \mathbf{y}) = \textstyle\sum_{i,j} x_{ij} y_{ij} t_{ij} \exp(\hat{\ell}_{nij}).$$

Note that there may be $\mathbf{x} \ne \mathbf{0}$ such that $\{\mathbf{x}, \mathbf{x}\} = 0$ or $(\mathbf{x}, \mathbf{x}) = 0$. But under the conditions imposed below, the probability goes to 1 that both $\{\cdot, \cdot\}$ and $(\cdot, \cdot)$ define inner products on $\Omega(n)$. Let the set of all n's where this happens be denoted by $\mathcal{N}$. The corresponding norms will then be defined by $|||\,\mathbf{x}\,|||^2 = \{\mathbf{x}, \mathbf{x}\}$ and $|\mathbf{x}|^2 = (\mathbf{x}, \mathbf{x})$. Finally, for every $n \in \mathcal{N}$ let

$$\begin{aligned} \lambda^*_{nij} &= \frac{P_{ij}}{E(t_{ij})} && \text{if} \quad E(t_{ij}) > 0, \\ &= \exp(\ell^0_{nij}) && \text{if} \quad E(t_{ij}) = 0, \end{aligned}$$

where ℓ^0_{nij} is defined below. Further, let ℓ^*_n be the vector with coordinates $\ell^*_{nij} = \log(\lambda^*_{nij})$

and define ℓ_n^0 to be the vector minimizing $\Sigma_{i,j}(\ell^*_{nij} - x_{ij})^2 P_{ij}$ over all $\mathbf{x}$ in $\Omega(n)$. If $C = \{(i, j) : P_{ij} > 0\}$, then $p_C(\ell_n^0) = P'_{p_C(\Omega(n))} p_C(\ell_n^*)$, where P' is the orthogonal projection from $p_C(R^{nI(n)})$ onto $p_C(\Omega(n))$ with respect to the inner product $\{\cdot, \cdot\}$. Note that the uniqueness of ℓ_n^0 is implied by $n \in \mathcal{N}$.

A third inner product $[\cdot, \cdot]$ is defined by $[\mathbf{x}, \mathbf{y}] = \Sigma x_{ij} y_{ij} t_{ij} \exp(\ell^0_{nij})$, and its corresponding norm is $\| \cdot \|$. For any $\mathbf{x}$ in $\Omega(n)$, let $f_n(\mathbf{x})$ be the vector defined by requiring that for any $\mathbf{y}$ in $\Omega(n)$,

$$[\mathbf{y}, f_n(\mathbf{x})] = \textstyle\sum_{i,j} y_{ij} \{I_{ij} - t_{ij} \exp(x_{ij})\}.$$

Further, let $g_n(\mathbf{x}) = \mathbf{x} + f_n(\mathbf{x})$, and $\mathbf{Z}_n = f_n(\ell_n^0)$. If $\hat{\ell}_n$ is the solution of the equation $g_n(\mathbf{x}) = \mathbf{x}$, then by the definition of g_n, $\hat{\ell}_n$ satisfies equation (2.9). If a unique MLE exists, it is therefore equal to the fixed point of g_n.

It will be assumed in the remainder of this paper that $n \in \mathcal{N}$; this involves no loss of generality.

3.3 Existence of Maximum Likelihood Estimates. Despite the large number of technical details involved in the proof of the asymptotic results, the main line of development is simple. It is to be expected that the MLE will converge in some sense to ℓ_n^0, the projection on $\Omega(n)$ of the vector of logarithms of the average interval hazard rates λ^*_{nij}. The approach taken to proving that the difference between ℓ_n^0 and $\hat{\ell}_n$ is small consists in investigating the consequences of letting ℓ_n^0 be the first approximation to the fixed point of g_n. In other words, we will study a sequence of vectors $\{\ell_{ni}\}$, $i = 0, 1, \cdots$ with $\ell_{no} = \ell_n^0$, and $\ell_{n,i+1} = g_n(\ell_{ni})$. It is important to note that $\ell_{n1} = \ell_n^0 + \mathbf{Z}_n$. To prove the existence of a MLE with probability 1 as $n \to \infty$, it suffices to show that $\{\ell_{ni}\}$ converges. To study the asymptotic behavior of the MLE, it is necessary to show that $\hat{\ell}_n$ is sufficiently well approximated by ℓ_{n1}, and then to investigate the behavior of $\mathbf{Z}_n$. Four conditions are needed:

CONDITION A. This condition concerns the distribution of the covariate values. Note that it relaxes the customary requirement that the values of the covariate be bounded. Let $\mathbf{v}(\mathbf{a})$ be the vector with components $v_j(\mathbf{a}) = \mathbf{a}'\mathbf{x}_j$, and let $m_i(\mathbf{a})$ and $d_i(\mathbf{a})$ be, respectively, the weighted mean and variance of $\{v_j(\mathbf{a})\}$, with $\{P_{ij}\}$ as weights. Define

$$P_{i+} = \textstyle\sum_j P_{ij}, \qquad P_{+j} = \sum_i P_{ij} \qquad \text{and } P_{++} = \sum_i \sum_j P_{ij}.$$

Then $\max_i m_i(\mathbf{a})$ and $\max_i d_i(\mathbf{a})$ are both bounded as $n \to \infty$,

$$\textstyle\sum_i P_{i+} d_i(\mathbf{a}) / \sum_i P_{i+} \to c(\mathbf{a}),$$

and

$$\max_{\mathbf{a}} \max_j v_j^2(\mathbf{a})/c(\mathbf{a}) = O(I'(n)), \tag{3.4}$$

where $I'(n)$ is the dimension of $\Omega(n)$, assumed to be of the same order of magnitude as $I(n)$. Also, the second moment of $\{\exp |(\boldsymbol{\beta}, \mathbf{x}_j)|\}$, either unweighted or with weights $\{P_{ij}\}$ or $\{P_{+j}\}$, is bounded.

CONDITION B. (i) $1/P_{++} = O(n-1)$, (ii) $\max_{i,i'} P_{i+}/P_{i'+} = O(1)$. (iii) If w_i is the length of the ith interval, then $\max_i w_i = o(1)$, and $\max_{i,j} w_i \exp(\boldsymbol{\beta}, \mathbf{x}_j) = O(1)$.
This condition puts another constraint on the distribution of the covariates, one which in most situations is stronger than the one expressed by (3.4). It is also specified that the intervals all go to zero in length, and all be of the same order of magnitude when measured in terms of the expected number of deaths in each. The next condition limits the rate of increase of the number of intervals.

CONDITION C.

$$\frac{\{I'(n)\}^3}{n} = o(1).$$

CONDITION. D. This condition basically specifies that the distance between ℓ_n^* and its projection on $\Omega(n)$ be small.

(i) Let $q_{nij} = \ell_{nij}^0 - \ell_{nij}^*$. Then the quantities $\max_{i,j}|q_{nij}|$, $\sum_{i,j} P_{ij}q_{nij}^4$, and $n^{-1}I'(n)^2 \sum_{i,j} P_{ij}q_{nij}^2$ are all of order $o(1)$.

(ii) The function $\lambda_0(t)$ is bounded on $(0, T]$.

It can be shown (Friedman, 1981) that Condition D is satisfied in a general class of situations involving basically the smoothness of $\lambda_0(t)$.

The development now leads to Lemmas 3.3, 3.4, and 3.5, which directly allow us to apply Theorem 3.2 to the present situation.

LEMMA 3.1. *Given Conditions A–D,*

$$\max_{\mathbf{z}\in\Omega(n)} \left| \frac{\|\mathbf{z}\|^2}{|||\mathbf{z}|||^2} - 1 \right| = o_p(1).$$

PROOF. The proof is given in Friedman (1981). For general comments, see Section 5.

LEMMA 3.2. *Define a sequence of random variables* $\{b_n\}$ *by*

$$b_n = c \max_{\mathbf{z}\in\Omega(n)} \max_{i,j} \frac{|z_{ij}|}{\|\mathbf{z}\|},$$

where $c > 1$. *Then given Conditions A–D,*

$$b_n^2 = O_p(I'(n)/n).$$

PROOF. Because Lemma 3.1 implies that $\max_{\mathbf{z}\in\Omega(n)} |||\mathbf{z}|||/\|\mathbf{z}\| = O_p(1)$, it suffices to demonstrate that

$$\max_{\mathbf{z}\in\Omega(n)} \max_{i,j} z_{ij}^2/|||\mathbf{z}|||^2 = O(I'(n)/n).$$

The remainder of the proof is presented in Section 5.

LEMMA 3.3 *Given A–D, there exists a sequence* $\{f_n\}$ *of real numbers greater than* 1 *such that* $I'(n)/f_n^2 = o(1)$ *and* $b_n f_n^2 = o_p(1)$.

PROOF. Let $f_n = \{nI'(n)\}1/8$. Then by Lemma 3.2, $b_n f_n^2 = O_p(I'(n)3/4/n1/4)$, and $I'(n)/f_n^2 = \{I'(n)\}3/4/n1/4$.

LEMMA 3.4. *Given Conditions A–D,* $\Pr(\|\mathbf{Z}_n\| \le f_n/2) \to 1$, *where* $\{f_n\}$ *is defined as in Lemma* 3.3.

PROOF. It is shown in Friedman (1981) that $E(\|\mathbf{Z}_n\|^2) = I'(n) + o(I'(n))$. The lemma then follows from Chebyshev's inequality.

LEMMA 3.5. *If Conditions A–D are satisfied,* $\{f_n\}$ *is defined as in Lemma* 3.3, *and* $G_n = \{\mathbf{x} \in \Omega(n) : \|\mathbf{x} - \ell_n^0\| \le f_n\}$, *then*

$$\Pr(\sup_{\mathbf{x}\in G_n} \|g'_{n,\mathbf{x}}\| \le b_n \|\mathbf{x} - \ell_n^0\|) \to 1.$$

PROOF. See Section 5.

The existence and some of the basic properties of the maximum likelihood estimates can now be proven.

THEOREM 3.3. *Let Conditions A–D be satisfied, and let* ℓ_n^0 *be the first approximation to the fixed point of* g_n. *Then as n goes to infinity, the probability approaches* 1 *that the sequence of approximations converges to a fixed point* $\hat{\ell}_n$.

PROOF. It is sufficient to demonstrate that the probability goes to 1 for the conditions specified in Theorem 3.2 to be satisfied. By Lemma 3.3, the probability goes to 1 that $f_n b_n \leq \frac{1}{2}$. Lemma 3.4 places the necessary bound on the distance between the first two approximations to the fixed point, and Lemma 3.5 guarantees that the probability goes to 1 that the bound on $\| g'_{n,\mathbf{x}} \|$ is satisfied.

THEOREM 3.4. *Under the conditions of Theorem* 3.3, *the probability goes to* 1 *that the* MLE $\hat{\ell}_n$, *if it exists, is unique.*

PROOF. See Section 5.

It can be concluded, moreover, from the proof of Theorem 3.2 that the probability goes to 1 that $\| \ell_{ni} - \ell_n^0 \| \leq f_n$, $i = 1, 2, \cdots$, and in particular that

$$\| \hat{\ell}_n - \ell_n^0 \| \leq f_n. \tag{3.5}$$

While f_n goes to infinity, the components of $\hat{\ell}_n - \ell_n^0$ become small.

COROLLARY 3.1. *If* $\hat{\ell}_n$ *exists, then* $\max_{i,j} | \hat{\ell}_{nij} - \ell_{nij}^0 | \to_P 0$.

PROOF. The corollary follows from (3.5) and Lemma 3.3.

3.4 Asymptotic normality of linear functionals of the Maximum Likelihood Estimate. The estimates of the regression parameters of a piecewise exponential model for survival data may be expressed as linear functionals of $\hat{\ell}_n$. In this subsection, general properties of linear functionals of $\hat{\ell}_n$ will be examined. Let h_n be a linear functional defined on $\Omega(n)$, so that for every $\mathbf{x}$ in $\Omega(n)$, $h_n(\mathbf{x}) = \sum_{i,j} c_{nij} x_{ij}$. Define the quantity

$$\sigma(h_n) = \sup_{\mathbf{x} \in \Omega(n)} \frac{| h_n(\mathbf{x}) |}{||| \mathbf{x} |||}.$$

For each h_n, there exists a vector $\mathbf{d}_n$ in $\Omega(n)$ such that for all $\mathbf{x}$ in $\Omega(n)$, $h_n(\mathbf{x}) = \{\mathbf{d}_n, \mathbf{x}\}$. Then by the Schwartz inequality, $\sigma(h_n) = ||| \mathbf{d}_n |||$. One additional lemma is required before the theorem on asymptotic normality can be presented.

LEMMA 3.6. *Given Conditions A–D, if* $\{\mathbf{a}_n\}$ *is any sequence of vectors in* $\Omega(n)$ *such that* $||| \mathbf{a}_n ||| \neq 0$, *then as n goes to infinity,*

$$\frac{\{\mathbf{a}_n, \mathbf{Z}_n\}}{||| \mathbf{a}_n |||} - \frac{[\mathbf{a}_n, \mathbf{Z}_n]}{||| \mathbf{a}_n |||} \to_P 0.$$

PROOF. See Friedman (1981).

THEOREM 3.5. *If Conditions A–D are satisfied, and if the sequence* $\{h_n\}$ *is such that* $\sigma(h_n) > 0$, *then*

$$\frac{h_n(\hat{\ell}_n) - h_n(\ell_n^0)}{\sigma(h_n)} \to_D N(0, 1).$$

PROOF. See Section 5.

Since $\sigma(h_n)$ is unknown, it must be estimated by $s(h_n) = \sup_{x \in \Omega(n)} | h_n(\mathbf{x}) | / | \mathbf{x} |$. The estimate $s(h_n)$ is equivalent to the estimate of the variance obtained in the usual way from the matrix of the second partial derivatives of L evaluated at $\hat{\ell}_n$.

THEOREM 3.6. *Let* $\{h_n\}$ *be any sequence of linear functionals on* $\Omega(n)$ *such that* $s(h_n) > 0$ *for each n. Then under the conditions of Theorem* 3.3, $\sigma(h_n)/s(h_n) \to_P 1$.

PROOF. See Section 5.

EXAMPLE 3.1. Consider the model (2.8) with $K = 1$, and for each n for which $\hat{\ell}_n$ exists, let its coordinates be written as $\hat{\ell}_{nij} = a_{ni}(\hat{\ell}_n) + b_n(\hat{\ell}_n)x_j$, where a_{ni} and b_n are linear functionals $b_n(\ell) = (\ell_{n11} - \ell_{n12})/(x_1 - x_2)$ and $a_{ni}(\ell) = (\ell_{n12}x_1 - \ell_{n11}x_2)/(x_1 - x_2)$.

COROLLARY 3.2. *If the conditions of Theorem* 3.3 *are satisfied, then*

$$b_n(\hat{\ell}_n) - b_n(\hat{\ell}^0_n) \to_P 0.$$

PROOF. It can be shown that $h_n(\hat{\ell}_n) - h_n(\ell^0_n) \to_P 0$ if $\Sigma |c_{nij}|/f_n \to 0$. It is easy to check that the functionals b_n satisfy this condition.

It is also desirable to prove that $\sqrt{n}\,(b_n(\ell^0_n) - \beta) \to 0$. Under the conditions alluded to above and described in Friedman (1981) which lead to Condition D being satisfied, it is necessary to impose the requirement that $n/I'(n^4) = o(1)$.

4. A Numerical Example. The MLE can be found in practice by the usual iterative methods; see Holford (1976), Friedman (1978). A data set, from Merrell and Shulman (1955), consisting of the survival times of 98 lupus erythematosus patients, was analyzed using model (2.8). The covariates are (1) sex; (2) race (white/non-white); (3) age at time of diagnosis; (4) time elapsed between estimated onset and diagnosis (≤ 2 years/> 2 years); and (5) recency of diagnosis (before/after July 1951). The age of patient 49 is unknown, and she was therefore not used in the analysis.

The models I–VI in Table 1 were fitted mainly to illustrate the variability of the regression parameters in the piecewise exponential model as the number of intervals and the intervals themselves change. Inspection of the table reveals that while the goodness of fit of the models may vary substantially, the estimates of the regression parameters do not change greatly relative to the magnitude of their estimated standard errors. This is comforting in view of the fact that despite the asymptotic results about the allowable rate of increase in $I(n)$, precise practical guidelines for choosing the number of intervals have not been formulated. In general, monotonically increasing or decreasing functions $\lambda_0(t)$ will lead to greatest bias. Calculations in Friedman (1978) show, however, that for a balanced two-sample study with $\lambda_2(t)/\lambda_1(t) = 2$ and $\lambda_0(t)$ proportional to t, the bias in $\hat{\beta}$ is only about 4% for a certain choice of intervals when $I = 5$, and decreases further as the number of intervals is increased.

It is recommended that an analysis start with a moderate I (5–7), followed by an examination of the $\hat{\alpha}_i$'s and their estimated standard error to identify sharp changes in the underlying hazard rate function, and especially to identify any monotone trend in $\lambda_0(t)$. When a monotone trend is found, a transformation of the time scale should be considered.

TABLE 1
Analysis of lupus erythematosus data

Model	Interval Boundaries (in years)	b_1	b_2	b_3	b_4	b_5	log likelihood at $\hat{\ell}$
I	.5, .8, 1.1, 1.7,	−.484	−.663	.0013	.497	1.125	−67.84
	2.5, 3.1	(.54)	(.39)	(.015)	(.38)	(.49)	
II	.3, .8, 1.0, 2.0, 3.0	−.457	−.653	.0008	.488	1.191	−61.94
III	.4, .9, 1.5, 2.5	−.483	−.658	.0009	.513	1.117	−68.53
IV	.3, 1.0, 2.0, 3.0	−.450	−.623	.0019	.464	1.244	−66.39
V	.4	−.473	−.648	.0011	.494	1.151	−70.41
VI	.3	−.459	−.635	.0011	.489	1.192	−67.78
VII	Cox Model	−.459	−.615	.0002	.503	1.151	

Numbers in parentheses are estimated standard errors for Model I.

For example, the estimates $\{\hat{\alpha}_i\}$ for Model IV, and their corresponding estimated standard errors in parentheses, are given below.

−1.170	−2.322	−3.285	−2.879	−2.284
(.29)	(.37)	(.59)	(.60)	(.60)

The hazard rate during the first three months after diagnosis is higher than at other times, but there is no evidence of a monotone trend.

Model VII is the Cox model, fitted using Procedure PHGLM of the Statistical Analysis System. This procedure uses the Breslow (1974) modification for tied data. It is encouraging that the estimates of the regression parameters in Models I–VI do not differ greatly from the Cox estimate. As a rough measure of this difference, the differences between the Cox estimates and the piecewise exponential estimates were divided by estimated standard errors given in Table 1, the ratios squared and summed over the five regression parameters. The results are .026, .019, .022, .060, .012, and .015.

Further discussions of and examples of the piecewise exponential methodology can be found in Holford (1976), Holford (1980) and Laird and Olivier (1981).

5. Further Proofs. This section contains proofs of several lemmas and theorems of Section 3, as well as an example showing that Condition D is satisfied in a general class of situations.

Proof of Lemmas 3.1 and 3.6. These lemmas are proven for the general model (2.4) with $K = 1$ in Friedman (1981). The proof involves defining P_n and Q_n to be diagonal matrices whose non-zero elements are respectively P_{ij} and $t_{ij}\exp(\ell^0_{nij})$, defining D_n to be the $n \times \{I(n) + 1\}$ matrix whose first $I(n)$ columns are the vectors defined by (2.7) and whose last column consists of the values of x, and showing that $|||\mathbf{Z}_n|||\lambda_{mn} = o_p(1)$, where λ^2_{mn} is the maximum of the squared eigenvalues of the matrix $I - (D_n'PD)^{-1}(D_n'Q_nD_n)$.

Proof of Lemma 3.2. Let the coordinates of each $\mathbf{z}$ be written as

$$z_{ij} = \textstyle\sum_m \alpha_m u_{mi} + \sum_k \gamma_k x_j^k = u_i(\mathbf{z}) + w_j(\mathbf{z}),$$

and $\Omega(n)$ be partitioned into

$$\Omega^1(n) = \{\mathbf{z} : |u_{i^*(\mathbf{z})}(\mathbf{z})| = \max_i |u_i(\mathbf{z})| > 2\max_j |w_j(\mathbf{z})| = 2|w_{j^*(\mathbf{z})}(\mathbf{z})|\}$$

and

$$\Omega^2(n) = \{\mathbf{z} : |u_{i^*(\mathbf{z})}| \le 2|w_{j^*(\mathbf{z})}(\mathbf{z})|\}.$$

Then

$$\max_{\mathbf{z}\in\Omega^1(n)} \max z_{ij}^2/|||\mathbf{z}|||^2 \le \max_{\mathbf{z}\in\Omega^1(n)} \frac{2\{u_{i^*(\mathbf{z})}(\mathbf{z})\}^2 + 2\{w_{j^*(\mathbf{z})}(\mathbf{z})\}^2}{\sum_j \{u_{i^*(\mathbf{z})}(\mathbf{z}) + w_j(\mathbf{z})\}^2 P_{i^*(\mathbf{z}),j}}$$

$$\le \max_{\mathbf{z}\in\Omega^1(n)} \frac{2.5\{u_{i^*(\mathbf{z})}(\mathbf{z})\}^2}{\left(\frac{1}{4}\right)\sum_j \{u_{i^*(\mathbf{z})}(\mathbf{z})\}^2 P_{i^*(\mathbf{z}),j}} \le 10 \min_i 1/\textstyle\sum_j P_{ij},$$

so that as a consequence of condition B,

$$\max_{\mathbf{z}\in\Omega^1(n)} \max_{i,j} z_{ij}^2/|||\mathbf{z}||| = O(I'(n)/n).$$

On the other hand,

$$\max_{\mathbf{z}\in\Omega^2(n)} \max_{i,j} z_{ij}^2/|||\mathbf{z}||| \le \max_{\mathbf{z}\in\Omega^2(n)} \max_j \frac{10\{w_{j^*(\mathbf{z})}(\mathbf{z})\}^2}{\sum_{i,j}\{w_j(\mathbf{z}) - \bar{w}_i(\mathbf{z})\}^2 P_{ij}} = O(I'(n)/n)$$

where $\bar{w}_i(\mathbf{z}) = \sum_j w_j P_{ij}/\sum_j P_{ij}$.

PROOF OF LEMMA 3.5. The proof of this lemma uses the easily verified fact that for every real number x,

$$|\exp(x) - 1| \le |x| \exp(|x|).$$

From the definition of b_n and Lemma 3.3 it follows that for any $\varepsilon > 0$,

$$\Pr(\sup_{\mathbf{x}\in G_n} \max_{i,j} |\mathbf{x}_{ij} - \ell^0_{nij}| < \varepsilon) \to 1. \tag{3.6}$$

Note that for a self-adjoint linear transformation from V to V, $\|T\| = \sup_{x\in V} |(\mathbf{x}, T\mathbf{x})| / |\mathbf{x}|^2$; see, e.g., Kantorovich and Akilov (1964). As a consequence of the fact that $[y, g'_{n,\mathbf{x}}(y)] = \sum y^2_{ij} t_{ij} [\exp(\ell^0_{nij}/x_{ij})]$, we have

$$\begin{aligned}\|g'_{n,\mathbf{x}}\| &= \max_{\mathbf{y}\in\Omega(n)} \left| \frac{\sum_{i,j} y^2_{ij} t_{ij} \{\exp(\ell^0_{nij}) - \exp(x_{ij})\}}{\sum_{i,j} y^2_{ij}\, t_{ij} \exp(\ell^0_{nij})} \right| \\ &= \max_{\mathbf{y}\in\Omega(n)} \left| \frac{\sum_{i,j} y^2_{ij} t_{ij} \exp(\ell^0_{nij}) \{1 - \exp(x_{ij} - \ell^0_{nij})\}}{\sum_{i,j} y^2_{ij} \exp(\ell^0_{nij})} \right| \\ &\le \max_{i,j} |1 - \exp(x_{ij} - \ell^0_{nij})| \le \max_{i,j} |x_{ij} - \ell^0_{nij}| \exp(|x_{ij} - \ell^0_{nij}|).\end{aligned}$$

Therefore,

$$\Pr(\sup_{\mathbf{x}\in G_n} \|g'_{n,\mathbf{x}}\| \le b_n \|\mathbf{x} - \ell^0_n\|)$$
$$\ge \Pr\left(\forall \mathbf{x} \in G_n, \frac{\max_{i,j} |x_{ij} - \ell^0_{nij}|}{\|\mathbf{x} - \ell^0_n\|} \le \frac{b_n}{\max_{i,j} \exp(|\mathbf{x}_{ij} - \ell^0_{nij}|)}\right).$$

The lemma then follows by (3.6) and the definition of b_n.

PROOF OF THEOREM 3.4. By Theorem 2.1, it needs to be shown that Condition 1 is satisfied for large n. Condition 1 is satisfied if and only if $\dim p_A[\Omega(n)] = \dim \Omega(n)$. Note that if $\Omega \subset \Omega^*$, then $\dim p_A\Omega^* = \dim \Omega^*$ implies $\dim p_A\Omega = \dim \Omega$. Let $\Omega^*(n) = \{\mathbf{x} : x_{ij} = \alpha_i + (\boldsymbol{\beta}, \mathbf{x}_j)\}$. It can be verified that $\dim p_A\Omega^*(n) = I(n) + K$ if (a) $\sum_j I_{I(n),j} > 0$, and if (b) there is no vector $\mathbf{a}$ such that $\mathbf{a}'\mathbf{x}_j = 1$ for all j, since $(1, j) \in A$ for all j. But Condition B implies that the probability goes to 1 that (a) is true, while Condition A implies the same for (b).

PROOF OF THEOREM 3.5. If $h_n(\mathbf{x}) = \{\mathbf{d}_n, \mathbf{x}\}$ for each $\mathbf{x}$ in $\Omega(n)$, then the theorem can be shown to be true if

$$X_n = \frac{[\mathbf{d}, \mathbf{Z}_n]}{|||\mathbf{d}_n|||} = \frac{\sum_{i,j} d_{nij} \{I_{ij} - t_{ij} \exp(\ell^0_{nij})\}}{|||\mathbf{d}_n|||} \longrightarrow_D N(0, 1).$$

This can be accomplished by examining the logarithm of the moment generating function, $\log E\{\exp(tX_n)\}$, which can be shown to converge to $t^2/2$; see Curtiss (1942). In Friedman (1981) it is shown that if we write $X_n = \sum Y_{nj}$, where

$$Y_{nj} = \sum_i d_{nij} \{I_{ij} - t_{ij} \exp((\ell^0_{nij})\} / |||\mathbf{d}_n|||,$$

and if Conditions A–D are satisfied, then

(a) there is a sequence $\{K_n\}$ such that $K_n = o(1)$ and

$\Pr(\max_j |Y_{nj}| < K_n) = 1$;

(b) $E(X_n) = \sum_j E(Y_{nj}) = o(1)$;

(c) $\sum_j E^2(Y_{nj}) = o(1/I'(n))$;

(d) $\mathrm{Var}(X_n) = \sum_j Var(Y_{nj}) = 1 + o(1)$.

The function $\log E\{\exp(tX_n)\}$ is then expanded in a Taylor series.

THEOREM 3.6. *Let* $s^*(h_n) = \sup_{x\in\Omega(n)}|h_n(\mathbf{x})|/\|\mathbf{x}\|$. It follows from Lemma 3.1 that $\sigma(h_n)/s^*(h_n) \to_P 1$. If it can also be shown that $s^*(h_n)/s(h_n) \to_P 1$, the conclusion will follow. But since the probability goes to 1 that (3.5) holds, Lemma 3.5 implies that

$$\Pr\left\{\max_{\mathbf{x}\in\Omega(n)}\left|\frac{\sum_{i,j} x_{ij}^2 t_{ij})\exp(\hat{\ell}_{nij})}{\sum_{i,j}x_{ij}^2 t_{ij}\exp(\ell_{nij}^0)} - 1\right| \le b_n\|\hat{\ell}_n - \ell_n^0\|\right\}$$
$$= \Pr\left\{\max_{\mathbf{x}\in\Omega(n)}\left|\frac{|\mathbf{x}|^2}{\|\mathbf{x}\|^2} - 1\right| \le b_n\|\hat{\ell}_n - \ell_n^0\|\right\} \to 1,$$

and the conclusion follows.

The following lemmas indicate that Condition D is satisfied in a general class of situations.

LEMMA 4.1. *Let* $H_j(t)$ *be the probability that the* jth *individual is not censored before time t. Assume that* $\lambda_0(t)$ *is bounded on* $(0, T]$, *and that both* $\lambda_0(t)$ *and the functions* $H_j(t)$ *are continuous and have continuous first and second derivatives which are uniformly bounded for all j. Then for each n there are constants* λ_{ni}, $i = 1, \cdots, I(n)$ *and a constant B, such that for large enough n,*

$$|\ell_{nij}^* - \log(\lambda_{ni}) - (\boldsymbol{\beta}, \mathbf{x}_j)| \le Bw_i^2\exp(\boldsymbol{\beta}, \mathbf{x}_j).$$

PROOF. The proof relies on a simple Taylor series expansion of λ_{nij}^* as a function of w_i, and is given in Friedman (1978).

LEMMA 4.2. *If* $R_j(t)$ *is the probability of being observed at time t, assume that for large enough n* $\sum_j \lambda_j(t)R_j(t)/n \ge z(t)$, *where* $z(t)$ *is a continuous function taking only positive values on* $[0, T]$. *Then* $\max_i w_i = O(1/I(n))$.

PROOF. See Friedman (1981).

With λ_{ni} as in Lemma 4.1, let ℓ_n^+ have coordinates $\ell_{nij}^+ = \log(\lambda_{ni}) + (\boldsymbol{\beta}, \mathbf{x}_j)$. As a consequence of Condition A, Lemma 4.1, and the definition of ℓ_n^0,

$$\frac{I^2(n)}{n}\sum_{i,j} P_{ij}q_{nij}^2 \le \frac{I^2(n)}{n}\sum_{i,j} P_{ij}(\ell_{nij}^+ - \ell_{nij}^*)^2$$
$$\le \frac{B^2I^2(n)P_{++}}{n}\max_i w_i^4\left[\frac{\sum_j P_{+j}\exp\{2(\boldsymbol{\beta}, \mathbf{x}_j)\}}{P_{++}}\right] = I^2(n)\max_i w_i^4 O(1).$$

Further, since $|||\ell_n^+ - \ell_n^0|||^2 \le \sum_{i,j} P_{ij}(\ell_{nij}^* - \ell_{nij}^+)^2$,

$$\max_{i,j}|q_{nij}| \le \max_{i,j}|\ell_{nij}^* - \ell_{nij}^+| + \max_{i,j}|\ell_{nij}^+ - \ell_{nij}^0|$$
$$\le B(\max_i w_i)\{\max_{i,j} w_i\exp(\boldsymbol{\beta}, \mathbf{x}_j)\} + \frac{\{I'(n)\}^{1/2}}{n^{1/2}}|||\ell_n^+ - \ell_n^0|||$$
$$= (\max_i w_i)o(1) + \{I'(n)\}^{1/2}(\max_i w_i^2)O(1),$$

and

$$\sum_{i,j} P_{ij}q_{nij}^4 \le \max_{i,j} q_{nij}^2\sum P_{ij}q_{nij}^2 = n(\max_i w_i^4)(\max_{i,j} q_{nij}^2).$$

Thus if $\max_i w_i = O(1/I(n))$, Condition D is satisfied for the model specified by (2.8) as long as $n/I^6(n) \to 0$.

Acknowledgement. This paper is based on the author's dissertation at the University of Chicago. The author is grateful to Professor Shelby J. Haberman for the encouragement and advice generously provided during the course of this dissertation. The author also thanks the editor and referees for helpful comments and suggestions.

REFERENCES

BRESLOW, N. (1974). Covariance analysis of censored survival data. *Biometrika* **30** 89–99.

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

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.

CURTISS, J. H. (1942). A note on the theory of moment generating functions. *Ann. Math. Statist.* **13** 430–433.

FEIGL, P. and ZELEN, M. (1965). Estimation of exponential probabilities with concomitant information. *Biometrics* **21** 826–838.

FRIEDMAN, M. (1978). Analysis of survival data with covariates and censoring using a piecewise exponential model. Ph.D. dissertation, Univ. of Chicago.

FRIEDMAN, M. (1981). Piecewise exponential models for survival data with covariates. Rutgers Univ. Tech. Report.

HABERMAN, S. H. (1973). Log-linear models for frequency data: Sufficient statistics and likelihood equations. *Ann Statist.* **1** 617–632.

HABERMAN, S. H. (1977). Maximum likelihood estimates in exponential responses models. *Ann Statist.* **5** 815–841.

HOLFORD, T. R. (1976). Life tables with concomitant information. *Biometrics* **32** 587–597.

HOLFORD, T. R. (1980). The analysis of rates and of survivorship using log-linear models. *Biometrics* **36** 299–305.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60** 267–278.

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

KANTOROVICH, L. V. and AKILOV, G. P. (1964). *Functional Analysis in Normed Spaces.* Pergamon, New York.

LAIRD, N. and OLIVIER, D. (1981). Covariance analysis of censored survival data using log-linear analysis technique. *J. Amer. Statist. Assoc.* **76** 231–240.

PRENTICE, R. L. and KALBFLEISCH, J. D. (1979). Hazard rate models with covariates. *Biometrics* **35** 25–39.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NJ 08903