Statistics and Its Interface Volume 5 (2012) 195–206

Additive hazards regression and partial likelihood estimation for ecological monitoring data across space

Feng-Chang Lin and Jun Zhu*

We develop continuous-time models for the analysis of environmental or ecological monitoring data such that subjects are observed at multiple monitoring time points across space. Of particular interest are additive hazards regression models where the baseline hazard function can take on flexible forms. We consider time-varying covariates and take into account spatial dependence via autoregression in space and time. We develop statistical inference for the regression coefficients via partial likelihood. Asymptotic properties, including consistency and asymptotic normality, are established for parameter estimates under suitable regularity conditions. Feasible algorithms utilizing existing statistical software packages are developed for computation. We also consider a simpler additive hazards model with homogeneous baseline hazard and develop hypothesis testing for homogeneity. A simulation study demonstrates that the statistical inference using partial likelihood has sound finite-sample properties and offers a viable alternative to maximum likelihood estimation. For illustration, we analyze data from an ecological study that monitors bark beetle colonization of red pines in a plantation of Wisconsin.

AMS 2000 subject classifications: Primary 62N02; secondary 62M10.
Keywords and phrases: Current status data, Grouped survival data, Maximum likelihood, Multiple monitoring times, Spatial autoregression, Spatial lattice.

1. INTRODUCTION

In many environmental and ecological monitoring programs, subjects are observed across space and repeatedly over time. The motivating example here is a study of insect-tree interactions in a red pine plantation of Wisconsin [18, 23]. The trees were planted on a regular grid of sites. Each site where a tree was present was visited by researchers on an annual basis from 1986 to 1992. Two types of bark beetles were of interest, namely, *Ips* species (predominantly *Ips pini* (Say) and to a lesser extent *Ips grandicollis* (Eichhoff)), a bark beetle that colonizes the main stem of a tree, and *Dendroctonus valens* (LeConte), a bark beetle known as turpentine beetle that colonizes the lower stem of a tree. [23] analyzed the data for the purpose of evaluating the relation between the two types of bark beetles and the survival of trees. [18] focused on a subset of the data and analyzed *Ips* species in relation to turpentine beetle.

While [23] proposed autologistic-type models, which assume that time is discrete and coincides with the monitoring times, [18] proposed an alternative, continuous-time model with additive hazards regression that accounts for multiple monitoring times. The continuous-time model was shown to have several advantages over the discrete-time model. Parameters of different continuous-time models, particularly regression coefficients, are comparable even with different sampling frequencies, but those of different discrete-time models are not always comparable. In addition, even though observations are made at discrete points in time, most environmental or ecological processes of interest are over continuous time. An underlying continuous-time process is not guaranteed to exist in the specification of a discrete-time model, while a continuous-time model does not have this issue. However, the modeling approach in [18] is fully parametric and requires that the baseline hazard be estimated using external data. Here we consider this continuous-time modeling framework, but develop an alternative, semiparametric model, such that the baseline hazard has a flexible form and thus does not require estimation using external data as in [18].

While nonparametric maximum likelihood estimation of a distribution function can be used to analyze such monitoring data [22, 8], semiparametric regression models offer a viable alternative [7, 5, 21, 12, 13]. Among the existing models and estimation methods, an additive hazards regression developed by [12] and [13] is particularly appealing, as time-varying covariates can be readily included in the model, which broadens its applicability in a wide variety of disciplines. In addition, the model can be transformed to a proportional hazards model and hence statistical inference can be carried out by most standard statistical software packages. However, the methods by [12] and [13] are limited to current status data, in that a subject is monitored only once at a random point in time.

Here we aim to extend the additive hazards regression for current status data to multiple monitoring times, in that

*Corresponding author.

a subject is followed more than once at pre-determined points in time. This type of data occur often in ecological monitoring programs, but it is not obvious how the methodology developed in [12] or [13] can be generalized to analyze such data. Thus we take a different approach. In particular, since each subject is monitored regularly and no lost-to-follow-up occurs during the study, the data can be viewed as a type of grouped failure time [15]. However, regression models for grouped failure time are mostly confined to multiplicative forms, while additive forms are largely unexplored. An exception is recent work by [19] who considered maximum likelihood estimation for additive hazards regression in an ecological monitoring study, although their data were not spatially referenced and thus there was no need to account for spatial dependence.

Our main contribution is to develop valid and feasible statistical inference for grouped failure time that has additive forms. More specifically, we propose to transform the additive hazard to proportional hazard at discrete monitoring times and apply partial likelihood estimation. We also compare partial likelihood with maximum likelihood estimation. Furthermore, since our method may be implemented in a standard statistical software package, it has computational advantage over, for example, [18] where statistical inference is via Bayesian hierarchical modeling and can be computationally intensive.

The remainder of the paper is organized as follows. In Section 2, we describe an additive hazards model of interest, along with notation to be used throughout the paper. For statistical inference, we propose partial likelihood estimation in Section 3 and consider maximum likelihood estimation in Section 4. Computational issues are addressed in Section 5. In Section 6, a related, more parsimonious additive hazards model is proposed. Simulation experiments and a red pine data example are reported in Section 7 to demonstrate the applicability of the partial likelihood and compare with maximum likelihood estimation. Conclusions and discussion are given in Section 8. Technical details including proofs of theorems are provided in the Appendix A–C.

2. ADDITIVE HAZARDS REGRESSION MODEL

For subject i ($i = 1, \ldots, n$) at time $t > 0$ in the kth follow-up period ($k = 1, 2, \ldots, K$), consider a continuous-time additive hazards model,

$$\lambda_{k,i}(t) = Y_{k,i}\{\lambda_0(t) + \psi' Z_{k,i}(t)\}, \tag{1}$$

where ψ is a q-dimensional vector of unknown regression coefficients and $Z_{k,i}(\cdot)$ is a q-dimensional vector of covariates which are possibly time-varying and assumed to be known up to time t. The non-negative baseline hazard function $\lambda_0(t)$ is left unspecified to provide more flexibility than a fully parametric approach. Further, $Y_{k,i} = 1$ if subject i is event free at the $(k-1)$th monitoring time and thus is at risk during the follow-up period k, and $Y_{k,i} = 0$ otherwise. Without loss of generality, we assume all subjects are event free at the start of the study and thus at risk in the first follow-up period ($k = 1$).

Moreover, we let $c_0 = 0 < c_1 < c_2 < \cdots < c_K < \infty$ denote the monitoring times. We let T_i denote the exact time of event occurrence, which can be observed to either surpass the last monitoring time c_K or to lie in $[c_{k-1}, c_k)$ for some $k \in \{1, \ldots, K\}$. Hence, the observations comprise of $\{c_k, Z_{k,i}(\cdot), \delta_{k,i}\}$, where $\delta_{k,i} = I(T_i \geq c_k)$ denotes the event status, $i = 1, \ldots, n_k$, $n_k = \sum_{i=1}^{n} Y_{k,i}$ is the total number of subjects at risk during the kth follow-up period, and $k = 1, \ldots, K$.

3. PARTIAL LIKELIHOOD ESTIMATION

3.1 $K = 1$: Current status data

We begin by considering current status data ($K = 1$) with one monitoring time. As is common in environmental and ecological studies, the monitoring time is assumed to be pre-determined (i.e. fixed) here. This is in contrast to the setup in [12] where the one monitoring time is assumed to be a random point in time.

The probability of no event between c_0 and c_1 is

$$p_{1,i}(\theta) = \exp(-\alpha_1 - \psi' \tilde{Z}_{1,i}), \tag{2}$$

where

$$\alpha_1 = \int_{c_0}^{c_1} \lambda_0(t)\mathrm{d}t \quad \text{and} \quad \tilde{Z}_{1,i} = \int_{c_0}^{c_1} Z_{1,i}(t)\mathrm{d}t,$$

with $\theta = [\alpha_1, \psi']'$. The probability $p_{1,i}(\theta)$ has a multiplicative form with $\exp(-\alpha_1)$ serving as the baseline hazard [12]. For the unspecified $\lambda_0(\cdot)$, we apply a partial likelihood,

$$\begin{aligned} \mathrm{pl}_1(\psi) &= \prod_{i=1}^{n} \left\{ \frac{p_{1,i}(\theta)}{\sum_{j=1}^{n} p_{1,j}(\theta)} \right\}^{\delta_{1,i}} \\ &= \prod_{i=1}^{n} \left\{ \frac{\exp(-\psi' \tilde{Z}_{1,i})}{\sum_{j=1}^{n} \exp(-\psi' \tilde{Z}_{1,j})} \right\}^{\delta_{1,i}}, \end{aligned} \tag{3}$$

to obtain an estimate of ψ under the assumption that the outcomes are independent conditional on the history of all the covariates in the follow-up period $[c_0, c_1)$ [3, 4].

In particular, consider the log-partial likelihood function,

$$\ell_1(\psi) = \sum_{i=1}^{n} \delta_{1,i} \left[-\psi' \tilde{Z}_{1,i} - \log\left\{ \sum_{j=1}^{n} \exp(-\psi' \tilde{Z}_{1,j}) \right\} \right],$$

and obtain a maximum partial likelihood estimate $\hat{\psi}_{\mathrm{pl}} = \mathrm{argmax}_{\psi} \ell_1(\psi)$. With

$$S_1^{(l)}(\psi) = n^{-1} \sum_{j=1}^{n} (-1)^l \tilde{Z}_{1,j}^{\otimes l} \exp\{-\psi' \tilde{Z}_{1,j}\},$$

$l = 0, 1, 2$, and for a vector a, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa'$, we formally establish the large-sample properties of $\hat{\psi}_{\mathrm{pl}}$ as follows.

Theorem 3.1. *Under conditions* (a)–(e) *in Appendix* A, *$\hat{\psi}_{\mathrm{pl}}$, maximizing $\ell_1(\psi)$, is consistent, and $n^{1/2}(\hat{\psi}_{\mathrm{pl}} - \psi)$ converges in distribution to a normal variable with mean zero and a covariance matrix that can be consistently estimated by*

$$\hat{\Sigma}_1(D) = \hat{I}_1(\hat{\psi}_{\mathrm{pl}})^{-1}\hat{D}_1(\hat{\psi}_{\mathrm{pl}})\hat{I}_1(\hat{\psi}_{\mathrm{pl}})^{-1},$$

where

$$\hat{I}_1(\psi) = n^{-1}\sum_{i=1}^{n}\delta_{1,i}\left\{\frac{S_1^{(2)}(\psi)}{S_1^{(0)}(\psi)} - \frac{S_1^{(1)}(\psi)^{\otimes 2}}{S_1^{(0)}(\psi)^2}\right\},$$

$$\hat{D}_1(\psi) = n^{-1}\sum_{i=1}^{n}\left[\left\{-\tilde{Z}_{1,i} - \frac{S_1^{(1)}(\psi)}{S_1^{(0)}(\psi)}\right\} \times \{\delta_{1,i} - \hat{\eta}_1\exp(-\psi'\tilde{Z}_{1,i})\}\right]^{\otimes 2},$$

and $\hat{\eta}_1 = \sum_{i=1}^{n}\delta_{1,i}/\sum_{i=1}^{n}\exp(-\psi'\tilde{Z}_{1,i})$ is a consistent estimate of $\exp(-\alpha_1)$.

Note that, for variance estimation, one may consider the inverse of the observed information matrix $\hat{I}_1(\hat{\psi}_{\mathrm{pl}})^{-1}$. However, this is only appropriate when the monitoring time is randomly assigned as in [12], independent of the time of event T_i, since the variance of the score function $\mathrm{var}\{\partial\ell_1(\psi)/\partial\psi\}$ is the same as the information matrix $E\{-\partial^2\ell_1(\psi)/\partial\psi\partial\psi'\}$. When the monitoring times are predetermined as is common in environmental and ecological monitoring programs, this argument may not hold since the second term of

$$\begin{aligned}
&\mathrm{var}\{\partial\ell_1(\psi)/\partial\psi\}\\
&= \sum_{i=1}^{n}E\left[\left\{-\tilde{Z}_{1,i} - \frac{S_1^{(1)}(\psi)}{S_1^{(0)}(\psi)}\right\}^{\otimes 2}p_{1,i}(\theta)\{1 - p_{1,i}(\theta)\}\right]\\
&= E\left\{-\frac{\partial^2\ell_1(\psi)'}{\partial\psi\partial\psi}\right\}\\
&\quad - \sum_{i=1}^{n}E\left[\left\{-\tilde{Z}_{1,i} - \frac{S_1^{(1)}(\psi)}{S_1^{(0)}(\psi)}\right\}^{\otimes 2}p_{1,i}^2(\theta)\right]
\end{aligned}$$

is generally not negligible. Thus we need to develop an alternative, more precise variance estimation for $n^{1/2}(\hat{\psi}_{\mathrm{pl}} - \psi)$.

One possibility is

$$\Sigma_1(C) = \hat{I}_1(\hat{\psi}_{\mathrm{pl}})^{-1}\hat{C}_1(\hat{\psi}_{\mathrm{pl}})\hat{I}_1(\hat{\psi}_{\mathrm{pl}})^{-1},$$

where

$$\hat{C}_1(\psi) = \hat{I}_1(\psi) - n^{-1}\sum_{i=1}^{n}\left\{-\tilde{Z}_{1,i} - \frac{S_1^{(1)}(\psi)}{S_1^{(0)}(\psi)}\right\}^{\otimes 2} \times \eta_1^2\exp(-2\psi'\tilde{Z}_{1,i}), \tag{4}$$

is a corrected variance estimation of the score function [16]. Computation of $\hat{C}_1(\cdot)$ can be challenging, as estimation of η_1 may not be straightforward.

Remarkably, when subjects are monitored only once at a fixed monitoring time, one can treat $p_{1,i}(\theta)$ as the intensity of a discrete-time counting process $\tau_i(c_k)$ indexed by $c_k \in \{c_0, c_1\}$ with $\tau_i(c_0) = 0$ and $\tau_i(c_1) = \delta_{1,i}$. The partial likelihood (3) is thus equivalent to the approximate partial likelihood by [2] for handling tied survival time at c_1 under a proportional hazards model [9, 16]. Hence the score function,

$$\begin{aligned}
\partial\ell_1(\psi)/\partial\psi &= \sum_{i=1}^{n}\left\{-\tilde{Z}_{1,i} - \frac{S_1^{(1)}(\psi)}{S_1^{(0)}(\psi)}\right\}\delta_{1,i}\\
&= \sum_{i=1}^{n}\left\{-\tilde{Z}_{1,i} - \frac{S_1^{(1)}(\psi)}{S_1^{(0)}(\psi)}\right\}\{\delta_{1,i} - p_{1,i}(\theta)\},
\end{aligned}$$

is an unbiased estimating function for ψ since $E\{\delta_{1,i} - p_{1,i}(\theta)\} = 0$. Therefore, instead of (4), an empirical estimate $\hat{D}_1(\hat{\psi}_{\mathrm{pl}})$ can be used to estimate $\mathrm{var}\{n^{-1/2}\partial\ell_1(\psi)/\partial\psi\}$. The variance estimation of $n^{1/2}(\hat{\psi}_{\mathrm{pl}} - \psi)$ thus becomes $\hat{\Sigma}_1(D) = \hat{I}_1(\hat{\psi}_{\mathrm{pl}})^{-1}\hat{D}_1(\hat{\psi}_{\mathrm{pl}})\hat{I}_1(\hat{\psi}_{\mathrm{pl}})^{-1}$. The robust estimate $\hat{\Sigma}_1(D)$ is more practical, since it can be attained from most statistical software packages that have the capability to implement Breslow's approximate partial likelihood with robust variance estimation. For this reason, we will focus on $\hat{\Sigma}_1(D)$ here.

3.2 $K > 1$: Multiple monitoring times

We now extend the methodology for current status data with one fixed monitoring time ($K = 1$) to multiple monitoring times ($K > 1$). Let $\mathcal{H}_{c_k} = \sigma\{\tilde{Z}_l, Y_l, \delta_{l-1} : l = 1, \ldots, k\}$ denote history up to c_k, which is a σ-algebra generated by the covariate processes $\tilde{Z}_l = [\tilde{Z}_{l,1}', \ldots, \tilde{Z}_{l,n}']'$, the at-risk processes $Y_l = [Y_{l,1}, \ldots, Y_{l,n}]'$, and the event history $\delta_{l-1} = [\delta_{l-1,1}, \ldots, \delta_{l-1,n}]'$, for $l = 1, \ldots, k$. Consider a discrete-time counting process $\tau_i(c_k) \equiv \sum_{l=1}^{k}Y_{l,i}\delta_{l,i}$, $k = 1, \ldots, K$, and $\tau_i(c_0) = 0$. We assume that $\tau_i(c_k)$ is adapted to $\mathcal{H}_{c_k}$ for each i and has a jump size $+1$ if no event occurred in $[c_{k-1}, c_k)$. That is, $\tau_i(c_k)$ can have multiple jumps. Conditional on the history $\mathcal{H}_{c_k}$, we assume that τ_i can independently have an increment at c_k with intensity,

$$Y_{k,i}p_{k,i}(\theta) = \Pr\{\tau_i(c_k) - \tau_i(c_k-) = 1|\mathcal{H}_{c_k}\}, \tag{5}$$

for $k = 1, \ldots, K$ and thus acts as a counting process with recurrent events.

Suppose that subject i is at risk in the kth follow-up period, the probability of no event between c_{k-1} and c_k under model (1) is $\exp(-\alpha_k - \psi'\tilde{Z}_{k,i})$, where

$$\alpha_k = \int_{c_{k-1}}^{c_k}\lambda_0(t)\mathrm{d}t, \quad \text{and} \quad \tilde{Z}_{k,i} = \int_{c_{k-1}}^{c_k}Z_{k,i}(t)\mathrm{d}t. \tag{6}$$

Equation (6) leads to a proportional hazards model,

$$p_{k,i}(\theta) = \exp(-\alpha_k - \psi'\tilde{Z}_{k,i}), \tag{7}$$

where $\theta = [\alpha_1, \ldots, \alpha_K, \psi']'$. Unlike $p_{1,i}(\theta)$ in (2), $p_{k,i}(\theta)$ has a different baseline hazard $\exp(-\alpha_k)$ for different k. Thus, we propose a stratified analysis on the discrete-time counting process τ_i [17]. That is, conditional on $\mathcal{H}_{c_k}$, we assume independent increment as (5) and that the event status $\delta_{k,i}$ are independent for $i = 1, \ldots, n_k$. A stratified partial likelihood can be written as a product of partial likelihood within a stratum,

$$\text{(8)} \quad \text{pl}_K(\psi) = \prod_{k=1}^{K} \prod_{i=1}^{n} \left\{ \frac{p_{k,i}(\theta)}{\sum_{j=1}^{n} Y_{k,j} p_{k,j}(\theta)} \right\}^{Y_{k,i}\delta_{k,i}} = \prod_{k=1}^{K} \prod_{i=1}^{n} \left\{ \frac{\exp(-\psi' \tilde{Z}_{k,i})}{\sum_{j=1}^{n} Y_{k,j} \exp(-\psi' \tilde{Z}_{k,j})} \right\}^{Y_{k,i}\delta_{k,i}},$$

where the follow-up periods are considered as the strata.

It follows that the log-partial likelihood function is

$$\ell_K(\psi) = \sum_{k=1}^{K} \sum_{i=1}^{n} Y_{k,i}\delta_{k,i} \left[-\psi' \tilde{Z}_{k,i} - \log\left\{ \sum_{j=1}^{n} Y_{k,j} \exp(-\psi' \tilde{Z}_{k,j}) \right\} \right]$$

and can be maximized to obtain a maximum partial likelihood estimate $\hat{\psi}_{\text{pl}}$ for ψ. In particular, when K is fixed, one can show that $\hat{\psi}_{\text{pl}}$ is a consistent estimate of ψ and asymptotically normal. Variance estimation has a sandwich form, as it needs to be corrected due to the discreteness of τ_i. Define $S_k^{(l)}(\psi) = n^{-1} \sum_{j=1}^{n} (-1)^l Y_{k,j} \tilde{Z}_{k,j}^{\otimes l} \exp\{-\psi' \tilde{Z}_{k,j}\}$, $l = 0, 1, 2$. We formally establish the large-sample properties of $\hat{\psi}_{\text{pl}}$ as follows.

Theorem 3.2. *Under conditions* (a)–(e) *in Appendix* A, $\hat{\psi}_{\text{pl}}$, *maximizing* $\ell_K(\psi)$, *is consistent, and* $n^{1/2}(\hat{\psi}_{\text{pl}} - \psi)$ *converges in distribution to a normal variable with mean zero and a covariance matrix that can be consistently estimated by*

$$\hat{\Sigma}_K(D) = \hat{I}_K(\hat{\psi}_{\text{pl}})^{-1} \hat{D}_K(\hat{\psi}_{\text{pl}}) \hat{I}_K(\hat{\psi}_{\text{pl}})^{-1},$$

where $\hat{I}_K(\psi) = \sum_{k=1}^{K} \hat{I}_k(\psi)$ *and* $\hat{D}_K(\psi) = \sum_{k=1}^{K} \hat{D}_k(\psi)$ *with*

$$\hat{I}_k(\psi) = n^{-1} \sum_{i=1}^{n} Y_{k,i}\delta_{k,i} \left\{ \frac{S_k^{(2)}(\psi)}{S_k^{(0)}(\psi)} - \frac{S_k^{(1)}(\psi)^{\otimes 2}}{S_k^{(0)}(\psi)^2} \right\},$$

$$\hat{D}_k(\psi) = n^{-1} \sum_{i=1}^{n} Y_{k,i} \left[\left\{ -\tilde{Z}_{k,i} - \frac{S_k^{(1)}(\psi)}{S_k^{(0)}(\psi)} \right\} \times \{\delta_{k,i} - \hat{\eta}_k \exp(-\psi' \tilde{Z}_{k,i})\} \right]^{\otimes 2},$$

and $\hat{\eta}_k = \sum_{i=1}^{n} Y_{k,i}\delta_{k,i} / \sum_{i=1}^{n} Y_{k,i} \exp(-\psi' \tilde{Z}_{k,i})$ *is a consistent estimate of* $\exp(-\alpha_k)$, $k = 1, \ldots, K$.

Theorem 3.2 is an extension of Theorem 3.1 for $K = 1$ to the case $K > 1$. We will show, in the following section, that $\hat{\eta}_k$ $(k = 1, \ldots, K)$ is a profile estimator for $\exp(-\alpha_k)$ under a profile likelihood. Consistency and asymptotic normality of $\hat{\eta}_k$ will be established in Section 6.

4. MAXIMUM LIKELIHOOD ESTIMATION

A maximum likelihood approach is also possible here, as the data are grouped survival with T_i either greater than c_K or between c_{k-1} and c_k, for $k = 1, \ldots, K$. Let $\bar{F}_i(c_x; \theta) = \prod_{k=1}^{x} p_{k,i}(\theta)$ denote the probability of no event up to the monitoring time c_x for subject i if $x \in \{1, \ldots, K\}$ and $\bar{F}_i(c_x) = 0$ if $c_x > c_K$, under the independent increment assumption in (5). A full likelihood function, therefore, is given by

$$L(\theta) = \prod_{i=1}^{n} \{\bar{F}_i(c_{x-1}; \theta) - \bar{F}_i(c_x; \theta)\} = \prod_{k=1}^{K} \prod_{i=1}^{n} \left[p_{k,i}(\theta)^{\delta_{k,i}} \{1 - p_{k,i}(\theta)\}^{1-\delta_{k,i}} \right]^{Y_{k,i}}.$$

The score functions are

$$\text{(9)} \quad \frac{\partial \log L(\theta)}{\partial \alpha_k} = \sum_{i=1}^{n} Y_{k,i} \{1 - p_{k,i}(\theta)\}^{-1} \{p_{k,i}(\theta) - \delta_{k,i}\},$$

for $k = 1, \ldots, K$, and

$$\text{(10)} \quad \frac{\partial \log L(\theta)}{\partial \psi} = \sum_{k=1}^{K} \sum_{i=1}^{n} Y_{k,i} \tilde{Z}_{k,i} \{1 - p_{k,i}(\theta)\}^{-1} \times \{p_{k,i}(\theta) - \delta_{k,i}\},$$

where $\theta = [\alpha_1, \ldots, \alpha_K, \psi']'$ as in (7).

We denote the maximum likelihood estimate (MLE) as $\hat{\theta}_m = [\hat{\alpha}_1, \ldots, \hat{\alpha}_K, \hat{\psi}_m']'$. By conventional maximum likelihood arguments, $\hat{\theta}_m$ is consistent, asymptotically normal, and efficient. Furthermore $\hat{I}_m(\hat{\theta}_m)^{-1}$ consistently estimates the variance of $n^{1/2}(\hat{\theta}_m - \theta)$, where $\hat{I}_m(\theta) = -\partial^2 \log L(\theta) / \partial\theta\partial\theta'$ is the observed information matrix. In particular,

$$\text{(11)} \quad \hat{I}_m(\theta) = \sum_{k=1}^{K} \sum_{i=1}^{n} Y_{k,i} \tilde{Z}_{k,i}^{*\otimes 2} \{1 - p_{k,i}(\theta)\}^{-2} p_{k,i}(\theta)(1 - \delta_{k,i}),$$

where $\tilde{Z}_{k,i}^* = [1_k', \tilde{Z}_{k,i}']'$ and 1_k is a K-dimensional vector with 1 in the kth element and 0 otherwise.

Maximizing $L(\theta)$, or equivalently, solving the estimating equations (9) and (10) simultaneously, is in principle feasible. However, in practice, it is feasible only when the number of monitoring times K is relatively small. When K is large, a profile estimate of ψ that treats $[\alpha_1, \ldots, \alpha_K]'$ as nuisance would be more manageable computationally. Dropping the term $\{1-p_{k,i}(\theta)\}^{-1}$ in (9) leads to an estimator for $\exp(-\alpha_k)$ as

$$\exp(-\hat{\alpha}_k) = \sum_{i=1}^{n} Y_{k,i}\delta_{k,i} \Big/ \sum_{i=1}^{n} Y_{k,i}\exp(-\psi'\tilde{Z}_{k,i}), \tag{12}$$

when fixing ψ [12]. Replacing $\exp(-\alpha_k)$ in (10) with $\exp(-\hat{\alpha}_k)$ and again dropping the term $\{1-p_{k,i}(\theta)\}^{-1}$, the score function becomes

$$\sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i}\delta_{k,i}\left\{-\tilde{Z}_{k,i} - \frac{\sum_{j=1}^{n} Y_{k,j}(-\tilde{Z}_{k,j})\exp(-\psi'\tilde{Z}_{k,j})}{\sum_{j=1}^{n} Y_{k,j}\exp(-\psi'\tilde{Z}_{k,j})}\right\}, \tag{13}$$

which is exactly the same as the stratified partial likelihood score function of (8).

The argument above brings out an interesting link between the partial likelihood and the maximum likelihood estimation. Intuitively, since the monitoring times here are fixed and the same for each subject, the partial likelihood estimate $\hat{\psi}_{\mathrm{pl}}$ is expected to be close to the maximum likelihood estimate and thus should have relatively high efficiency. However, we also observe that the partial likelihood estimation is equivalent to the one when dropping the term $\{1-p_{k,i}(\theta)\}^{-1}$ in both (9) and (10). Hence, it does not achieve the semiparametric information bound [13].

5. COMPUTATIONAL ASPECT

Maximum likelihood estimation can be computed by a Newton-Raphson algorithm. Starting from an initial value $\hat{\theta}_m^{(0)}$, we update

$$\hat{\theta}_m^{(s+1)} = \hat{\theta}_m^{(s)} + \hat{I}_m(\hat{\theta}_m^{(s)})^{-1} U_m(\hat{\theta}_m^{(s)}),$$

where U_m is the score function defined in (9) and (10), and $\hat{I}_m$ is the observed information matrix defined in (11). We iterate until convergence. However, the observed information matrix $\hat{I}_m$ may be difficult to invert when the number of follow-up periods K is large. One way to deal with this is to use a symmetric 2×2 partition matrix

$$\hat{I}_m^{-1} = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}^{-1} = \begin{bmatrix} I_{11}^{-1} + AB^{-1}A' & -AB^{-1} \\ -B^{-1}A' & B^{-1} \end{bmatrix},$$

where $A = I_{11}^{-1}I_{12}$ and $B = I_{22} - I_{21}A$ [15]. More specifically, let $d_{k,i} = (1-p_{k,i})^{-2}p_{k,i}(1-\delta_{k,i})$, where $p_{k,i} = p_{k,i}(\theta)$. Then $I_{11} = \mathrm{diag}\{\sum_{i=1}^{n} Y_{k,i}d_{k,i}\}$ is a $K\times K$ diagonal matrix, and $I_{11}^{-1} = \mathrm{diag}\{(\sum_{i=1}^{n} Y_{k,i}d_{k,i})^{-1}\}$. Also,

$$I_{21} = \left[\sum_{i=1}^{n} Y_{1,i}d_{1,i}\tilde{Z}_{1,i}, \ldots, \sum_{i=1}^{n} Y_{K,i}d_{K,i}\tilde{Z}_{K,i}\right]$$

is a $p\times K$ matrix with $I_{12} = I_{21}'$ and $I_{22} = \sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i}d_{k,i}\tilde{Z}_{k,i}^{\otimes 2}$ is a $p\times p$ positive definite matrix. Thus, it suffices to invert the smaller matrix B.

As mentioned before, an advantage of partial likelihood estimation is that an estimate from (8) can be attained by standard statistical software packages. The following preprocessing is needed, however.

- Step 1: Compute $\tilde{Z}_{k,i} = \int_{c_{k-1}}^{c_k} Z_{k,i}(t)\mathrm{d}t$ as a time-independent covariate.
- Step 2: Assign $(c_k - c_{k-1})$ as the "survival" time for those subjects who are at risk in the kth follow-up period (i.e. $Y_{k,i} = 1$).
- Step 3: Assign those subjects who are event free at the kth monitoring time to the "failure" category (i.e. survival time observed) and those subjects who have had an event at the kth monitoring time to the "censored" category (i.e. survival time censored).
- Step 4: Select the option for stratified analysis and include a variable for strata in the data set.
- Step 5: Select the option for Breslow's estimation for tied survival time with robust variance estimation.

Step 1 can be easily implemented when $Z_{k,i}$ is time invariant, since $\tilde{Z}_{k,i} = (c_k - c_{k-1})Z_{k,i}$. Step 3 needs to be implemented with care. Keep in mind that for those subjects who are at risk in the kth follow-up period and event free at the end, the event indicator $\delta_{k,i}$ is 1 so their "survival time" $(c_k - c_{k-1})$ is exactly observed. Therefore, we need to assign an event indicator to those who actually have no event occurrence. As for Step 5, since our partial likelihood is equivalent to the Breslow's approximate partial likelihood when dealing with tied survival time, selecting the option of Breslow's method is needed, as well as a robust variance estimation. However, some packages, such as `coxph()` in `R`, set Efron's estimating equation [6] for tied survival time as the default, which needs to be changed as Efron's estimation does not give the same score function as (13). Finally, the total number of observations in the data set summed over each stratum is $\sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i}$, which could be smaller than Kn.

Taking `R` for example, let `zz` be assigned to a one-dimensional covariate $\tilde{Z}_{k,i}$ in the data set, `d_k` to $(c_k - c_{k-1})$, `delta` to $\delta_{k,i}$, and `monitoring` to indicators for strata, the following code suffices to provide the desired results.

```
coxph(Surv(d_k,delta)~zz+strata(monitoring),
      method="breslow",robust=T)
```

Note that, the coefficient estimates from the statistical packages is negative of our proposed partial likelihood estimates $\hat{\psi}_{\mathrm{pl}}$, as $-\psi$ in (8) is the reparameterized coefficient for the time-independent covariate $\tilde{Z}_{k,i}$. However, the variance (or standard deviation) estimates from packages can be directly applied since the negative sign does not affect the variance estimation.

6. INFERENCE FOR BASELINE HOMOGENEITY

6.1 Periodic model

Thus far we have assumed that the baseline hazard $\lambda_0(\cdot)$ varies by time t, which is defined as the time elapsed from the start c_0. A common scenario is that the baseline hazard is either constant over time or is periodically renewed at certain points of time [18]. This gives rise to a different model specification, which nonetheless is a special case of our previous model. Suppose the duration of each follow-up period is the same and the baseline hazard function is periodic and renewed at the time of monitoring, an additive hazards model, similar to (1), is of interest,

$$\lambda_{k,i}(t) = Y_{k,i}\{\lambda_0(t - c_{k-1}) + \psi' Z_{k,i}(t)\}. \tag{14}$$

When the duration of each period is equal with $c_k - c_{k-1} \equiv c$ $(k = 1, \ldots, K)$, integration of the baseline hazard in each stratum, $\alpha_1, \ldots, \alpha_K$, is the same across strata, as

$$\alpha_k = \int_{c_{k-1}}^{c_k} \lambda_0(t - c_{k-1})\mathrm{d}t = \int_0^c \lambda_0(u)\mathrm{d}u,$$

which is free of k. We call the resulting model a homogeneous model, as versus a heterogeneous model where at least one of the α_k's is different.

Let $\alpha_k = \alpha$, for $k = 1, \ldots, K$. The model parameter θ only consists of α and ψ. The score functions hence become

$$\frac{\partial \log L(\theta)}{\partial \alpha} = \sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i}\{1 - p_{k,i}(\theta)\}^{-1}\{p_{k,i}(\theta) - \delta_{k,i}\}, \tag{15}$$

and

$$\frac{\partial \log L(\theta)}{\partial \psi} = \sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i}\tilde{Z}_{k,i}\{1 - p_{k,i}(\theta)\}^{-1} \times \{p_{k,i}(\theta) - \delta_{k,i}\}. \tag{16}$$

Dropping the term $\{1 - p_{k,i}(\theta)\}^{-1}$ in both (15) and (16), a profile estimate of $\exp(-\alpha)$ can be attained as

$$\exp(-\hat{\alpha}) = \sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i}\delta_{k,i} \Big/ \sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i}\exp(-\psi'\tilde{Z}_{k,i}).$$

The estimating function for ψ is

$$\sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i}\delta_{k,i} \times \left\{-\tilde{Z}_{k,i} - \frac{\sum_{l=1}^{K}\sum_{j=1}^{n} Y_{l,j}(-\tilde{Z}_{l,j})\exp(-\psi'\tilde{Z}_{l,j})}{\sum_{l=1}^{K}\sum_{j=1}^{n} Y_{l,j}\exp(-\psi'\tilde{Z}_{l,j})}\right\},$$

which is in fact a partial likelihood function when treating all subjects and outcomes as independent. In other words, when the baseline hazard function is periodic, the model is more parsimonious, and either the partial likelihood or the full likelihood can be simplified. The results above are applicable to the model with constant baseline hazard as it is a special case of (14).

6.2 Test for homogeneity

When a homogeneous model is true, a heterogeneous model obviously over-stratifies and hence is less efficient. We derive a formal test in order to select between these two competing models with the null hypothesis $H_0 : \alpha_1 = \cdots = \alpha_K$. Two quadratic tests are possible, namely likelihood ratio test when a maximum likelihood approach is performed, and a direct comparison among $\exp(-\hat{\alpha}_k)$ for $k = 1, \ldots, K$ under a partial likelihood approach.

In the maximum likelihood approach, the homogeneous model is nested within the heterogeneous one. Thus it is natural to use a likelihood ratio test with predetermined size, where the likelihood ratio is

$$Q_n^{(1)} = L(\hat{\alpha}_1, \ldots, \hat{\alpha}_K, \hat{\psi}_{m_K}) \big/ L(\hat{\alpha}, \hat{\psi}_{m_1}),$$

where $[\hat{\alpha}_1, \ldots, \hat{\alpha}_K, \hat{\psi}'_{m_K}]'$ are the solutions to (9) and (10), while $[\hat{\alpha}, \hat{\psi}'_{m_1}]'$ are the solutions to (15) and (16). Under H_0, it is well known that $-2\log Q_n^{(1)}$ converges in distribution to a χ^2-variate with $K - 1$ degrees of freedom as $n \to \infty$ [20]. Thus, we reject H_0 at the level of a $(0 < a < 1)$ when $-2\log Q_n^{(1)} > \chi^2_{K-1,a}$, where $\chi^2_{K-1,a}$ is the $(1-a)$-percentile of a χ^2-distribution with $K - 1$ degrees of freedom.

With the partial likelihood approach, a counterpart likelihood ratio test is not obvious, since the α_k's are treated as nuisance. However, a direct comparison among $\exp(-\hat{\alpha}_k)$, $k = 1, \ldots, K$, from profile estimates (12) is possible. Let $\eta_k = \exp(-\alpha_k)$, $\eta = [\eta_1, \ldots, \eta_K]'$, and $\eta_\triangle = [\eta_1 - \eta_2, \ldots, \eta_{K-1} - \eta_K]'$. Denote $\hat{\eta}_k(y) = \exp(-\hat{\alpha}_k)|_{\psi=y}$ in (12), $\hat{\eta}(y) = [\hat{\eta}_1(y), \ldots, \hat{\eta}_K(y)]'$, and $\hat{\eta}^* = \hat{\eta}(\hat{\psi}_{\mathrm{pl}})$. We establish the large-sample properties of $\hat{\eta}^*$ in the following theorem.

Theorem 6.1. *Under conditions* (a), (b), *and* (d) *in Appendix* A, $\hat{\eta}^*$ *is a consistent estimate of* η, *and* $n^{1/2}(\hat{\eta}^* - \eta)$ *converges in distribution to a normal variable with mean zero and a covariance matrix* Ω. *The matrix* Ω *can be consistently estimated by a symmetric matrix* $\hat{\Omega}$, *where the* (u, v)*th entry* $(u, v \in \{1, \ldots, K\}$ *and* $u \leq v)$ *of* $\hat{\Omega}$ *is* $B_u'\hat{\Sigma}_K(D)B_v + \hat{\sigma}^2_{uv}\{S_u^{(0)}(\hat{\psi}_{\mathrm{pl}})S_v^{(0)}(\hat{\psi}_{\mathrm{pl}})\}^{-1}$, *where* $B_u = \bar{\delta}_u S^{(1)}(\hat{\psi}_{\mathrm{pl}})/S^{(0)}(\hat{\psi}_{\mathrm{pl}})^2$ *with* $\bar{\delta}_u = n^{-1}\sum_{i=1}^{n} Y_{u,i}\delta_{u,i}$ *and*

$$\hat{\sigma}^2_{uv} = n^{-1}\sum_{i=1}^{n} Y_{u,i}Y_{v,i}\{\delta_{u,i} - \hat{\eta}_u(\hat{\psi}_{\mathrm{pl}})\exp(-\hat{\psi}'_{\mathrm{pl}}\tilde{Z}_{u,i})\} \times \{\delta_{v,i} - \hat{\eta}_v(\hat{\psi}_{\mathrm{pl}})\exp(-\hat{\psi}'_{\mathrm{pl}}\tilde{Z}_{v,i})\}.$$

Table 1. Simulation results of stratified partial likelihood and maximum likelihood when the true model has a homogeneous baseline. †: multiplied by 10^4

		Partial Likelihood with Stratified Analysis								Test for homogeneity
		ψ_1				ψ_2				Size(%) $Q_n^{(2)}$
K	m	Bias†	EV†	AVE†	CP	Bias†	EV†	AVE†	CP	
1	10	14.24	43.92	44.79	0.931	−11.99	14.29	12.78	0.937	–
	20	6.57	11.89	12.04	0.949	−2.66	1.87	2.00	0.948	–
	30	3.48	5.30	5.58	0.945	4.90	0.87	0.82	0.952	–
3	10	16.34	24.82	24.31	0.946	−12.27	3.16	3.17	0.952	3.7
	20	3.03	7.13	7.16	0.957	−3.02	0.56	0.55	0.957	4.6
	30	2.10	3.57	3.43	0.950	−1.66	0.25	0.25	0.948	3.7
6	10	16.86	17.08	16.28	0.947	−10.88	2.72	2.48	0.930	10.1
	20	−3.69	4.91	5.16	0.953	−2.07	0.48	0.45	0.943	5.8
	30	2.08	2.51	2.49	0.946	−3.04	0.20	0.20	0.953	4.9

		Maximum Likelihood with Stratified Analysis								Test for homogeneity
		ψ_1				ψ_2				Size(%) $Q_n^{(1)}$
K	m	Bias†	EV†	AVE†	CP	Bias†	EV†	AVE†	CP	
1	10	122.21	24.79	55.59	0.990	−5.24	6.07	16.38	0.998	–
	20	20.44	8.02	11.84	0.983	9.78	1.68	1.93	0.970	–
	30	8.23	4.44	5.32	0.980	1.23	0.75	0.75	0.951	–
3	10	47.67	14.34	25.43	0.989	−31.48	1.78	3.40	0.993	3.0
	20	10.15	5.22	6.69	0.978	−6.89	0.40	0.51	0.974	3.4
	30	2.55	2.95	3.15	0.971	−2.66	0.20	0.22	0.963	4.8
6	10	40.88	11.24	18.19	0.984	−49.47	1.54	2.70	1.000	3.5
	20	12.16	4.22	4.79	0.974	−8.86	0.35	0.42	0.967	5.8
	30	−5.52	2.14	2.25	0.956	−2.86	0.17	0.18	0.964	5.7

Thus, under H_0, the test statistic

$$Q_n^{(2)} = n(e'\hat{\eta}^*)'\hat{V}^{-1}e'\hat{\eta}^*$$

is a χ^2-variate with $K-1$ degrees of freedom as $n \to \infty$, where e is a $(K-1) \times K$ matrix satisfying $\eta_\triangle = e'\eta$, and $\hat{V} = e'\hat{\Omega}e$. Similarly, we reject H_0 at the level of a when $Q_n^{(2)} > \chi^2_{K-1,a}$.

7. NUMERICAL EXAMPLES

7.1 Simulation study

We conduct simulation to assess the performance of partial likelihood estimation and compare with maximum likelihood estimation. Data sets are generated according to the setup of the red pine data example described in Section 1. Suppose an $m \times m$ spatial lattice. To mimic the situation that some trees had already been colonized in the beginning of the study, we randomly generate a 0–1 indicator with probability 0.1 of *Ips* colonization by year $k = 0$. Therefore, the sample size n varies in each simulation but on average is $0.9m^2$. In year $k = 1, \ldots, K$, the number of turpentine beetle colonization is independently generated by a Poisson distribution at a rate of 0.8. The outcome of no event in year k ($\delta_{k,i} = 1$) is generated according to model (1) with probability $p_{k,i}(\theta)$ in (7) when the tree at site i ($i = 1, \ldots, n_k$) is at risk.

To account for spatial dependence, we define the kth-order neighbor of a given site as those sites that are the kth nearest neighbors. For example, the first-order neighborhood on a square lattice has the four nearest neighbors in the north, south, west, and east, and the second-order neighborhood has the four second-nearest neighbors in the northwest, southwest, northeast, and southeast, etc. We let the number of *Ips* colonization in the neighborhood in year $(k-1)$ be a covariate for year k, which is summed up to the fifth-order neighbors. That is, we model spatial dependence via autoregression in space from a previous time point.

We set the total number of strata to $K = 1, 3, 6$ and the grid size to $m = 10, 20, 30$. The cumulative baseline hazard is set to $\alpha_k = 0.1k$ under heterogeneity, and a constant $\alpha_k = 0.1$ under homogeneity. The coefficients are set to $\psi = [\psi_1, \psi_2]' = [0.03, 0.03]'$. A total of 1,000 repeated samples are simulated for each combination of K and m.

Table 1 compares the simulation results using partial and maximum likelihood analysis under a homogeneous model. We report the bias of estimation (Bias) defined as the average of the replicated estimates minus the true value, the empirical variance (EV) defined as the sample variance of

Table 2. Simulation results of stratified and non-stratified partial likelihood when the true model has a heterogeneous baseline. †: multiplied by 10^4

Partial Likelihood with Stratified Analysis									
		ψ_1				ψ_2			
K	m	Bias†	EV†	AVE†	CP	Bias†	EV†	AVE†	CP
1	10	7.44	46.49	44.72	0.920	11.41	13.64	12.67	0.943
	20	17.42	12.54	12.19	0.941	4.89	1.86	1.97	0.949
	30	−17.02	5.00	5.51	0.957	2.03	0.89	0.82	0.939
3	10	11.66	32.15	31.03	0.944	−11.51	4.22	4.36	0.959
	20	9.43	8.79	8.79	0.949	−8.41	0.71	0.73	0.950
	30	10.00	4.20	4.12	0.937	−1.95	0.33	0.33	0.946
6	10	21.34	31.00	28.25	0.934	−5.03	4.48	3.99	0.928
	20	−2.90	8.20	8.21	0.943	−4.85	0.70	0.68	0.946
	30	1.05	3.90	3.85	0.943	−2.33	0.31	0.31	0.944

Partial Likelihood with Non-Stratified Analysis									
		ψ_1				ψ_2			
K	m	Bias†	EV†	AVE†	CP	Bias†	EV†	AVE†	CP
3	10	10.63	32.40	31.29	0.945	122.31	2.59	2.34	0.885
	20	8.77	8.83	8.85	0.947	106.67	0.36	0.38	0.600
	30	10.42	4.25	4.14	0.938	105.53	0.16	0.15	0.238
6	10	27.50	30.95	29.35	0.944	80.38	2.81	2.02	0.868
	20	−4.52	8.51	8.42	0.948	66.37	0.43	0.35	0.780
	30	1.37	3.96	3.94	0.943	69.06	0.17	0.14	0.546

the replicated estimates, the average of the replicated variance estimates (AVE), and empirical coverage probability (CP) at a 95% nominal level. The size of the test for homogeneity is also reported. For maximum likelihood estimation, $Q_n^{(1)}$ is applied, while for partial likelihood estimation, $Q_n^{(2)}$ is applied. The significance level $a = 5\%$ is assigned for both tests. Table 2 summarizes the simulation results by stratified and non-stratified partial likelihood estimation under a heterogenous model to assess the effect of model misspecification.

For partial likelihood in Table 1, the estimation appears to be consistent and the variance estimation is close to the empirical variance when n is large, even when $K = 1$. However, when $K = 1$ and n is small, the partial likelihood estimation has a slightly lower empirical coverage probability than the nominal level. On the other hand, the maximum likelihood estimation appears to be consistent, but apparently over-estimates the variance of $\hat{\psi}_1$ by the information matrix when either K or n is small. The variance estimation is close to the empirical variance only when $K = 6$ and n is large. Compared with the partial likelihood that performs well for a small n, the maximum likelihood estimation needs a relatively larger sample size to achieve less bias, which suggests that the partial likelihood is more robust for a smaller sample size. The empirical size of both tests for homogeneity is close to 5% when n becomes larger, indicating that both $Q_n^{(1)}$ and $Q_n^{(2)}$ are applicable. However, when n is under 100 ($m = 10$) and $K = 6$, the empirical size of $Q_n^{(2)}$ in the partial likelihood estimation is about twice that of the true significance level. This suggests preference for $Q_n^{(1)}$ in cases where the sample size is small and subjects are followed up at a large number of monitoring times.

Furthermore, our experience is such that maximum likelihood via a Newton-Raphson algorithm can be numerically rather unstable, due to near-singularity of the information matrix $\hat{I}_m$ during the iterations toward convergence. Besides, the probability $p_{k,i}$, as a function of $\hat{\theta}_m^{(s)}$, should be bounded above by 1 in each iteration, but this can fail when the true $p_{k,i}$ is close to 1. Such failures in computing the maximum likelihood estimates occur more often when the sample size is relatively small, which explains why the empirical coverage probability is higher than the nominal level in Table 1. In contrast, the partial likelihood estimation is more stable numerically, since the restriction on the parameter space is rather minimal. In addition, the information matrix can be inverted with more ease, as its dimension is determined by the number of covariates and, unlike maximum likelihood estimation, is not affected by the number of monitoring times K.

In Table 2, when the true model is heterogeneous, parameter estimation is biased and has low empirical coverage probability for ψ_2 by a non-stratified analysis. However, the estimation of parameter ψ_1 that represents the effect of a covariate seems unaffected. This gives empirical evidence that the partial likelihood approach is robust against model misspecification for the baseline hazards. Finally, partial likelihood with stratification has consistent estimation of both the model parameter and the variance, which induces empirical coverage rates that are close to the nominal level.

Table 3. *Analysis results of the red pine data. †: standard error; ‡: compared to a χ^2 distribution with degree of freedom 4; *: significantly different from 0 at the level of 5%*

Partial Likelihood				
	Turpentine(SE†)	*Ips*(SE)	η_k(SE)	$Q_n^{(2)\ddagger}$
Stratified	0.067(0.0153)*	0.042(0.0035)*	0.984(0.0032)*	107.5
			0.995(0.0039)*	
			1.015(0.0025)*	
			0.984(0.0031)*	
			1.003(0.0029)*	
Non-Stratified	0.068(0.0153)*	0.041(0.0035)*	0.996(0.0018)*	

Maximum Likelihood				
	Turpentine(SE†)	*Ips*(SE)	η_k(SE)	$Q_n^{(1)\ddagger}$
Stratified	0.068(0.0142)*	0.038(0.0028)*	0.992(0.0019)*	25.8
			0.994(0.0016)*	
			1.000(0.0004)*	
			0.991(0.0020)*	
			0.996(0.0014)*	
Non-Stratified	0.069(0.0142)*	0.038(0.0028)*	0.995(0.0007)*	

7.2 Red pine data example

For the red pine data analysis, we set the first year of insect survey as the initial period $k = 0$ when $n_1 = 2,683$ trees had not been colonized by *Ips*. In the following 5 years, there were $n_2 = 2,599$, $n_3 = 2,508$, $n_4 = 2,484$, and $n_5 = 2,418$ trees at risk of *Ips* colonization. As described in the previous section, we let $Z_{k,i}(t) = [X_{k-,i}, N_{k-1,i}]'$, where for tree i, $X_{k-,i}$ is the number of turpentine beetles colonized at the end of year k, and $N_{k-1,i}$ is the cumulative number of *Ips* colonization in the neighborhood up to the fifth order in year $(k-1)$. The additive hazards model of interest is

$$\lambda_{k,i}(t) = Y_{k,i}\{\lambda_0(t) + \psi_1 X_{k-,i} + \psi_2 N_{k-1,i}\}.$$

Table 3 presents the analysis result by both partial likelihood and maximum likelihood estimation. We report both parameter and baseline hazard estimates with their standard errors. We also report the χ^2-test result for the null hypothesis that the model is homogeneous. The two different methods give very similar results on the parameter estimation of ψ_1. Partial likelihood has a slightly larger standard error than maximum likelihood. Regardless of the approach, colonization of the turpentine beetle is shown to have had a significant positive effect on the risk of *Ips* colonization. *Ips* colonization in the previous year in the neighbors also played a significant role in increasing the risk of *Ips* colonization in the current year. Estimates of the cumulative baseline hazards, η_k, are similar using both partial and maximum likelihood estimation. However, a departure from the homogeneous model is statistically significant due to small standard errors, even though the parameter estimates appear to be similar across the follow-up periods.

8. CONCLUSIONS AND DISCUSSION

We have considered additive hazards regression for the analysis of environmental or ecological monitoring data such that subjects are observed at multiple monitoring time points and the baseline hazard function can take on flexible forms. The covariates can be time-varying and we have used autoregression to take into account spatial dependence. We have developed partial likelihood estimation for statistical inference of the regression coefficients and the baseline hazard. Asymptotic properties, including consistency and asymptotic normality, have been established for the maximum partial likelihood estimates under suitable regularity conditions. We have proposed feasible algorithms utilizing existing statistical software packages for computation and thus our method has distinct computational advantage compared with some existing methods. We have also developed hypothesis testing for homogeneity of the baseline hazard. Using numerical examples from both simulation and an ecological monitoring study, we have demonstrated that the partial likelihood inference is feasible and has sound finite-sample properties.

Our results are comparable to the findings in [18], although [18] required that external data be available to estimate the baseline hazard function. In contrast, our approach does not require external data which may not always be available or suitable for estimating the baseline hazard in practice. Furthermore, these results are comparable to the spatial-temporal autologistic approach in [23] where spatial dependence is modeled by autoregressive terms within a given year but the computation is far more intensive due to an unknown normalizing constant in the loglikelihood function. Our approach here is also autoregression but on a previous year, which has led to much faster computation. For the red pine data example, the scientific conclusions drawn from both approaches are remarkably similar. In general, however, the results may differ; we plan to explore and compare alternative approaches to modeling spatial dependence.

We contend that partial likelihood estimation offers a viable alternative to maximum likelihood estimation for at least two reasons. One, maximum likelihood is feasible only for small number of monitoring times K. When K is large, the computing cost is substantially increased due to a large number of parameters, with one extra intercept for each additional monitoring time. One exception is when the baseline intensity is constant with only one intercept for the baseline, but this case is overly restrictive. In contrast, the number of parameters in the partial likelihood (8) is not affected by the number of monitoring times, as the baseline parameters are treated as nuisance. Two, our simulation experiments indicate that partial likelihood estimation is numerically more stable to compute in general. It also provides more robust results than maximum likelihood estimation particularly with a small number of monitoring times and/or a small number of subjects. Taken together, these results suggest that, even though maximum likelihood estimation is more efficient than the partial likelihood counterpart in theory, there are compelling reasons to consider partial likelihood estimation developed here in practice for both a small and large number of monitoring times.

Although the monitoring times are regular in our data example, they can be irregular in other studies such that the monitoring times may not be regularly-spaced or have different schemes for each subject. [14], also applied in [19], proposed an idea of mapping the irregular monitoring times to a regular grid, which can be applied to extend our method. Let $c_{k_0} < c_{k_1} < \cdots < c_{k_{m_i}}$ denote the mapped set of ordered monitoring times of the subject at the ith site with $k_0 = 0$ and $k_j \in \{1, 2, \ldots, K\}$ for $j = 1, \ldots, m_i$. A slight modification of (7), defined by

$$p_{k_j,i} = \exp\{-\alpha_{k_j} - \psi' \tilde{Z}_{k_j,i}\},$$

can again be utilized to estimate ψ with the log-likelihood function

$$\sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_{k_j,i}\{\delta_{k_j,i} \log p_{k_j,i} + (1 - \delta_{k_j,i}) \log(1 - p_{k_j,i})\},$$

where $\alpha_{k_j} = \sum_{k=k_{j-1}+1}^{k_j} \alpha_k$, and $\tilde{Z}_{k_j,i}$, $Y_{k_j,i}$, and $\delta_{k_j,i}$ are defined in a similar manner to $\tilde{Z}_{k,i}$, $Y_{k,i}$, and $\delta_{k,i}$, respectively. We conjecture that partial likelihood estimation can be devised by appropriate stratification. In addition, it would be interesting to extend our method to deal with monitoring times that are randomly assigned. We leave such extensions for future research.

APPENDIX A. NOTATION AND REGULARITY CONDITIONS

Let $\mathbb{K} = \{1, \ldots, K\}$ and $\dot{f}(\psi) \equiv \partial f / \partial \psi$. Let $\theta_0 = [\eta_1^0, \ldots, \eta_K^0, \psi_0]'$ denote the true values of $\theta = [\eta_1, \ldots, \eta_K, \psi]'$ and Ψ denote a compact closure surrounding the true parameter ψ_0. Consider the following regularity conditions.

(a) $K < \infty$ and $0 < \min_{k \in \mathbb{K}} \eta_k \le \max_{k \in \mathbb{K}} \eta_k < \infty$.

(b) For $\psi \in \Psi$, there exists continuous functions $s_k^{(l)}(\psi)$, $l = 0, 1, 2$, such that

$$\sup_{\psi \in \Psi} \max_{k \in \mathbb{K}} \|S_k^{(l)}(\psi) - s_k^{(l)}(\psi)\| \to_p 0.$$

(c) Define $R_k^{(0)}(\psi) = n^{-1} \sum_{i=1}^{n} Y_{k,i} \exp(-2\psi' \tilde{Z}_{k,i})$. There exists continuous function $r_k^{(0)}(\psi)$ such that $\|R_k^{(0)}(\psi_0) - r_k^{(0)}(\psi_0)\| \to_p 0$ for each $k \in \mathbb{K}$.

(d) For all $\psi \in \Psi$ and $k \in \mathbb{K}$, $s_k^{(1)}(\psi) = \dot{s}_k^{(0)}(\psi)$, $s_k^{(2)}(\psi) = \dot{s}_k^{(1)}(\psi)$, $s_k^{(l)}(\psi)$ are bounded for $l = 0, 1, 2$, and $s_k^{(0)}(\psi)$ is bounded away from 0. Define

$$I_k(\psi_0) = \eta_k^0 s_k^{(0)}(\psi_0) \left\{ \frac{s_k^{(2)}(\psi_0)}{s_k^{(0)}(\psi_0)} - \frac{s_k^{(1)}(\psi_0)^{\otimes 2}}{s_k^{(0)}(\psi_0)^2} \right\},$$

and

$$D(\psi_0) = \lim_{n \to \infty} n^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n} \left\{ -\tilde{Z}_{k,i} - \frac{s_k^{(1)}(\psi_0)}{s_k^{(0)}(\psi_0)} \right\}^{\otimes 2} \times p_{k,i}(\theta_0)\{1 - p_{k,i}(\theta_0)\}.$$

Assume $I(\psi_0) = \sum_{k=1}^{K} I_k(\psi_0)$ is positive definite.

(e) Lindeberg condition. For any $\varepsilon > 0$,

$$n^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n} Y_{k,i} |\tilde{Z}_{k,ij}|^2 I(n^{-1/2} |\tilde{Z}_{k,ij}| > \varepsilon) \times p_{k,i}(\theta_0)\{1 - p_{k,i}(\theta_0)\} \to_p 0$$

for $j \in \{1, \ldots, q\}$, where the subscript j denotes the jth element of the vector.

APPENDIX B. PROOF OF THEOREMS 3.1 AND 3.2

Our proof is constructed for a general K and thus covers both Theorems 3.1 and 3.2.

Proof. The log partial likelihood function in (8) can be rewritten as

$$\ell_K(\psi) = \sum_{k=1}^{K} \sum_{i=1}^{n} Y_{k,i} \delta_{k,i} \left[-\psi' \tilde{Z}_{k,i} - \log\{n S_k^{(0)}(\psi)\} \right],$$

where $\psi \in \Psi$. It is not hard to show that the function $n^{-1}\{\ell_K(\psi) - \ell_K(\psi_0)\}$ has the same probability limit as

$n^{-1}\{\Lambda_K(\psi) - \Lambda_K(\psi_0)\}$, where

$$\Lambda_K(\psi) = -\psi' \sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i} p_{k,i}(\theta_0)\tilde{Z}_{k,i} - \sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i} p_{k,i}(\theta_0) \log\{nS_k^{(0)}(\psi)\}.$$

The $n^{-1}\{\Lambda_K(\psi) - \Lambda_K(\psi_0)\}$ converges in probability to a continuous function

$$-(\psi - \psi_0)' \sum_{k=1}^{K} \eta_k^0 s_k^{(1)}(\psi_0) - \sum_{k=1}^{K} \log\left\{\frac{s_k^{(0)}(\psi)}{s_k^{(0)}(\psi_0)}\right\}\eta_k^0 s_k^{(0)}(\psi_0),$$

by conditions (a), (b), and (d), which has first-order derivative zero at $\psi = \psi_0$ and second-order derivative $-\sum_{k=1}^{K} I_k(\psi_0)$ as the negative of a positive definite matrix. The consistency of $\hat{\psi}_{\rm pl}$ thus is guaranteed by a classic convex theorem [1].

The asymptotic normality of $n^{1/2}(\hat{\psi}_{\rm pl} - \psi_0)$ is shown by first proving the asymptotic normality of $n^{-1/2}U_K(\psi_0)$, where $U_K(\psi_0) = \dot{\ell}_K(\psi) \mid_{\psi=\psi_0}$. Let

$$E_k(\psi) = S_k^{(1)}(\psi) / S_k^{(0)}(\psi).$$

By definition,

$$n^{-1/2}U_K(\psi_0) = n^{-1/2} \sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i}\{-\tilde{Z}_{k,i} - E_k(\psi_0)\} \times \{\delta_{k,i} - p_{k,i}(\theta_0)\},$$

which has a zero mean. When $n \to \infty$, for Lindeberg condition to hold, it suffices to show that

$$n^{-1} \sum_{k=1}^{K}\sum_{i=1}^{n} Y_{k,i}|E_{kj}(\psi_0)|^2 I\{n^{-1/2}|E_{kj}(\psi_0)| > \varepsilon\} p_{k,i}(\theta_0)\{1 - p_{k,i}(\theta_0)\} \to_p 0,$$

which can be easily checked under conditions (a)–(c). Thus, $n^{-1/2}U_K(\psi_0)$ converges in distribution to a random variable with mean 0 and limiting variance $D(\psi_0)$ as defined in condition (d). By Taylor's expansion of $U_K(\hat{\psi}_{\rm pl})$ around ψ_0,

$$n^{1/2}(\hat{\psi}_{\rm pl} - \psi_0) = -\hat{I}_K(\psi^\dagger)^{-1} n^{-1/2} U_K(\psi_0)$$

has the same limiting distribution as $-I(\psi_0)^{-1}n^{-1/2} \times U_K(\psi_0)$, where $\hat{I}_K$ is the observed information accumulated over all strata and $\psi^\dagger$ is on the line segment between $\hat{\psi}_{\rm pl}$ and ψ_0. The asymptotic normality of $n^{1/2}(\hat{\psi}_{\rm pl} - \psi_0)$ thus follows the asymptotic normality of $n^{-1/2}U_K(\psi_0)$ with mean 0 and variance $\Sigma = I(\psi_0)^{-1}D(\psi_0)I(\psi_0)^{-1}$. □

APPENDIX C. PROOF OF THEOREM 6.1

Proof. For the kth element in $\hat{\eta}^*$, it is straightforward to show that $\hat{\eta}_k(\hat{\psi}_{\rm pl})$ is a consistent estimator of η_k^0 for each k, since $\hat{\eta}_k(\hat{\psi}_{\rm pl}) - \hat{\eta}_k(\psi_0) \to_p 0$ and $\hat{\eta}_k(\psi_0) - \eta_k^0 \to_p 0$ in the decomposition $\hat{\eta}_k(\hat{\psi}_{\rm pl}) - \eta_k^0 = \hat{\eta}_k(\hat{\psi}_{\rm pl}) - \hat{\eta}_k(\psi_0) + \hat{\eta}_k(\psi_0) - \eta_k^0$. Convergence of the first component in probability to zero follows from a Taylor's expansion, condition (a), and the consistency of $\hat{\psi}_{\rm pl}$ for ψ_0. Convergence of the second component follows from the unbiasedness of $\hat{\eta}_k(\psi_0)$ as

$$E\left\{\sum_{i=1}^{n} Y_{k,i}\delta_{k,i} \Big/ \sum_{i=1}^{n} Y_{k,i}\exp(-\psi_0'\tilde{Z}_{k,i})\right\} = EE\left\{\sum_{i=1}^{n} Y_{k,i}\delta_{k,i} \Big/ \sum_{i=1}^{n} Y_{k,i}\exp(-\psi_0'\tilde{Z}_{k,i}) \,\Big|\, \mathcal{H}_{c_k}\right\} = \eta_k^0$$

and $\mathrm{var}\{\hat{\eta}_k(\psi_0)\} \to 0$ as $n \to \infty$.

The proof of asymptotic normality of $\hat{\eta}^*$ is somewhat more complicated. Let $b = \eta_k^0 s^{(1)}(\psi_0) / s^{(0)}(\psi_0)$. We can see that, for each $k = 1, \ldots, K$, $n^{1/2}\{\hat{\eta}_k(\hat{\psi}_{\rm pl}) - \eta_k^0\} - n^{1/2}b_k'(\hat{\psi}_{\rm pl} - \psi_0)$ has the same limiting distribution as $n^{1/2}\sum_{i=1}^{n} m_{k,i}(\theta_0) / s_k^{(0)}(\psi_0)$, where $m_{k,i}(\theta_0) = Y_{k,i}\{\delta_{k,i} - p_{k,i}(\theta_0)\}$ is a centralized discrete-time counting process in period k that is adapted to $\mathcal{H}_{c_k}$. The asymptotic normality of $n^{1/2}\sum_{i=1}^{n} m_{k,i}(\theta_0)$ holds, since $\delta_{k,i}$ is simply a binary random variable with conditional mean $p_{k,i}(\theta)$. However, derivation of the limiting variance is challenging, as the dependence between outcomes at different times may be difficult to handle. To resolve this, we utilize a theory developed in [11] for a single counting process with multiple jumps which satisfies the conditional independent increment assumption (5). That is, we naively treat $\delta_{u,i}$ and $\delta_{v,i}$ $(u \neq v)$ as independent outcomes even though marginally those two terms may be correlated when the subject is at risk during both follow-up periods. The (u, v)th entry of the limiting variance Ω thus is $\omega_{uv} = b_u'\Sigma b_v + \sigma_{uv}^2 / s_u^{(0)}(\psi_0)s_v^{(0)}(\psi_0)$, where Σ is the limiting variance of $n^{1/2}(\hat{\psi}_{\rm pl} - \psi_0)$ and σ_{uv}^2 is the limiting covariance of $n^{-1/2}\sum_{i=1}^{n} Y_{k,i}m_{k,i}(\theta_0)$ at $k = u, v$. That is, $\sigma_{uv}^2 = \lim_{n\to\infty} n^{-1}E\sum_{i=1}^{n} Y_{u,i}Y_{v,i}m_{u,i}(\theta_0)m_{v,i}(\theta_0)$, assuming that σ_{uv}^2 exists. To prove this, note that for $u, v \in \mathbb{K}$,

$$\Omega = \lim_{n\to\infty} \mathrm{cov}\big[n^{1/2}\{\hat{\eta}_u(\hat{\psi}_{\rm pl}) - \eta_u\}, n^{1/2}\{\hat{\eta}_v(\hat{\psi}_{\rm pl}) - \eta_v\}\big] = A + B + C + D,$$

where

$$A = \lim_{n\to\infty} nE\big[\{\hat{\eta}_u(\hat{\psi}_{\rm pl}) - \hat{\eta}_u(\psi_0)\}\{\hat{\eta}_v(\hat{\psi}_{\rm pl}) - \hat{\eta}_v(\psi_0)\}'\big] = b_u'\Sigma b_v,$$

$$B = \lim_{n\to\infty} nE\big[\{\hat{\eta}_u(\hat{\psi}_{\rm pl}) - \hat{\eta}_u(\psi_0)\}\{\hat{\eta}_v(\psi_0) - \eta_v^0\}'\big] = 0,$$

$$C = \lim_{n\to\infty} nE\big[\{\hat{\eta}_u(\psi_0) - \eta_u^0\}\{\hat{\eta}_v(\hat{\psi}_{\rm pl}) - \hat{\eta}_v(\psi_0)\}'\big] = 0,$$

and

$$D = \lim_{n\to\infty} nE\left[\{\hat{\eta}_u(\psi_0) - \eta_u^0\}\{\hat{\eta}_v(\psi_0) - \eta_v^0\}'\right]$$
$$= \sigma_{uv}^2 \big/ s_u^{(0)}(\psi_0) s_v^{(0)}(\psi_0).$$

Thus the result holds. □

ACKNOWLEDGMENTS

We thank Professor Ken Raffa for providing the red pine data used for illustration of our proposed methodology. Funding has been provided for this research from a USDA Cooperative State Research, Education and Extension Service (CSREES) Hatch project and an NSF project.

Received 31 March 2011

REFERENCES

[1] Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics*, **10**, 1100–1120. MR0673646

[2] Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics*, **30**, 89–99.

[3] Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B*, **34**, 187–220. MR0341758

[4] Cox, D. R. (1975). Partial likelihood. *Biometrika*, **62**, 269–276. MR0400509

[5] Diamond, I. D. and McDonald, J. W. (1991). Analysis of current-state data. In *Demographic Applications of Event History Analysis*, Eds. J. Trussell, R. Hankinson and J. Tilton, Oxford University Press, Oxford, 231–252.

[6] Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, **72**, 557–565. MR0451514

[7] Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, **42**, 845–854. MR0872963

[8] Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhauser, Basel. MR1180321

[9] Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. Wiley, New York. MR1924807

[10] Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd ed. Springer, New York.

[11] Lin, F. and Fine, J. P. (2009). Pseudomartingale estimating equations for modulated renewal process models. *Journal of the Royal Statistical Society Series B*, **71**, 3–23. MR2655521

[12] Lin, D. Y., Oakes, D. and Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika*, **85**, 289–298. MR1649115

[13] Martinussen, T. and Scheike, T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika*, **89**, 649–658. MR1929169

[14] Pipper, C. B. and Ritz, C. (2007). Checking the grouped data version of the Cox model for interval-grouped survival data. *Scandinavian Journal of Statistics*, **34**, 405–418. MR2346647

[15] Prentice, R. L. and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, **34**, 57–67.

[16] Prentice, R. L. and Kalbfleisch, J. D. (2003). Mixed discrete and continuous Cox regression model. *Lifetime Data Analysis*, **9**, 195–210. MR1970405

[17] Prentice, R. L., Williams, B. J. and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, **68**, 373–379. MR0626396

[18] Rasmussen, J. G., Møller, J., Aukema, B. H., Raffa, K. F. and Zhu, J. (2007). Bayesian inference for multivariate point processes observed at sparsely distributed times. *Journal of the Royal Statistical Society, Series B*, **69**, 701–713. MR2370076

[19] Ritz, C., Pipper, C., Yndgaard, F., Fredlund, K. and Steinrücken, G. (2010). Modelling flowering of plants using time-to-event methods. *European Journal of Agronomy*, **32**, 155–161.

[20] Shao, J. (2003). *Mathematical Statistics*, 2nd ed. Springer, New York. MR2002723

[21] Shiboski, S. C. and Jewell, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *Journal of the American Statistical Association*, **87**, 360–372.

[22] Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored and truncated data. *Journal of the Royal Statistical Society Series B*, **38**, 290–295. MR0652727

[23] Zhu, J., Rasmussen, J., Møller, J., Aukema, B. H. and Raffa, K. F. (2008). Spatial-temporal modeling of forest gaps generated by colonization from below- and above-ground bark beetle species. *Journal of the American Statistical Association*, **103**, 162–177. MR2420224

Feng-Chang Lin
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill
NC 27599
USA
E-mail address: flin@bios.unc.edu

Jun Zhu
Department of Statistics and Department of Entomology
University of Wisconsin at Madison
Madison
WI 53706
USA
E-mail address: jzhu@stat.wisc.edu