Harvard University
Harvard University Biostatistics Working Paper Series

Year 2006 *Paper* 40

Survival Analysis with Change Point Hazard Functions

Melody S. Goodman* Yi Li†
Ram C. Tiwari‡

*Harvard University, mgoodman@hsph.harvard.edu
†Harvard University and Dana Farber Cancer Institute, yili@jimmy.harvard.edu
‡National Cancer Institute, tiwarir@mail.nih.gov

Survival Analysis with Change Point Hazard Functions

Melody S. Goodman,[1,*] **Yi Li**[2,3] **and Ram C. Tiwari**[4]

[1]Department of Biostatistics, Harvard University, Boston, MA 02115, U.S.A.

[2]Department of Biostatistics, Harvard University, MA 02115, U.S.A.

[3]Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, MA 02115, U.S.A.

[4]Statistical Research and Applications Branch, National Cancer Institute, Bethesda, MD 20892, U.S.A.

Abstract

Hazard functions are an important component of survival analysis as they quantify the instantaneous risk of failure at a given time point. Increasing attention has been given to testing the assumption of a constant failure rate against a failure rate that changes at a single point in time. We expand the set of alternatives to allow for the consideration of multiple change-points, and propose a model selection approach using sequential testing that considers two types of models: the piecewise constant hazard model and piecewise linear hazard model. The latter model will easily accommodate the addition of covariates. These methods are data driven and allow us to estimate not only the trend in the hazard function but where those changes in trend occur. Such an analysis is valuable as it has implications in health care policy decisions. Methods, like the ones proposed in this paper, that estimate the overall survival trends for an entire population allow researchers and clinicians a better understanding of how changing medical practice affects the survival experience for a patient population. We illustrate our methods by applying them to the NIH SEER prostate and breast cancer data sets, and the NCHS birth cohort data.

---

[*] *Corresponding author's email address:* mgoodman@hsph.harvard.edu

Key words: Alpha spending, Change-point, Hazard estimation, Multiple comparisons Survival analysis

1. Introduction

Cancer is the second leading cause of death in the United States and a major burden to health care, therefore, medical progress against cancer is a major public health goal. The breakthroughs in clinical trials usually significantly improve the survival of cancer patients. To better understand the impact of medical breakthroughs, treatments or interventions, on the survival experience for the patient population, it is useful to provide a general picture of cancer survival trend on the population at large. The methodology developed in the paper estimates the trend of the hazard function and where the changes in trend occur and is applied to survival data in two types of cancer. We also examine the hazard for infant mortality one of the most salient indicators of a population's health.

The hazard function is an important component of survival analysis since it describes the immediate risk of failure at a given time point. Although common survival methods such as the Cox proportional hazards model do not require explicit estimation of the hazard function, as they are more concerned with the effects of the covariates on the hazard function, there are several situations where explicit estimation of the hazard function is useful. One such case is change-point hazard rate models. These models assume a function with different hazard rates that change at a few time points. These times points are often referred to as the change points, and are unknown and need to be estimated.

There is much work in the literature regarding estimation and testing in a piecewise constant model with one change point (see for example, Loader, 1991; Matthews and Farewell, 1982; Gijbels and Gürler, 2003; Nguyen, Rogers and Walker, 1984; Yao, 1986; and Pons, 2002). However, we are interested in testing for the presence of multiple change points in the hazard function as there are some public health examples that suggest that, due to improvement in treatments or diagnosis, there may be two or more changes in the hazard rate. This paper is motivated by two examples, the first is an interest in examining prostate cancer mortality rates among black and white men in the United States. The second is examining the infant mortality rate among black and white infants born in the United States in 1998.

Gray (1996) proposed methodology for nonparametric estimation of a hazard function using ordinary nonparametric regression smoothers. This approach provides a smooth estimate of the hazard function and allows for covariates. Although finding the appropriate value of the smoothing parameter is non-trivial in some cases, the hazard function is not required to be linear. In this approach the knots and the number of knots are to be prespecified by the data analyst. The use of ordinary smoothers provide great insight into the trend, including changes in the trend, of the hazard function. However, in many contexts there is a need to know when changes in the trend occur. For example, when trying to develop cost effective recommendations for disease screening there is an interest in knowing how the hazard changes and when these changes occur. We are interested in detecting the number of change points in the hazard function, and estimating all of the unknown parameters in the model, including the change points, using the data to determine the number and value of the change points.

There has been some debate about the use of likelihood ratio statistics to test the change point model against a constant hazard model (Matthews and Farewell, 1982; Nguyen, Rogers and Walker, 1984; Henderson, 1990; and Matthews and Farewell, 1985). Although other methods have been proposed (see for example, Nguyen, Rogers and Walker, 1984; Müller and Wang, 1990; and Henderson, 1990), the maximum likelihood ratio test has been shown to work under certain conditions (Matthews and Farewell, 1985; Yao, 1986).

Let us assume that the maximum number of change points in the model is some finite number K. Our aim is to find the model, with the number of change points k ($k = 0, ..., K$) that best fits our data. We can think of this as a model selection process, where we start with the model with no change points and perform a hypothesis test to compare it to the model with one change point (Loader, 1991; Gijbels and Gürler, 2003; and Matthews and Farewell, 1982). We can then compare the model with one change point, to the model with two change points, and so forth, only moving on to the next step if we reject the null hypothesis in the previous step. We formulate the likelihood ratio test statistic to compare multiple change point models by extending the methodology from the single

change point model (Loader, 1991; Gijbels and Gürler, 2003; and Matthews and Farewell, 1982). We also formulate a Wald type test statistic to test the same hypothesis and discuss the advantages and disadvantages of both types of statistics later.

The rest of this article is structured as follows. We define the piecewise constant model with multiple change points and develop both likelihood ratio test (see the Appendix A.1) and Wald type test statistics in Section 2 . In Section 3 we define the piecewise linear model with multiple change points and develop likelihood ratio test (see the Appendix A.2 for derivation) and Wald type test statistics for this model. We discuss the issue of multiple comparisons and develop an alpha spending function to preserve the overall Type I error in Section 4. For methodology assessment simulation studies to investigate the estimation of parameters, Type I error, and power, are presented in Section 5. Illustration of our proposed methods is conducted through application in three data examples and presented in Section 6. We conclude with a general discussion in Section 7.

2. Piecewise Constant Multiple Change Point Model

Let $X_1, ..., X_n$ denote independent identically distributed survival times, and $C_1, ..., C_n$ be the censoring times which are assumed to be independent of X. We only observe the pairs $(T_i, \delta_i), i = 1, 2, ..., n$, where $T_i = \min(X_i, C_i)$ and $\delta_i = 1$ if $X_i \leq C_i$ and zero otherwise. Consider the following change point model:

$$\lambda(t) = \begin{cases} \alpha_1 & 0 \leq t < \tau_1 \\ \alpha_2 & \tau_1 \leq t < \tau_2 \\ \vdots & \\ \alpha_{K+1} & t \geq \tau_K, \end{cases} \tag{1}$$

where $0 < \tau_1 < \ldots < \tau_K$, K is the number of change points in the model, and α_j is the value of the hazard function between the time points τ_{j-1} and τ_j. The $\tau_j's$ can be thought of as the order statistics for the change points in the hazard function.

Since $f(t) = \lambda(t)\exp[-\int_o^t \lambda(u)du]$, for model (1)

$$f(t) = \begin{cases} \alpha_1 \exp[-\alpha_1 t] & 0 \leq t < \tau_1 \\ \alpha_2 \exp[-\alpha_1\tau_1 - \alpha_2(t-\tau_1)] & \tau_1 \leq t < \tau_2 \\ \vdots & \\ \alpha_{K+1}\exp[-\alpha_1\tau_1 - \alpha_2(\tau_2-\tau_1) - \ldots - \alpha_{K+1}(t-\tau_K)] & t \geq \tau_K, \end{cases}$$

is a piecewise exponential density function. Let $X(t)$ denote the number of deaths observed up to time t. $X(t)$ is defined by

$$X(t) = \sum_{i=1}^{n} I(T_i < t)\delta_i,$$

where δ_i is an indicator for non-censoring. Note that $X(\tau_j) = \sum_{i=1}^{n} I(T_i < \tau_j)\delta_i$ is the number of observed deaths up to change-point τ_j (Gijbels and Gürler, 2003).

We frame our problem in a similar manner as a sequential analysis problem, where we perform a hypothesis test and if we reject the null hypothesis we will continue on to the next hypothesis test. If we fail to reject the null hypothesis, we stop and conclude that we have found the final model. We test $H_0 : \alpha_{k-1} = \alpha_k$ versus $H_1 : \alpha_{k-1} \neq \alpha_k$ which is equivalent to testing the null hypothesis that $\tau_{k-1} = 0$ for $k = 2, ..., K$. Extending the existing methodology (Matthews and Farewell, 1982; Loader, 1991; Gijbels and Gürler, 2003), we develop a likelihood ratio test statistic to perform this test (see §A.1).

Since the reduced model has two less parameters than the full model, a naive application of asymptotic likelihood ratio theory would lead one to conclude that the likelihood ratio test statistic should have a χ^2_2 distribution (Matthews and Farewell, 1982). Matthews and Farewell (1982) have shown that, although asymptotic likelihood ratio theory does not properly apply, based on simulation, the percentiles of the χ^2_2 distribution appear to agree quite well even for censored data. Worsley (1988) reported percentage points for the same statistic based on the exact null distribution that were much larger than those reported by Matthews and Farewell (1982), and noted that the percentage points do not appear to tend toward a finite limit as the sample size approaches infinity (Worsley, 1988).

2.1 *Wald Test for the Piecewise Constant Model*

We are interested in testing for change points in model (1). An equivalent hypothesis would be to test if the hazard before and after such points are equivalent. To test the null hypothesis of no change points against the alternative of one change point, we can test the null hypothesis that $\alpha_1 = \alpha_2$ or $\alpha_1 - \alpha_2 = 0$. We propose the use of a Wald test statistic for testing a linear combination. Let $\theta' = [\alpha_1, \alpha_2, ...\alpha_{K+1}, \tau_1, \tau_2, ...\tau_K]$, to test a hypothesis of the form $H_0 : C'\theta = M$, we can use the following Wald test statistic

$$X_W = (C'\hat{\theta} - M)'[C'\hat{\Sigma}_{\hat{\theta}}C]^{-1}(C'\hat{\theta} - M) \sim \chi^2_s,$$

where C' is an $s \times p$ matrix, $s \leq p$, and M is the $s \times 1$ solution vector (Lachin, 2000). Although our proposed method is a multiple testing procedure, it is a stepwise process. Therefore, we only test one hypothesis at a time. To test $H_0 : \alpha_{k-1} - \alpha_k = 0$ versus $H_1 : \alpha_{k-1} - \alpha_k \neq 0$ we use a Wald type test statistic of the form,

$$X_W = \frac{(\hat{\alpha}_{k-1} - \hat{\alpha}_k)^2}{Var(\hat{\alpha}_{k-1} - \hat{\alpha}_k)} \sim \chi^2_1. \tag{2}$$

In order to calculate the variance in the denominator of the test statistic we use a partitioned Hessian matrix containing only those parameters of θ in the test statistic (2). Yao (1986) proves that the elements of θ are independent and therefore there is zero covariance between the elements of θ.

2.2 *Estimation in the Piecewise Constant Model*

We estimate the parameters in the model using an optimization function in R, based on the Nelder-Mead Simplex algorithm, to minimize the negative log-likelihood function evaluated at the maximum likelihood estimates of the α_j and finding those values of the τ_j that minimize the function. The maximum likelihood estimates of the τ_j are those values returned by the optimization function that minimize the negative log-likelihood.

3. Piecewise Linear Multiple Change Point Model

The piecewise linear model is slightly more comprehensive than the piecewise constant model and is one that we believe will be encountered more often in practice. This model is log linear and has a continuous hazard function with changes in trends occurring at the change points. It also allows for the addition of covariates into the model.

Suppose we observe data (T_i, δ_i), let λ be the hazard function of T_i, the time to some event. Let $\eta = \ln \lambda$, where η is a piecewise linear spline function with knots at $\tau_1, ..., \tau_K$ defined by

$$\eta(t) \equiv \eta(t; \alpha_0, \alpha_1, ..., \alpha_{K+1}, \tau_1, .., \tau_K, \boldsymbol{\beta}) = \alpha_0 + \alpha_1 t + \sum_{k=1}^{K} \alpha_{k+1}(t - \tau_k)_+ + \mathbf{Z}'\boldsymbol{\beta},$$

for fixed K, where $x_+ \equiv \max(0, x)$, $\mathbf{Z}$ is the covariate vector, and $\boldsymbol{\beta}$ is a vector of the parameter estimates for the effects of the covariates (Cai, Hyndman and Wand, 2002). This is also known as a linear jointpoint regression model with jointpoints $\tau_1, ..., \tau_K$ (see, for example, Kim et al. (2000)).

The piecewise linear change point model is defined by η, where

$$\eta(t) = \begin{cases} \alpha_0 + \alpha_1 t + \mathbf{Z}'\boldsymbol{\beta} & 0 \le t < \tau_1 \\ \alpha_0 + \alpha_1 t + \alpha_2(t - \tau_1)_+ + \mathbf{Z}'\boldsymbol{\beta} & \tau_1 \le t < \tau_2 \\ \vdots & \\ \alpha_0 + \alpha_1 t + \alpha_2(t - \tau_1)_+ + \ldots + \alpha_{K+1}(t - \tau_K)_+ + \mathbf{Z}'\boldsymbol{\beta} & t \ge \tau_K. \end{cases}$$

Note that we can rewrite $\eta(t)$ as

$$\eta(t) = \alpha_{0i} + \alpha_{1i} t + \mathbf{Z}'\boldsymbol{\beta} \qquad \tau_i \le t < \tau_{i+1},$$

where $\alpha_{0i} = \alpha_0 - \sum_{j=1}^{i+1} \alpha_j \tau_{j-1}$ and $\alpha_{1i} = \sum_{j=1}^{i+1} \alpha_j, i = 0, 1, ..., K$, and $\tau_0 \equiv 0$.

The log-likelihood for the data is

$$logL \equiv logL(\alpha_0, \alpha_1, ..., \alpha_{K+1}, \tau_1, .., \tau_K, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \delta_i \eta(T_i) - \int_0^{T_i} e^{\eta(u)} du \right\}, \qquad (3)$$

where δ_i is an indicator of non-censoring, $T_i = \min(X_i, C_i)$ and C_i is the censoring time (Cai, Hyndman and Wand, 2002).

We can test for the existence of change point τ_{k-1} by testing the null hypothesis $\tau_{k-1} = 0, k = 2, ..., K$. The likelihood ratio test statistic to test this hypothesis is equal to $logL(\alpha_0, \alpha_1, ..., \alpha_k, \tau_1, ..\tau_{k-1}, \boldsymbol{\beta}) - logL(\alpha_0, \alpha_1, ..., \alpha_{k-1}, \tau_1, ..\tau_{k-2}, \boldsymbol{\beta})$ evaluated at the maximum likelihood estimates of the parameters, where the $logL$ is defined by equation (3). Calculation of the likelihood ratio test statistic can be found in A.2.

3.1 *Wald Test for the Piecewise Linear Model*

Let $\theta' = [\alpha_0, \alpha_1, ...\alpha_{K+1}, \tau_1, \tau_2, ...\tau_k, \boldsymbol{\beta}]$, a test of the hypothesis for the jth element of θ is,

$$X_W = \frac{(\hat{\theta}_j - \theta_0)^2}{\widehat{Var}(\hat{\theta}_j)} \sim \chi^2_1,$$

where $\hat{V}(\hat{\theta}_j) = [I(\hat{\theta})^{-1}]_{jj}$ (Lachin, 2000). In the piecewise linear model a Wald test for the null hypothesis of no change point versus the alternative of one change point is equivalent to testing the hypothesis that $\alpha_2 = 0$, as the model with one change point reduces to the model with no change points when α_2 is equal to zero. A test for the existence of change point τ_k can be conducted by testing, $H_0 : \alpha_{k+1} = 0$ versus $H_1 : \alpha_{k+1} \neq 0$ using a Wald type test statistic of the form,

$$X_W = \frac{\alpha^2_{k+1}}{\widehat{Var}(\hat{\alpha}_{k+1})} \sim \chi^2_1. \tag{4}$$

3.2 *Estimation in the Piecewise Linear Model*

Estimation in the piecewise linear model was conducted in a similar manner to that of the piecewise constant model. Using a multidimensional optimization function in R based on the Nelder-Mead Simplex algorithm, we minimize the negative log-likelihood function and find those values of α_j, τ_j, and β_j that optimize the likelihood function. The maximum likelihood estimates of the parameters are those optimal values which minimize the negative log-likelihood function.

4. Preserving the Type I Error

Since our procedure has multiple comparisons, we have to make some correction to preserve the overall Type I error. A common approach to this problem would be to use a Bonferroni correction but this is dependent on prior knowledge of the number of hypothesis test being conducted and is often too conservative. To preserve the level of the test at α, we borrow methodology from the group sequential analysis literature. Lan and DeMets (1983) proposed an alpha spending technique in which the nominal significance level needed to reject the null hypothesis at each analysis is $\leq \alpha$ and increases as the study progresses. Thus, for the test, it is more difficult to reject the null at the earliest analysis but easier later on.

They proposed to use

$$\alpha^*(k) = \alpha s^*(k),$$

where $\alpha^*(k)$ is the significance level for the kth hypothesis and $s^*(k) = \frac{k}{K}$ is the spending function. Here, $\alpha^*(1) < \alpha^*(2) < ... < \alpha^*(K)$. In order to find a parsimonious model we want strong evidence for choosing a more complicated model, one with more change points, over a simpler one. Therefore, we are interested in a decreasing alpha spending function, where $\alpha^*(1) > \alpha^*(2) > ... > \alpha^*(K)$. With a decreasing alpha spending function, the test for each additional change point will be conducted at a more stringent α level than the one before it. If the overall significance level is α, let $\alpha^*(k) = \frac{\alpha}{2^{k-1}}$, where $\alpha^*(k)$ is the significance level for the kth hypothesis test. An advantage of this alpha spending function is that it is not dependent on the overall number of hypothesis test being conducted. Therefore, one does not need to know in advance how many hypothesis test they will need to conduct before reaching a final model.

The Type I error can be thought of as incorrectly choosing a model that has more change points than the true model. We show that the type I error rate will not exceed α by calculating the probability of choosing a model with one or more change points given the true model has no change points, the probability of choosing a model with k change points

given the true model has no change points, and the probability of choosing a model with more than k change points given the true model has k change points are all at most α.

Let M_i be the event that the model has i change points. Then

$$P(M_{i\geq 1}|M_0) = 1 - P(M_0|M_0) = 1 - (1-\alpha) = \alpha, \tag{5}$$

$$P(M_k|M_0) = \left(1 - \frac{\alpha}{2^k}\right) \prod_{i=1}^{k} \frac{\alpha}{2^{i-1}} < \frac{\alpha^k}{\prod_{i=1}^{k} 2^{i-1}} < \alpha,$$

and for $k \geq 1$

$$P(M_{i>k}|M_k) = \sum_{j=1}^{\infty} \left[\left(1 - \frac{\alpha}{2^{k+j}}\right) \prod_{i=k+1}^{k+j} \frac{\alpha}{2^{i-1}} \right] < \alpha.$$

5. Simulation

5.1 *Simulation for the Piecewise Constant Multiple Change Point Model*

We conducted a simulation study to investigate the estimation of parameters, the overall Type I error rate, and the power of our proposed methodology. These studies were conducted as follows. We simulated survival times from a piecewise constant hazard model with two change points by inverting the CDF and using the probability integral transformation. We used the uniform distribution to generate censoring times. In each case 5,000 independent data sets were generated, for each data set we use the Wald type test statistic (2) to compare the null model with no change points to an alternative model with one change point. If we reject the null we continue to the next hypothesis and test the null model of one change point to the alternative model of two change points. The α level for each test was determined using the alpha spending function proposed in Section 4. For each data set a final model was chosen and the parameters were estimated by maximizing the likelihood function for the final model. Table 1 displays the results of our simulations for $n = 500$, various values of the parameters, and different percentages of censoring.

The mean estimated value is the average estimated parameter value from all 5,000 simu-

Table 1

Model Estimation for Piecewise Constant Model with Two Change Points
Based on 5,000 simulations (n=500)

Censor	Parameters	Parameter Value	Mean Estimated Value	Std. Error Estimated Value	Model Based Std. Dev.	Coverage Probability
0%	α_1	0.95	0.953	0.053	0.046	0.949
	α_2	0.55	0.545	0.077	0.076	0.941
	α_3	0.15	0.145	0.032	0.034	0.923
	τ_1	2.0	1.992	0.118		
	τ_2	4.0	3.968	0.182		
5%	α_1	0.15	0.149	0.014	0.013	0.940
	α_2	0.55	0.558	0.039	0.035	0.947
	α_3	0.35	0.342	0.032	0.036	0.937
	τ_1	2.0	2.009	0.176		
	τ_2	4.0	4.018	0.170		
20%	α_1	0.15	0.148	0.021	0.017	0.933
	α_2	0.35	0.352	0.030	0.023	0.934
	α_3	0.55	0.579	0.059	0.060	0.951
	τ_1	1.0	1.023	0.218		
	τ_2	3.5	3.556	0.243		
25%	α_1	0.65	0.655	0.040	0.038	0.946
	α_2	0.35	0.346	0.048	0.048	0.937
	α_3	0.15	0.133	0.035	0.045	0.896
	τ_1	1.5	1.502	0.140		
	τ_2	3.5	3.486	0.208		
30%	α_1	0.15	0.148	0.019	0.019	0.935
	α_2	0.35	0.352	0.033	0.028	0.942
	α_3	0.55	0.569	0.048	0.045	0.946
	τ_1	1.0	1.012	0.092		
	τ_2	2.5	2.535	0.144		
35%	α_1	0.15	0.158	0.015	0.013	0.945
	α_2	0.55	0.550	0.051	0.046	0.943
	α_3	0.95	1.065	0.158	0.128	0.942
	τ_1	2.0	2.009	0.059		
	τ_2	4.0	4.055	0.169		
50%	α_1	0.15	0.126	0.014	0.017	0.558
	α_2	0.95	0.821	0.087	0.074	0.512
	α_3	0.45	0.387	0.071	0.084	0.874
	τ_1	2.0	2.005	0.017		
	τ_2	4.0	3.761	0.392		

Table 2
Type I Error Analysis for the Piecewise Constant Model
Based on 5,000 simulations (n=500)

% Censoring	Type I Error
0	0.0482
5	0.0478
10	0.0462
15	0.0496
20	0.0480
25	0.0500
30	0.0484
40	0.0494
45	0.0526
50	0.0526
60	0.0476
80	0.0530

lation runs, and the standard error is the standard deviation of these estimates. Based on simulation studies our method estimates the change points and the value of the hazard quite well even with a moderate amount of censoring (see Table 1). Although we used a sample size of 500 to ensure that there would be enough observed events, censoring percentages greater than 40% make parameter estimation less accurate. The coverage probability represents the relative frequency of the simulation runs whose 95% confidence interval includes the true parameter value. In most cases the coverage probability is close to the nominal level of 95%. However, as the percentage of censoring increases above 40% the coverage probability decreases significantly (see Table 1).

To test the overall Type I error rate of our methodology, we simulated data with no change points and implemented our stepwise approach. The Type I error is the number of times our methodology chooses a final model with one or more change points instead of the null model of no change points. Table 2 displays a sample of our results for samples of size 500, simulation results for samples of size 100 were quite similar. In all cases we observe a Type I error rate similar to the nominal significance level of the test (see Equation 5).

Table 3
Power Analysis for the Piecewise Constant Model with Two Change Points
Based on 5,000 simulations (n=500)

$\alpha_2 - \alpha_1$	$\alpha_3 - \alpha_2$	$\tau_2 - \tau_1$	% censoring	power
0.4	0.4	2	1	0.98
0.4	-0.2	2	3	0.98
-0.4	-0.4	2	16	0.95
0.2	0.2	1.5	23	0.93
-0.6	0.4	2	26	0.92
-0.3	-0.2	1.5	27	0.93
0.2	0.2	1.5	27	0.96
0.4	0.4	2	36	0.81
0.8	-0.5	2	49	0.90
0.8	-0.5	2	57	0.69

To conduct power analysis we simulated data under the alternative of two change points and applied our methodology to find the final model. The power is essentially a measure of the accuracy of our method in choosing the final model. Table 3 displays the results of the power analysis for samples of size 500. Power was most affected by sample size, the difference between α_{k-1} and α_k, and the difference between τ_{k-1} and τ_k. With samples of size 500, $|\alpha_{k-1} - \alpha_k| > 0.2$, $|\tau_{k-1} - \tau_k| > 1.5$, and $< 30\%$ censoring, we observed power of at least 90% , however as the censoring percentage increase above 30% the power decreases (see Table 3). We ran similar simulations for smaller sample sizes and found that estimation was not accurate and sometimes not possible for sample sizes less than 100 if there was censoring because there were not enough events in each interval (i.e., $t < \tau_1, \tau_1 < t < \tau_2, t > \tau_2$). For samples of size 200 we found the estimation of parameters and the coverage probabilities to be quite good, however as one would expect there was a significant loss in power compared to samples of size 500.

5.2 *Simulation for the Piecewise Linear Multiple Change Point Model*

Simulation in the piecewise linear model was conducted in a similar manner to that of the piecewise constant model. We simulated 5,000 data sets from a piecewise linear model

Table 4

Model Estimation for the Piecewise Linear Model with Three change Points Based on 5,000 simulations (n=1000)

Censor	Parameters	Simulated Value	Mean Estimated Value	Std. Error Estimated Value	Model Based Std. Dev.	Coverage Probability
0%	α_0	0.10	0.112	0.028	0.024	0.891
	α_1	0.20	0.215	0.028	0.018	0.828
	α_2	0.30	0.313	0.027	0.023	0.901
	α_3	0.45	0.458	0.027	0.047	0.990
	α_4	0.70	0.709	0.028	0.038	0.999
	τ_1	0.50	0.503	0.014		
	τ_2	1.50	1.520	0.056		
	τ_3	2.50	2.552	0.107		
	β_1	0.50	0.512	0.028	0.034	0.930
	β_2	0.75	0.761	0.028	0.019	0.851
20%	α_0	0.10	0.109	0.032	0.026	0.921
	α_1	0.20	0.211	0.032	0.026	0.878
	α_2	0.30	0.313	0.032	0.038	0.935
	α_3	0.45	0.464	0.030	0.045	0.997
	α_4	0.70	0.713	0.025	0.060	0.999
	τ_1	0.50	0.506	0.016		
	τ_2	1.50	1.533	0.066		
	τ_3	2.50	2.487	0.106		
	β_1	0.50	0.511	0.031	0.034	0.950
	β_2	0.75	0.765	0.031	0.021	0.861
50%	α_0	0.10	0.113	0.028	0.026	0.913
	α_1	0.20	0.216	0.029	0.024	0.869
	α_2	0.30	0.313	0.027	0.035	0.944
	α_3	0.45	0.460	0.028	0.031	0.999
	α_4	0.70	0.710	0.028	0.027	0.999
	τ_1	0.50	0.501	0.014		
	τ_2	1.50	1.512	0.054		
	τ_3	2.50	2.523	0.091		
	β_1	0.50	0.512	0.027	0.034	0.950
	β_2	0.75	0.753	0.029	0.021	0.858
75%	α_0	0.10	0.113	0.028	0.027	0.908
	α_1	0.20	0.217	0.028	0.024	0.847
	α_2	0.30	0.315	0.028	0.034	0.922
	α_3	0.45	0.460	0.029	0.089	0.999
	α_4	0.70	0.710	0.027	0.081	0.999
	τ_1	0.50	0.500	0.015		
	τ_2	1.50	1.509	0.054		
	τ_3	2.50	2.522	0.096		
	β_1	0.50	0.512	0.028	0.033	0.941
	β_2	0.75	0.762	0.028	0.019	0.832

Table 5
Type I Error Analysis for the Piecewise Linear Model
Based on 5,000 simulations (n=1000)

% Censoring	Type I Error
5	0.0486
10	0.0524
20	0.0454
55	0.0526
75	0.0464

with three change points and two covariates, one dichotomous and one continuous. For each data set we used a Wald type test statistic (Equation 4) to determine the final model. The parameter values for the final model were estimated via maximum likelihood. A sample of the results for $n = 1,000$ and various censoring percentages are displayed in Table 4. Based on our simulation studies estimation of model parameters is quite good even with a high censoring percentage. Similar results were found for $n = 500$, however estimation was not always feasible with high censoring percentages.

To test the overall Type I error rate in the piecewise linear model we simulated data with no change points and two covariates, one continuous and one dichotomous, and implemented our stepwise approach. The Type I error is the number of times a final model with one or more change points is chosen. Table 5 displays results for samples of size 1,000. Similar results were found for samples of size 500 with less than 30% censoring. Estimation was not always feasible for $n = 500$ with higher censoring percentages. In all cases the Type I error rate is close to the nominal significance level.

To conduct power analysis we simulated data under the three change point alternative and applied our method to choose a final model. Table 6 displays the results of our power analysis for samples of size 1,000 and different percentages of censoring. Even with high levels of censoring we have approximately 90% power. There was a significant loss in power for samples of size 500.

Table 6

Power Analysis for the Piecewise Linear Model with Three Change Points
Based on 5,000 simulations (n=1000)

α_2	α_3	α_4	$\tau_2 - \tau_1$	$\tau_3 - \tau_2$	% censoring	power
0.30	0.45	0.70	1.0	1.0	0	0.99
0.40	0.60	0.70	1.0	1.0	9	0.99
0.30	0.40	0.60	1.0	1.0	25	0.99
0.30	0.45	0.70	1.0	1.0	53	0.89
0.30	0.45	0.70	1.0	1.0	75	0.92

6. Applications

We apply our proposed methodology to three data examples; prostate cancer mortality, breast cancer mortality, and infant mortality. We restrict the change points to be larger than the first survival time and smaller than the second to last survival time, assuming these are non-censored time points, to avoid singularity, $T_{(1)} < \tau_1 < \ldots < \tau_k < T_{(n-1)}$ (Yao, 1986; Müller and Wang, 1994).

6.1 *Prostate Cancer Data*

To examine prostate cancer we use the Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Public-Use Data (1973-2001), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2004, based on the November 2003 submission. This data contains cancer incidence and survival for cases diagnosed from 1973 to 2001, follow-up continued until December 31, 2001. For the purpose of analysis we excluded subjects with unknown follow-up time and censored all data at 25 years of follow-up.

We are interested in finding the number of change points and estimating the location of the change points. We define an event as death from prostate cancer. If a subject dies from another cause they are censored at the time of their death. We restrict our analysis

to men who were diagnosed with localized or regional stage prostate cancer between 1973 and 1980. By restricting the year of diagnosis we limit our analysis to the SEER-9 registries comprising approximately 10% of the U.S. population. Preliminary analysis shows different trends in the hazard for each race, therefore race is not an additive variable in this model, and thus we stratified our analysis by race.

There were 21,099 white men who fit our restriction criteria in the SEER data set with 8,419 events and 71.5% of the observations being censored. The estimated log hazard function for white men has two change points (namely, at 2.42 and 18.53) and is defined by Equation 6.

$$\eta(t) = -3.66 + 0.26t - 0.30(t - 2.40)_+ - 0.05(t - 18.53)_+, \tag{6}$$

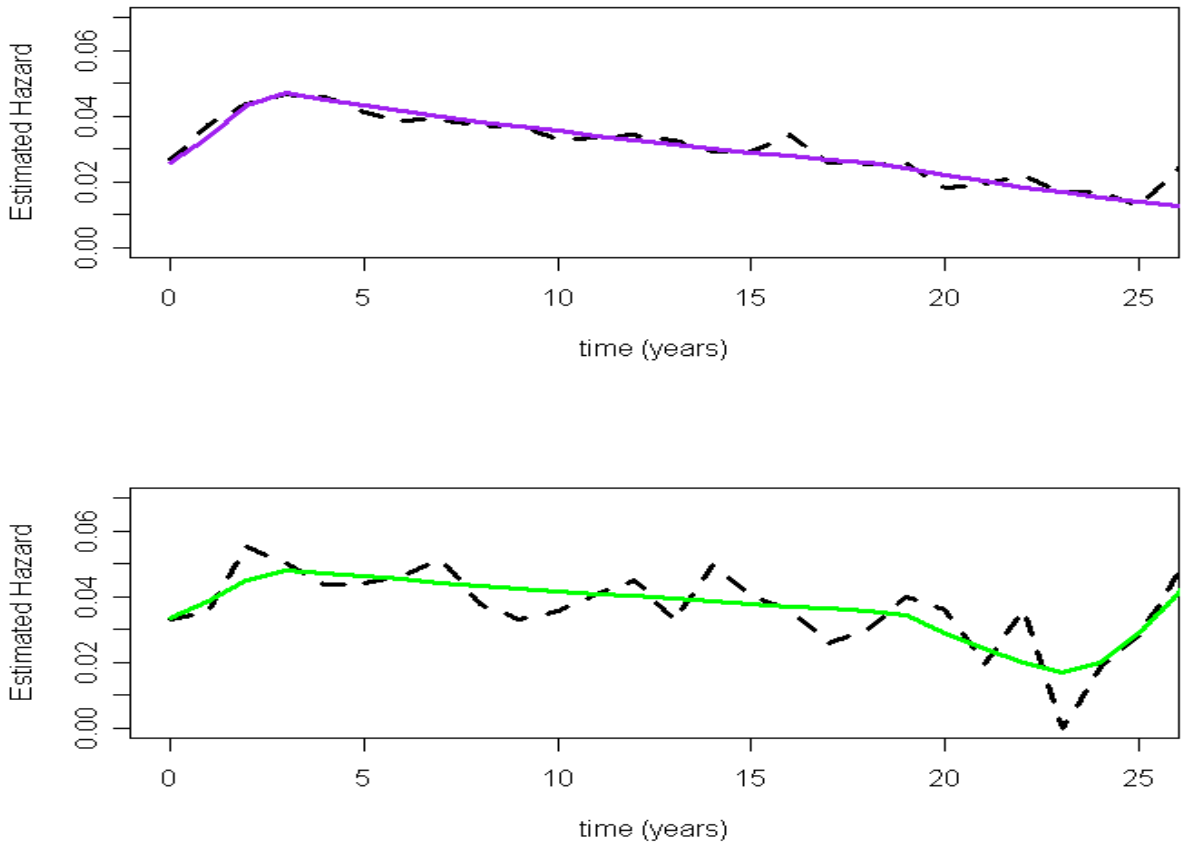
where $t \geq 0$.

The top graph in Figure 1 displays the estimated hazard function for white men. The dashed line is an estimate of the hazard function using the life table method of PROC LIFETEST in SAS. The solid line is the estimated hazard function using our method. From the graph we observe that our method is basically a smoothed version of the SAS estimate. The hazard function for white men diagnosed with prostate cancer between 1973 and 1980 increases until the first change point, and then begins to decrease until the second change point, followed by a slightly sharper decline until the end of the follow-up period.

There were 2,058 black men who fit our restriction criteria in the SEER data set with 955 events and 68.3% of the observations being censored. We estimate the log hazard for black men to have three change points (2.49, 18.94,and 23.34) and be defined by Equation 7. The sample size for blacks is considerably less than that of whites, the estimated hazard function from SAS (the dashed line in the bottom graph of Figure 1) is quite jagged and even drops down to zero for one time point where there were no events. The estimated hazard function from our method is the solid line in the bottom graph of Figure 1.

$$\eta(t) = -3.40 + 0.15t - 0.17(t - 2.49)_+ - 0.16(t - 18.94)_+ + 0.54(t - 23.34)_+, \tag{7}$$

Figure 1. Estimated Prostate Cancer Mortality Hazard Function for Men Diagnosed 1973-1980 by Race



The solid line is the estimated hazard function using our proposed methodology. The dashed line is the estimated hazard function using the life table method in SAS PROC LIFETEST. The top graph is the estimated hazard function for white men, and the bottom graph is the estimated hazard function for black men.

where $t \geq 0$.

The hazard function for black men has a sharp increase until the first change point followed by a gradual increase until the second change point after which there is a sharp decline until the third change point, followed by an increase until the end of the follow-up period.

These estimated hazard functions suggest that the prostate cancer process for black and whites may be different. For whites, the hazard reaches it first peak around two and a half years after diagnoses and then gradually declines over the next twenty years. The sample size for blacks is approximately 10% that of whites. Although we believe the sample size for blacks was sufficient to conduct analysis a larger sample size would have provided better insight into the prostate cancer process for this group, allowing us to determine with better certainty if the process for this group is different from that of their white counterparts.

6.2 *Breast Cancer Data*

Using the SEER data set we examine the hazard for breast cancer mortality. We define an event as death from breast cancer. We limited our analysis to women that were diagnosed with localized stage breast cancer between 1973 and 1980, were less than 50 years old at time of diagnosis, and had known follow-up time. We stratified our analysis by race. There were 7,224 white women in the SEER data set that fit our restriction criteria of these there were 1,812 events and the remaining 74.9% of the observations are censored. There were 683 black women who fit our restriction criteria in the SEER data set, 199 events and 70.9% of the observations are censored.

The estimated log hazard function for white women has one change point at 1.29 years after diagnoses and is defined by Equation 8. The hazard of breast cancer mortality for white women increases for the first three years after diagnosis and then begins to decline gradually through the end of the follow-up period, demonstrating that the first three years are the most critical after which the longer the time since diagnosis the lower the risk of

mortality. The top graph in Figure 2 displays the estimated hazard function for white women.

$$\eta(t) = -5.83 + 1.72t - 1.82(t - 1.29)_+, \quad t \geq 0 \tag{8}$$

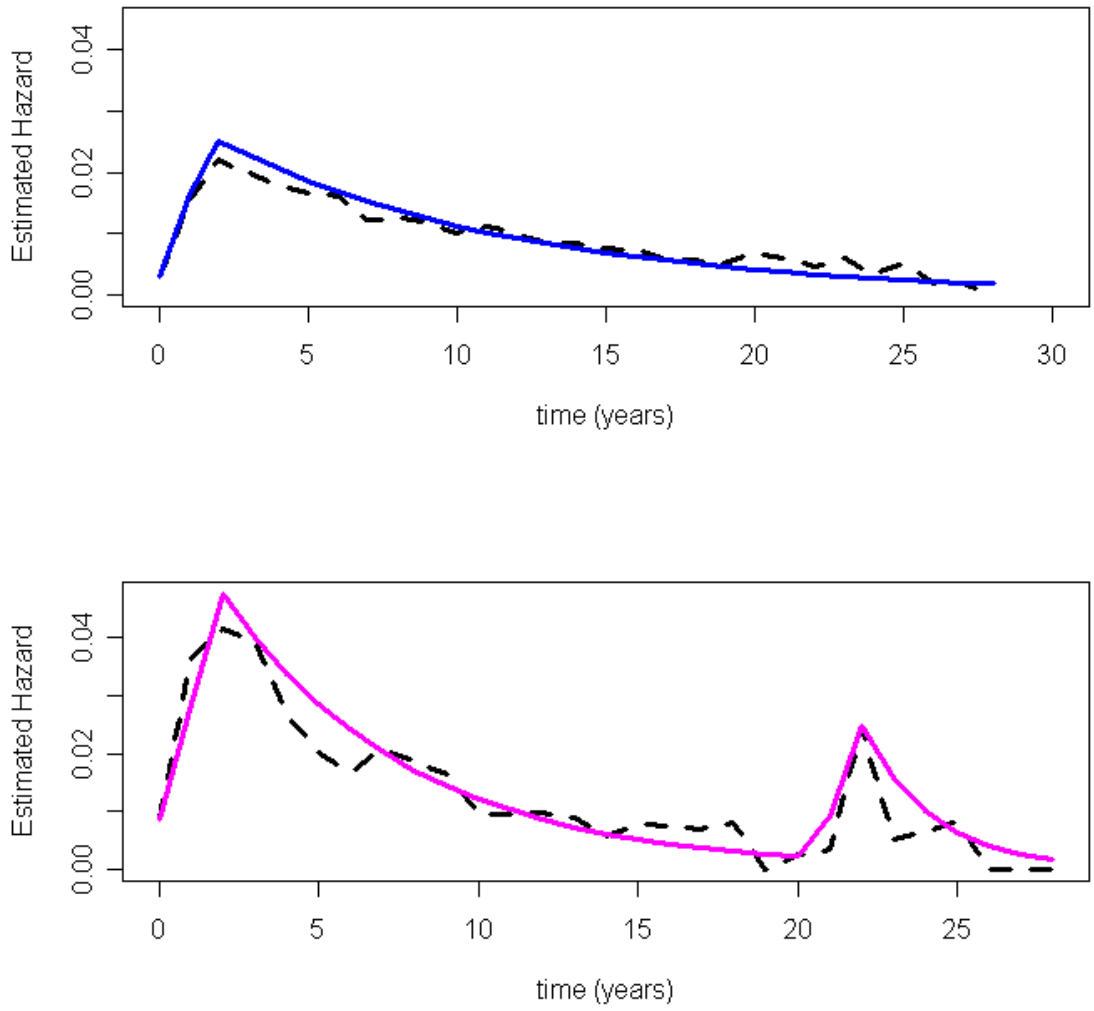
$$\eta(t) = -4.76 + 1.2t - 1.37(t - 1.5)_+ + 1.49(t - 19.94)_+ - 1.77(t - 21.82)_+, \quad t \geq 0 \tag{9}$$

The estimated log hazard function for black women has three change points (namely, at 1.5, 19.94, and 21.82) and is defined by Equation 9. The hazard function for black women increases up to the first change point, followed by a decline until the second change point. It begins to increase again until the third change point after which it decreases until the end of the follow-up period. The trend in the hazard for blacks is similar to that of whites except there is an unexplained spike at 22 years after diagnosis. This spike can be attributed to six breast cancer deaths 22 years after diagnosis, while there was only one event in every other year from 20 to 25 years after diagnosis. There was also a simultaneous decrease in the risk set as those women diagnosed in 1980 reach the end of follow-up for the study. The bottom graph in Figure 2 displays the estimated hazard function for black women.

6.3 *Infant Mortality Data*

To examine infant mortality in the Unites States we use the 1998 Birth Cohort linked birth/infant death data set from the National Center for Health Statistics (Hyattsville, Maryland, 1998). This data set contains information from the birth certificate for all births in the Unites States in 1998. These children are followed for one year from their date of birth, if they die within that year, the information from their death certificate is linked to the birth certificate information. We will only examine singleton births as the mortality rate differs with plurality. We also limit our analysis to mothers who self identified their race as White or Black. We are left with a sample of 3,618,498 of which there are 22,628 events. Although over 99% of our sample is censored, the large sample

Figure 2. Estimated Breast Cancer Mortality Hazard Function by Race



The solid line is the estimated hazard function using our proposed methodology. The dashed line is the estimated hazard function using the life table method in SAS PROC LIFETEST. The top graph is the estimated hazard function for white women, the bottom graph is the estimated hazard function for black women.

Table 7
Infant Mortality Rates by Gender and Race

Gender	Race	n	Mortality	Infant Mortality Rate
Male	White	1,551,213	8709	5.6 per 1000 population
Male	Black	300,349	4010	13.4 per 1000 population
Female	White	1,476,809	6741	4.6 per 1000 population
Female	Black	290,127	3168	10.9 per 1000 population

size allows for ample events to conduct analysis. We examine infant mortality rates by gender and race (see Table 7) and decide to adjust for these covariates in our analysis.

Using the life table method in SAS PROC LIFETEST we obtain an estimated hazard function stratified by gender of the child and race of the mother. As shown in Figure 3, the stratified analysis shows piecewise linear hazard functions with similar trends. We fit a proportional hazards regression model with gender and race as covariates and used the resulting estimates as initial values for the effects of the covariates.

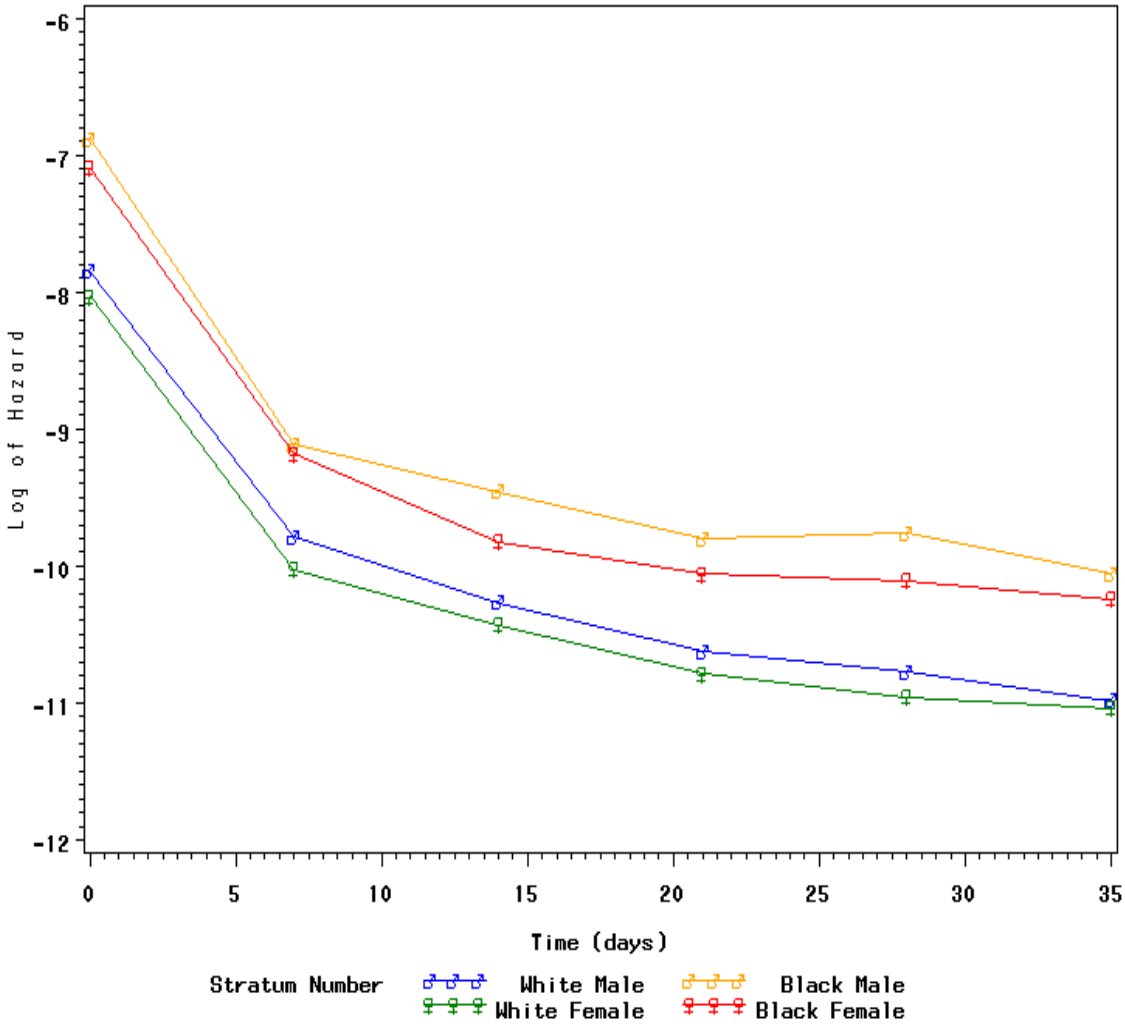
Applying our proposed methodology we estimate the hazard function to have three change points 2.22, 2.80, and 16.58 days, respectively. The estimated log hazard function is,

$$\eta(t) = -1.76 - 5.37t + 2.57(t-2.22)_+ + 2.18(t-2.80)_+ - 0.67(t-16.58)_+ - 1.37female + 1.81black,$$

where $t \geq 0$.

Our analysis shows that black males have the highest hazard, followed by black females, white males, and white females. The estimated log hazard is a decreasing function, with the sharpest decline before the first change point followed by less substantial decreases in all subsequent intervals. Our estimated hazard function indicates that the fist two days of life are the most critical and after surviving the first two weeks the hazard of infant mortality is very close to zero.

Figure 3. Estimated Log Infant Mortality Hazard Function By Gender and Race for the First 35 days of life



7. Discussion

There is some debate in the literature regarding the use of a likelihood ratio test statistic to test the null hypothesis that the model has no change points versus the alternative that the model has one change point. Since these models are nested, a likelihood ratio test seems appropriate as we know these test are most powerful in this setting. However, asymptotic likelihood theory does not apply making the distribution of the test statistic and the proper choice of cut-off points uneasy to identify.

We were able to develop a Wald type test statistic to test the same hypothesis. The test has a known asymptotic distribution and therefore the cut off points are easily identifiable. The use of a Wald test statistic avoids issues surrounding the likelihood ratio test statistic and appropriate cut off values. One drawback of this approach is that calculation of the variance in the denominator of the test statistic requires differentiation of the likelihood. In some instances these calculations are not trivial, we used an approximation function provided in the R statistical software package to calculate the Hessian matrix in these cases.

In the simulations coverage probabilities and model based standard deviations were calculated for all parameters excluding the change points, τ_j. The variances needed for these calculations were calculated using likelihood based methods. Differentiation of the likelihood with respect to the change points is not possible, so these methods do not apply in calculation of variance for change points. Yao (1986) proved the independence of the parameters of $\boldsymbol{\theta} = (\alpha_1, ..., \alpha_{K+1}, \tau_1, ..., \tau_K)$ in the piecewise constant model, we believe these properties to hold in the piecewise linear model as well. Inference about the change-point in both the piecewise constant and piecewise linear model is a nontrivial issue previously discussed in the literature (see, for example Hinkley, 1971; Hinkley, 1970; and Hinkley, 1969).

In most settings our method demonstrated good properties with regard to estimation, model selection, Type I error, and power. However, for samples of size less than 500 with

more than 40% censoring, this is not the case. Estimation, model selection, and power are all effected by the amount and location of censored observations. As the number of change points in the model increases, an increase in the number of uncensored observations is necessary to ensure there are enough uncensored events in each interval between change points. When censoring occurs near the change points valid estimation of these points is not possible.

Our methodology requires use of the Nelder-Mead (Simplex) optimization algorithm, this algorithm requires the user to provide initial starting values for the parameter estimates. From our simulation studies we found these values need to be reasonable but not precise. Although our method is capable of handling an unlimited number of covariates, it assumes that the covariates are additive variables in the model and merely shift the hazard function. This does not allow for the number and value of change points to differ by values of the covariate. When the hazard does not appear to be additive with respect to the covariates a stratified analysis must be done. If the covariate of interest is continuous it must be converted to a categorical variable to allow for a stratified analysis. We believe the development of methods that do not require the covariates to be additive in the model is an area for future research.

An advantage of our method is its ability to appropriately handle multiple comparisons without prior knowledge of the number of hypothesis test needed before reaching the final model. Other approaches would require the analyst to make a guess about the maximum number of change points in the model. If there were more change points than originally assumed, conducting additional hypothesis test would make the overall Type I error exceed α. If fewer test are needed to reach a final model than the α level for each test would be too conservative.

Although traditional smoothing techniques (see, for example Gray (1996)) can be applied to estimate the hazard function. When using these methods the knots and the number of knots are to be prespecified by the data analyst. Our approach uses the data to determine the number and value of the change points. In settings where the is an interest in not

only knowing the trend of the hazard function but where the changes in trend occur, this is a major advantage of our methodology.

Our proposed methods are easily implemented, using maximum likelihood estimation and Wald type test statistics, and can be applied to other important applications. The resulting estimate of the hazard function can also be used for predictive purposes and allows for non-parametric extrapolation of very long term survival by extrapolating the trend from the last change point. Despite its limitations we feel our method has several appealing advantages, coupled with the lack of literature in the area, making it a preferable means of analysis for survival data with multiple change points in the hazard. In future research the authors are interested in adding a spatial component to the model. By adding frailty terms to the model we can model spatial dependence across regions, extending the work of Li and Ryan (2002) to the multiple change point model.

Appendix

A. Likelihood Ratio Test Statistics

A.1 *Likelihood Ratio Test for the Piecewise Constant Model*

In the piecewise constant model, when $\alpha_{k-1} \neq \alpha_k$, the log-likelihood function is

$$logL(\alpha_1, ..., \alpha_k, \tau_1, ..., \tau_{k-1}) = X(\tau_1)\log\alpha_1 + [X(\tau_2) - X(\tau_1)]\log\alpha_2 + ... + [n_u - X(\tau_{k-1})]\log\alpha_k$$
$$- \alpha_1 \sum_{i=1}^{n}(T_i \wedge \tau_1) - \alpha_2 \sum_{i=1}^{n}(T_i \wedge \tau_2 - \tau_1)I(T_i > \tau_1) - ... - \alpha_k \sum_{i=1}^{n}(T_i - \tau_{k-1})I(T_i > \tau_{k-1}),$$

where n_u is the total number of non-censored events. For fixed $\tau_j's$, the maximum likelihood estimates (MLE's) of the parameters $\alpha_1, ..., \alpha_k$ are given by, $\hat{\alpha}_1 = \frac{X(\tau_1)}{\sum_{i=1}^{n}(T_i \wedge \tau_1)}$, $\hat{\alpha}_2 = \frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^{n}(T_i \wedge \tau_2 - \tau_1)I(T_i > \tau_1)}$,..., $\hat{\alpha}_{k-1} = \frac{X(\tau_{k-1}) - X(\tau_{k-2})}{\sum_{i=1}^{n}(T_i \wedge \tau_{k-1} - \tau_{k-2})I(T_i > \tau_{k-2})}$, and

$\hat{\alpha_k} = \frac{n_u - X(\tau_{k-1})}{\sum_{i=1}^{n}(T_i - \tau_{k-1})I(T_i > \tau_{k-1})}$. Substituting the MLE's into $logL$ gives,

$$\begin{aligned}\ell(\tau_1, ..., \tau_{k-1}) = X(\tau_1)\log\left[\frac{X(\tau_1)}{\sum_{i=1}^{n}(T_i \wedge \tau_1)}\right] \\ + [X(\tau_2) - X(\tau_1)]\log\left[\frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^{n}(T_i \wedge \tau_2 - \tau_1)I(T_i > \tau_1)}\right] + ... \\ + [X(\tau_{k-1}) - X(\tau_{k-2})]\log\left[\frac{X(\tau_{k-1}) - X(\tau_{k-2})}{\sum_{i=1}^{n}(T_i \wedge \tau_{k-1} - \tau_{k-2})I(T_i > \tau_{k-2})}\right] \\ + [n_u - X(\tau_{k-1})]\log\left[\frac{n_u - X(\tau_{k-1})}{\sum_{i=1}^{n}(T_i - \tau_{k-1})I(T_i > \tau_{k-1})}\right] - n_u.\end{aligned}$$

On the other hand, when $\alpha_{k-1} = \alpha_k$, the log-likelihood function is

$$logL(\alpha_1, ..., \alpha_{k-1}, \tau_1, ..., \tau_{k-2}) = X(\tau_1)\log\alpha_1 + [X(\tau_2) - X(\tau_1)]\log\alpha_2 + ... + [n_u - X(\tau_{k-2})]\log\alpha_{k-1}$$
$$- \alpha_1\sum_{i=1}^{n}(T_i \wedge \tau_1) - \alpha_2\sum_{i=1}^{n}(T_i \wedge \tau_2 - \tau_1)I(T_i > \tau_1) - ... - \alpha_{k-1}\sum_{i=1}^{n}(T_i - \tau_{k-2})I(T_i > \tau_{k-2}).$$

For fixed $\tau_j's$, the MLE's of the parameters $\alpha_1, ..., \alpha_{k-1}$ are given by, $\hat{\alpha}_1 = \frac{X(\tau_1)}{\sum_{i=1}^{n}(T_i \wedge \tau_1)}$, $\hat{\alpha}_2 = \frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^{n}(T_i \wedge \tau_2 - \tau_1)I(T_i > \tau_1)}$,..., $\hat{\alpha}_{k-1} = \frac{n_u - X(\tau_{k-2})}{\sum_{i=1}^{n}(T_i - \tau_{k-2})I(T_i > \tau_{k-2})}$. Substituting the MLE's into $logL$ gives,

$$\begin{aligned}\ell(\tau_1, ..., \tau_{k-2}) = X(\tau_1)\log\left[\frac{X(\tau_1)}{\sum_{i=1}^{n}(T_i \wedge \tau_1)}\right] \\ + [X(\tau_2) - X(\tau_1)]\log\left[\frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^{n}(T_i \wedge \tau_2 - \tau_1)I(T_i > \tau_1)}\right] + ... \\ + [n_u - X(\tau_{k-2})]\log\left[\frac{n_u - X(\tau_{k-2})}{\sum_{i=1}^{n}(T_i - \tau_{k-2})I(T_i > \tau_{k-2})}\right] - n_u.\end{aligned}$$

The likelihood ratio statistic is equal to $\ell(\tau_1, ..., \tau_{k-1}) - \ell(\tau_1, ..., \tau_{k-2})$ which can be simplified to

$$\begin{aligned}[X(\tau_{k-1}) - X(\tau_{k-2})]\log\left[\frac{X(\tau_{k-1}) - X(\tau_{k-2})}{\sum_{i=1}^{n}(T_i \wedge \tau_{k-1} - \tau_{k-2})I(T_i > \tau_{k-2})}\right] \\ + [n_u - X(\tau_{k-1})]\log\left[\frac{n_u - X(\tau_{k-1})}{\sum_{i=1}^{n}(T_i - \tau_{k-1})I(T_i > \tau_{k-1})}\right] \\ - [n_u - X(\tau_{k-2})]\log\left[\frac{n_u - X(\tau_{k-2})}{\sum_{i=1}^{n}(T_i - \tau_{k-2})I(T_i > \tau_{k-2})}\right].\end{aligned}$$

A.2 *Likelihood Ratio Test for the Piecewise Linear Model*

In the piecewise liner model when $\alpha_{k-1} \neq \alpha_k$, the log-likelihood function is

$$\begin{aligned}
logL(\alpha_0, \alpha_1, ..., \alpha_k, \tau_1, ..\tau_{k-1}, \boldsymbol{\beta}) &= \sum_{i=1}^{n} \Bigg\{ \delta_i(\alpha_0 + \alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + ... + \alpha_k (T_i - \tau_{k-1})_+ + \mathbf{Z}'\boldsymbol{\beta}) \\
&\quad - \int_0^{T_i} \exp(\alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + ... + \alpha_k (u - \tau_{k-1})_+ + \mathbf{Z}'\boldsymbol{\beta}) du \Bigg\} \\
&= \sum_{i=1}^{n} \Bigg\{ \delta_i(\alpha_0 + \alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + ... + \alpha_k (T_i - \tau_{k-1})_+ + \mathbf{Z}'\boldsymbol{\beta}) \\
&\quad - \Bigg[\frac{e^{\alpha_0 + \mathbf{Z}'\boldsymbol{\beta}}}{\alpha_1} (e^{\alpha_1 (T_i \wedge \tau_1)} - 1) + \frac{e^{\alpha_0 + \mathbf{Z}'\boldsymbol{\beta}}}{\alpha_1 + \alpha_2} (e^{\alpha_1 (T_i \wedge \tau_2) + \alpha_2 (T_i \wedge \tau_2 - \tau_1)} - e^{\alpha_1 \tau_1}) I(T_i > \tau_1) + ... \\
&\qquad + \frac{e^{\alpha_0 + \mathbf{Z}'\boldsymbol{\beta}}}{\alpha_1 + \alpha_2 + ... + \alpha_k} \Bigg(e^{\alpha_1 T_i + \alpha_2 (T_i - \tau_1) + ... + \alpha_k (T_i - \tau_{k-1})} \\
&\qquad\quad - e^{\alpha_1 \tau_{k-1} + \alpha_2 (\tau_{k-1} - \tau_1) + ... + \alpha_{k-1} (\tau_{k-1} - \tau_{k-2})} \Bigg) I(T_i > \tau_{k-1}) \Bigg] \Bigg\}
\end{aligned}$$

On the other hand, when $\alpha_{k-1} = \alpha_k$, the log-likelihood function is

$$\begin{aligned}
logL(\alpha_0, \alpha_1, ..., \alpha_{k-1}, \tau_1, ..\tau_{k-2}, \boldsymbol{\beta}) &= \sum_{i=1}^{n} \Bigg\{ \delta_i(\alpha_0 + \alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + ... + \alpha_{k-1} (T_i - \tau_{k-2})_+ + \mathbf{Z}'\boldsymbol{\beta}) \\
&\quad - \int_0^{T_i} \exp(\alpha_0 + \alpha_1 u + \alpha_2 (u - \tau_1)_+ + ... + \alpha_{k-1} (u - \tau_{k-2})_+ + \mathbf{Z}'\boldsymbol{\beta}) du \Bigg\} \\
&= \sum_{i=1}^{n} \Bigg\{ \delta_i(\alpha_0 + \alpha_1 T_i + \alpha_2 (T_i - \tau_1)_+ + ... + \alpha_{k-1} (T_i - \tau_{k-2})_+ + \mathbf{Z}'\boldsymbol{\beta}) \\
&\quad - \Bigg[\frac{e^{\alpha_0 + \mathbf{Z}'\boldsymbol{\beta}}}{\alpha_1} (e^{\alpha_1 (T_i \wedge \tau_1)} - 1) + \frac{e^{\alpha_0 + \mathbf{Z}'\boldsymbol{\beta}}}{\alpha_1 + \alpha_2} (e^{\alpha_1 (T_i \wedge \tau_2) + \alpha_2 (T_i \wedge \tau_2 - \tau_1)} - e^{\alpha_1 \tau_1}) I(T_i > \tau_1) + ... \\
&\qquad + \frac{e^{\alpha_0 + \mathbf{Z}'\boldsymbol{\beta}}}{\alpha_1 + \alpha_2 + ... + \alpha_{k-1}} \Bigg(e^{\alpha_1 T_i + \alpha_2 (T_i - \tau_1) + ... + \alpha_{k-1} (T_i - \tau_{k-2})} \\
&\qquad\quad - e^{\alpha_1 \tau_{k-2} + \alpha_2 (\tau_{k-2} - \tau_1) + ... + \alpha_{k-2} (\tau_{k-2} - \tau_{k-3})} \Bigg) I(T_i > \tau_{k-2}) \Bigg] \Bigg\}.
\end{aligned}$$

From this, the likelihood ratio test statistic is

$$
\begin{aligned}
LRT = \sum_{i=1}^{n} \Bigg\{ & \delta_i(\hat{\alpha}_0 + \hat{\alpha}_1 T_i + \hat{\alpha}_2(T_i - \hat{\tau}_1)_+ + ... + \hat{\alpha}_k(T_i - \hat{\tau}_{k-1})_+ + \mathbf{Z}'\hat{\boldsymbol{\beta}}) \\
& - \Bigg[\frac{e^{\hat{\alpha}_0 + \mathbf{Z}'\hat{\boldsymbol{\beta}}}}{\hat{\alpha}_1}(e^{\hat{\alpha}_1(T_i \wedge \hat{\tau}_1)} - 1) + \frac{e^{\hat{\alpha}_0 + \mathbf{Z}'\hat{\boldsymbol{\beta}}}}{\hat{\alpha}_1 + \hat{\alpha}_2}(e^{\hat{\alpha}_1(T_i \wedge \hat{\tau}_2 + \hat{\alpha}_2(T_i \wedge \hat{\tau}_2 - \hat{\tau}_1)} - e^{\hat{\alpha}_1 \hat{\tau}_1}) I(T_i > \tau_1) + ... \\
& \qquad + \frac{e^{\hat{\alpha}_0 + \mathbf{Z}'\hat{\boldsymbol{\beta}}}}{\hat{\alpha}_1 + \hat{\alpha}_2 + ... + \hat{\alpha}_k} \Bigg(e^{\hat{\alpha}_1 T_i + \hat{\alpha}_2(T_i - \hat{\tau}_1) + ... + \hat{\alpha}_k(T_i - \hat{\tau}_{k-1})} \\
& \qquad\qquad - e^{\hat{\alpha}_1 \hat{\tau}_{k-1} + \hat{\alpha}_2(\hat{\tau}_{k-1} - \hat{\tau}_1) + ... + \hat{\alpha}_{k-1}(\hat{\tau}_{k-1} - \hat{\tau}_{k-2})} \Bigg) I(T_i > \tau_{k-1}) \Bigg] \\
& \qquad - \delta_i(\tilde{\alpha}_0 + \tilde{\alpha}_1 T_i + \tilde{\alpha}_2(T_i - \tilde{\tau}_1)_+ + ... + \tilde{\alpha}_{k-1}(T_i - \tilde{\tau}_{k-2})_+ + \mathbf{Z}'\tilde{\boldsymbol{\beta}}) \\
& + \Bigg[\frac{e^{\tilde{\alpha}_0 + \mathbf{Z}'\tilde{\boldsymbol{\beta}}}}{\tilde{\alpha}_1}(e^{\tilde{\alpha}_1(T_i \wedge \tilde{\tau}_1)} - 1) + \frac{e^{\tilde{\alpha}_0 + \mathbf{Z}'\tilde{\boldsymbol{\beta}}}}{\tilde{\alpha}_1 + \tilde{\alpha}_2}(e^{\tilde{\alpha}_1(T_i \wedge \tilde{\tau}_2) + \tilde{\alpha}_2(T_i \wedge \tilde{\tau}_2 - \tilde{\tau}_1)} - e^{\tilde{\alpha}_1 \tau_1}) I(T_i > \tau_1) + ... \\
& \qquad + \frac{e^{\tilde{\alpha}_0 + \mathbf{Z}'\tilde{\boldsymbol{\beta}}}}{\tilde{\alpha}_1 + \tilde{\alpha}_2 + ... + \tilde{\alpha}_{k-1}} \Bigg(e^{\tilde{\alpha}_1 T_i + \tilde{\alpha}_2(T_i - \tilde{\tau}_1) + ... + \tilde{\alpha}_{k-1}(T_i - \tilde{\tau}_{k-2})} \\
& \qquad\qquad - e^{\tilde{\alpha}_1 \tilde{\tau}_{k-2} + \tilde{\alpha}_2(\tilde{\tau}_{k-2} - \tilde{\tau}_1) + ... + \tilde{\alpha}_{k-2}(\tilde{\tau}_{k-2} - \tilde{\tau}_{k-3})} \Bigg) I(T_i > \tau_{k-2}) \Bigg] \Bigg\},
\end{aligned}
$$

where $(\hat{\alpha}_0, \hat{\alpha}_1, ..., \hat{\alpha}_k, \hat{\tau}_1, ..\hat{\tau}_{k-1}, \hat{\boldsymbol{\beta}})$ are the MLE's of $(\alpha_0, \alpha_1, ..., \alpha_k, \tau_1, ..\tau_{k-1}, \boldsymbol{\beta})$ when $\alpha_{k-1} \neq \alpha_k$, and $(\tilde{\alpha}_0, \tilde{\alpha}_1, ..., \tilde{\alpha}_{k-1}, \tilde{\tau}_1, ..\tilde{\tau}_{k-2}, \tilde{\boldsymbol{\beta}})$ are the MLE's of $(\alpha_0, \alpha_1, ..., \alpha_{k-1}, \tau_1, ..\tau_{k-2}, \boldsymbol{\beta})$ when $\alpha_{k-1} = \alpha_k$.

References

Cai, T., Hyndman, R. J. and Wand, M. (2002). Mixed model-based hazard estimation. *Journal of Computational & Graphical Statistics* **11**, 784–798.

Gijbels, I. and Gürler, U. (2003). Estimation of a change point in a hazard function based on censored data. *Lifetime Data Analysis* **9**, 395–411.

Gray, R. J. (1996). Hazard rate regression usinf ordinary nonparmetric regression smoothers. *Journal of Computational and Graphical Statistics* **5**, 190–207.

Henderson, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika* **77**, 835–843.

Hinkley, D. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56**, 495–504.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1–17.

Hinkley, D. V. (1971). Inference in two-phase regression. *Journal of the American Statistical Association* **66**, 736–743.

Kim, H.-J., Fay, M. P., Feuer, E. J. and Midthune, D. N. (2000). Permutation test for jointpoint regression with applications to cancer rates. *Statistics in Medicine* **19**, 335–351.

Lachin, J. M. (2000). *Biostatistical Methods*. Wiley Series in Probability and Statistics. John Wiley & Sone, Inc., New York.

Lan, K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659–63.

Li, Y. and Ryan, L. (2002). Modeling spatial survival data using semiparametric frailty models. *Biometrics* **58**, 287–297.

Loader, C. R. (1991). Inference for a hazard rate change point. *Biometrika* **78**, 749–757.

Matthews, D. E. and Farewell, V. T. (1982). On testing for a constant hazard against a change-point alternative (Corr: V41 p1103). *Biometrics* **38**, 463–468.

Matthews, D. E. and Farewell, V. T. (1985). On a singularity in the likelihood for a change-point hazard rate model. *Biometrika* **72**, 703–704.

Müller, H.-G. and Wang, J.-L. (1990). Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models. *Biometrika* **77**, 305–314.

Müller, H.-G. and Wang, J.-L. (1994). Change-point models for hazard functions. In Carlstein, E. G., Müller, H.-G. and Siegmund, D., editors, *Change-point problems*, pages 2224–240. Institute of Mathematical Statistics.

Nguyen, G. S., Rogers, G. and Walker, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika* **71**, 299–304.

Pons, O. (2002). Estimation in a Cox regression model with a change-point at an unknown time. *Statistics* **36**, 101–124.

Worsley, K. J. (1988). Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. *Biometrics* **44**, 259–263.

Yao, Y.-C. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Communications in Statistics: Theory and Methods* **15**, 2455–2466.