BOOTSTRAP DISTRIBUTION FOR TESTING A CHANGE IN THE COX PROPORTIONAL HAZARD MODEL

LAM YUK FAI

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy
in
Statistics

June 2000

THE CHINESE UNIVERSITY OF HONG KONG
GRADUATE SCHOOL

The undersigned certify that we have read a thesis, entitled "Bootstrap Distribution for Testing a Change in the Cox Proportional Hazard Model" submitted to the Graduate School by Yuk-Fai Lam ( ) in partial fulfillment of the requirements for the degree of Master of Philosophy in Statistics. We recommend that it be accepted.

Prof. M.G. Gu,
Supervisor

Prof. T.S. Lau

Dr. K.H. Li

Prof B.Y. Jing,
External Examiner

DECLARATION

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

ACKNOWLEDGMENT

I would like to take this opportunity to express my sincere gratitude to my supervisor, Prof. Minggao Gu, for his guidance and generous assistance throughout this thesis.

ABSTRACT

This paper presents a new test and it variations for the fit of proportional hazards model. The procedure is based on the likelihood ratio principle and is inspired by the work in change-point problems. In order to specify the alternative, a change-point hazard ratio model is introduced and the new tests can also be viewed as tests for a change in this model. Sampling distribution of the statistics can be generated through simulations if the baseline hazard function is known. Sampling distributions in a simple case for some small to moderate sample sizes are also given. Large sample properties are discussed and we show that essentially they are the same as those in the change-point problem. We also note that unless the sample size is really large, the use of the asymptotic distribution can be misleading. We apply the bootstrap algorithm and show it works even for small sample size. The tests are compared with the procedures of Wei (1984) and found that they complement each other. We apply the new procedures to real data sets such as male mice data and CGD data. The new procedures are favorably compared and should be adopted in practice.

簡介

本論文介紹一個新的驗證辦法及其變種來驗證比例危險模型的恰當性。 這程序是基於概似比原理， 而靈感是來自轉點問題。 備擇假設將命爲轉點危險比例模型，因此這驗證辦法可視爲轉點存在的驗證。 如果基線危險函數是已知， 抽樣分佈可籍模擬法形成， 我們將提供小型到中等樣本大小的抽樣分佈。 新的驗證辦法的大樣本理論將被探討， 除非樣本大小真的很大， 否則漸近分佈會被誤用。 當這驗證辦法與 Wei 的驗證辦法作一比較， 發覺他們互補不足。 我們應用這新的驗證辦法於一些真實例子當中， 發覺效果更佳及應被使用在實例中。

INTRODUCTION

Analysis of Survival Data is important in clinical trials. This thesis is consist of 3 chapters. Chapter 1 presents some ingredients for analyzing survival data related to the contents in chapter 2. Chapter 2 and chapter 3 are self-contained presenting the main purpose of this thesis. Chapter 3 presents the large sample properties of our proposed test and some final comments. One can skip the basic concepts in chapter 1 and start the reading from the chapter 2.

Contents

Chapter 1

Basic Concepts

1.1 Survival data

Survival data is a dataset describing the time until an event occur on a subject (e.g. a patient, a male mice). The time variable we interested is called survival time. The event of interest usually is (1) death, (2) incidence of a disease or (3) recovering from a disease. Sometime it is termed as failure even a positive event is of interest. The corresponding survival time may represent (1) the number of years patients remain alive until their death, (2) the number of years the patients are healthy or (3) the time interval the patients stay in remission. Survival data normally is incomplete due to the procedure of data collection. Special statistical analysis have to be employed in extracting information from such a incomplete dataset. The coming example illustrates some concepts of survival data.

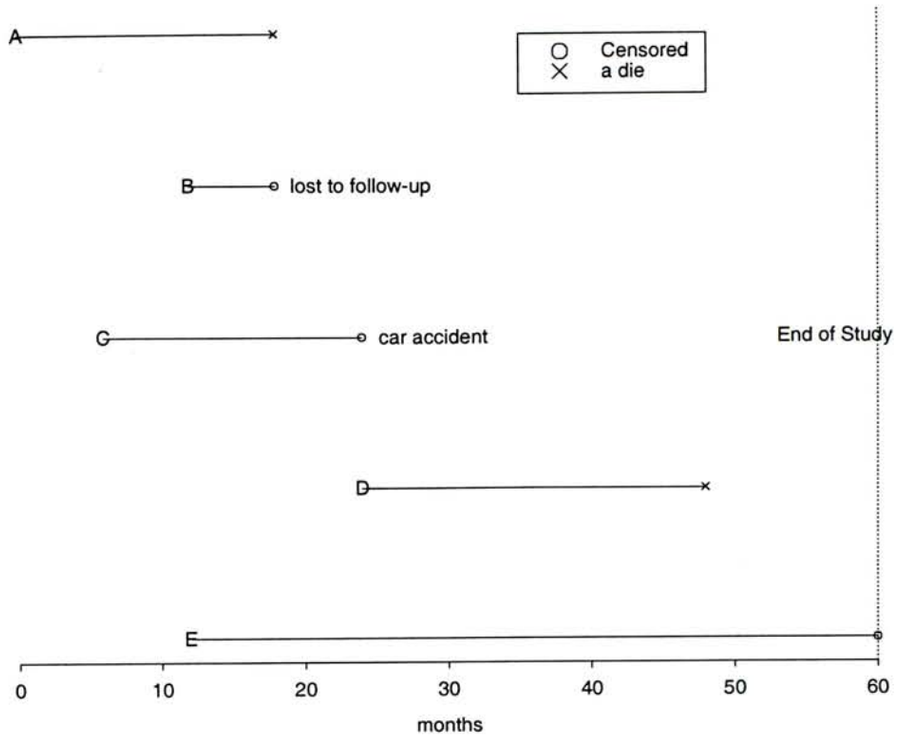
1.1.1 An example

For patients who received a heart transplant, one may want to know how long can the patient survive or how large the risk of death at some specific time after the transplantation. Because of the limitation of time and money, investigation normally has restricted to a fixed period of time, say 5 years. Fig 1.1 shows 5 patients entering the study at different time in the 5-year study period.

Patient A receives a heart transplant at month 0 and die at month 18. He survived 18 months after the transplantation.

Patient B receives a heart transplant at month 12. Because of some reasons, he is unwilling to continue the study and lost to follow-up at month 18. In this case, we can only say that he survived at least 6 months but may be longer. It is

Figure 1.1: A graphical presentation of survival data



an example of incomplete data point in describing the survival time.

Patient C receives a heart transplant at month 6. Because of car accident, he die at month 24. Since we are interested in survival time due to the heart transplant, we say patient C survived at least 18 months. It is an incomplete data point in reference to the survival time of interest.

Patient D receives a heart transplant at month 24 and die at month 48. He survived 24 months after the transplantation.

Patient E receives a heart transplant at month 12. He still survives until the end of study. We say patient E survived at least 48 months. It is an incomplete data point in reference to the survival time of interest.

For the patient A and D, we know exactly their survival time that is the time period between they received a heart transplant and their death. However, for the patient B, C and E, we do not know the exact survival time. Instead, we know they survived up to certain months. These 3 incomplete data points are called censored data points. To be more specific, it is right-censored data since the survival time is incomplete at the right. The reasons of censoring are roughly described as follow:

- no event occur before the end of study (e.g. Patient E),
- the patient, for some reasons, is unwilling to continue the study before an event occur (e.g. Patient B),
- due to the unrelated death (e.g. Patient C).

In the heart transplant example, the period of study is fixed in the data collection plan. The number of censored data is thus a random quantity. It is called random censorship. Under the random censorship, survival data can be best described in this way: Let n be the number of individuals in the study, T_i be the actual survival time for the ith individual we want to investigate, C_i be the censoring variable for ith individual. Because of censoring, we can only observe a vector $(X_i, \delta_i), i = 1, \cdots, n$, where $X_i = \min(T_i, C_i)$, $\delta = 1$ if $X_i = T_i$, $\delta = 0$ if $X_i = C_i$. In this setting, $\delta = 1$ indicate the actual survival time is equal to the observed survival time. While $\delta = 0$ indicate the actual survival time is larger than the observed survival time. Thus the variable, δ, is an indicator variable to show the data is censored or not. The data set corresponding to the heart transplant example is shown in Table 1.1:

Table 1.1: A presentation of survival data in a table

Patient	X	δ
A	18	1
B	6	0
C	18	0
D	24	1
E	48	0

We only focus on the survival data which are right-censored, telling us the data are incomplete or truncated at the right-side. On the other hand, for the left-censored survival data, the data are incomplete or truncated at the left-side. We will not consider left-censored data in this thesis.

1.2 Some important functions

In the survival analysis, it is more natural to express information in survival function, $S(t)$ and hazard function, $\lambda(t)$. Similarly to the probability density

function, $f(t)$, and the cumulative density function, $F(t)$, the distribution of a non-negative random variable, T, can be uniquely specified by the survival function or the hazard function.

1.2.1 Survival function

Survival function, $S(t)$, is the probability that a person can survive longer than time t, i.e. $Pr(T > t)$. It is a non-increasing function with $S(0) = 1$ and $S(\infty) = 0$ similar to Fig 1.2. With this definition, we can show that $S(t) = 1 - F(t)$.

1.2.2 Hazard function

Hazard function is defined as:

$$\lambda(t) = \lim_{\Delta \to 0+} \frac{Pr(t \leq T < t + \Delta \mid T \geq t)}{\Delta}.$$

It is the probability of die at time t given that a person has survived up to time t. This function tells us the simultaneous risk attaching to the subject who still alive at time t. Fig 1.2 show an example of hazard function for exponential distribution ($\lambda = 1$)

With the definitions just presented, the following equations can be derived:

$$\begin{aligned} S(t) &= 1 - F(t), \\ \lambda(t) &= f(t)/S(t), \\ f(t) &= \lambda(t) \exp[-\int_0^t \lambda(u)du]. \end{aligned}$$

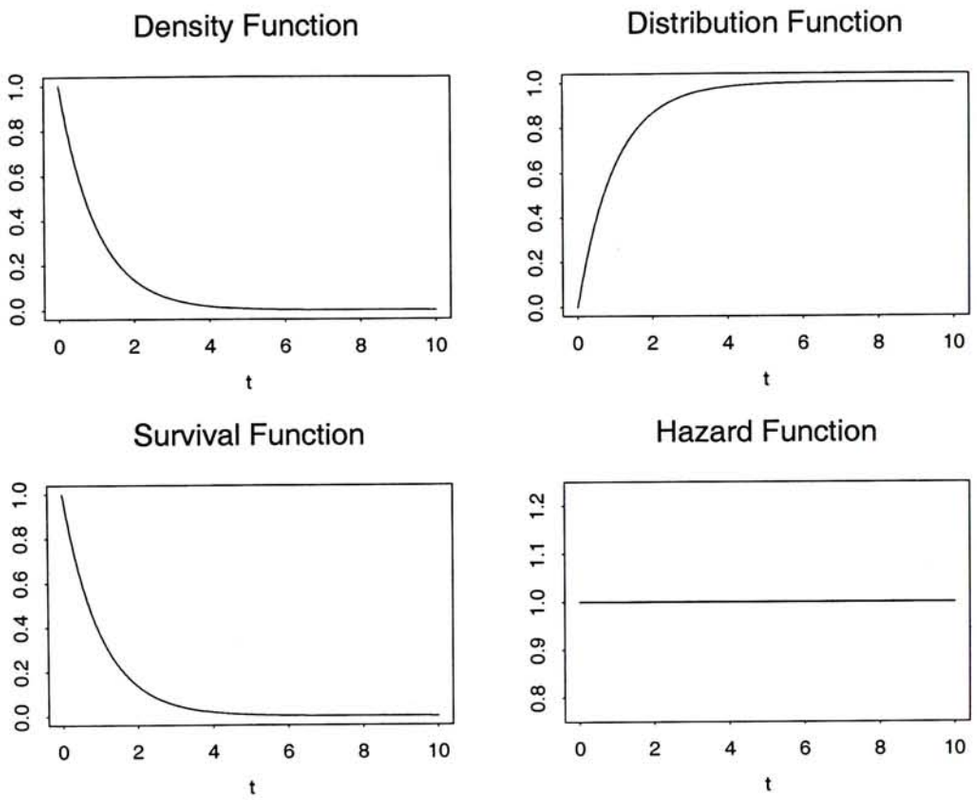
A typical example is a continuous non-negative random variable, T, with $Exponential(\lambda)$ distribution. The four interested functions can be easily written as:

$$\begin{aligned} f(t) &= \lambda \exp(-\lambda t), \\ F(t) &= 1 - \exp(-\lambda t), \\ S(t) &= \exp(-\lambda t), \\ \lambda(t) &= \lambda. \end{aligned}$$

The random variable, T, can be characterized by any one of the four functions above. Fig 1.2 gives a visualization. The hazard plot shows the risk is constant

over the life time for the exponential distribution. This is the reason why exponential distribution is no good to serve as a model for the life time of a machine, which should have an increasing hazard function.

Figure 1.2: Four Important Functions for Exponential(1)



1.3 Cox Proportional Hazards Model

Some variables may have an effect on the survival time. For example, a treatment may prolong one's survival time. The heart transplant example presented in section 1, higher the age may result in higher the hazard rate. This is a typical regression problem for the survival data. In 1972, Cox introduced the Proportional Hazards Model. Denote the survival time by T and the covariates by Z. The model assumes that given $Z = z$ the hazard function at t is,

$$\lambda(t \mid z) = \lambda_0(t) \exp(\beta' Z).$$

It is a semiparametric model since $\lambda_0(t)$ is not specified. Some features of the model are listed below:

- $\lambda_0(t)$ is called the baseline hazard function which depends only on t,

- Z is assumed to be time-independent covariates and so $\exp(\beta' Z)$ is constant over t,
- An exponential linkage is assumed between the covariates and the hazard function. In other words, the covariates link to the logarithm of the hazard function in additive way,
- hazard ratio between two individuals is a constant over t, that is, $\lambda(t \mid z_1)/\lambda(t \mid z_2) = \exp(\beta'(z_1 - z_2))$.

Estimation of the parameters β is based on the maximization of the partial likelihood function:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} \Big[\frac{\lambda(X_i \mid z_i)}{\sum_{j=1}^{n} \lambda(X_i \mid z_j) 1_{[X_j \geq X_i]}}\Big]^{\delta_i} \\ &= \prod_{i=1}^{n} \Big[\frac{\lambda_0(X_i)\exp(\beta z_i)}{\sum_{j=1}^{n} \lambda_0(X_i)\exp(\beta z_j) 1_{[X_j \geq X_i]}}\Big]^{\delta_i} = \prod_{i=1}^{n} \Big[\frac{\exp(\beta z_i)}{\sum_{j=1}^{n} \exp(\beta z_j) 1_{[X_j \geq X_i]}}\Big]^{\delta_i}. \end{aligned}$$

The unspecified baseline hazard is canceled out and partial likelihood depends only on β and the data values. Newton-Raphson Algorithm can be applied to maximize the function $L(\beta)$ by iterating

$$\beta^{i+1} = \beta^i + I^{-1}(\beta^i) U(\beta^i),$$

where

$$\begin{aligned} U(\beta) &= \frac{\partial}{\partial \beta} \log L(\beta), \\ I(\beta) &= -\frac{\partial^2}{\partial \beta^2} \log L(\beta). \end{aligned}$$

1.3.1 A special case

Let consider a case of one covariate which takes values either 0 or 1. It is used as an indicator variable for treatment. The proportional hazard model can be applied to investigate the treatment effect on the survival time.

For $x = 0$, representing the control (or placebo) group, the hazard function is,

$$\lambda(t \mid z = 0) = \lambda_0(t).$$

For $x = 1$, representing the treatment group, the hazard function is,

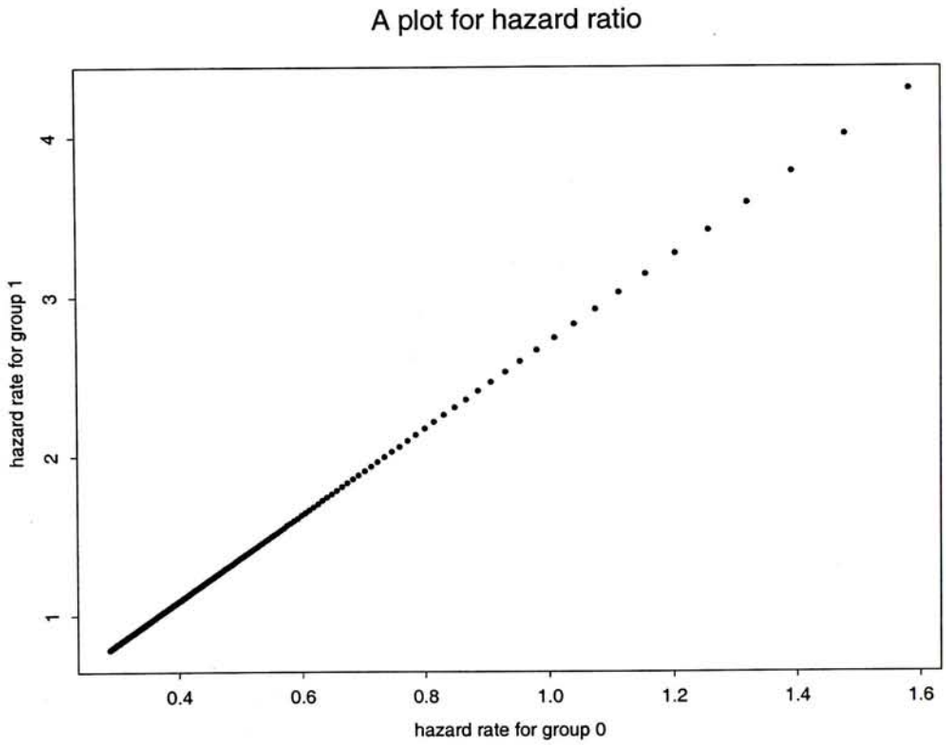
$$\lambda(t \mid z = 1) = \lambda_0(t) \exp(\beta).$$

Consider their hazard ratio,

$$\frac{\lambda(t \mid z = 1)}{\lambda(t \mid z = 0)} = \exp(\beta),$$

which is a constant over t. Fig 1.3 is a plot for $\lambda(t \mid z = 1)$ against $\lambda(t \mid z = 0)$. For this special case, $x = 0$ or 1, visual inspection like Fig 1.3 can indicate a fit of the Cox proportional hazards model.

Figure 1.3: A straight line representing a fit of Cox proportional hazards model



1.3.2 An example (continued)

In the previous heart transplant example, higher the age may result in higher the hazard rate. For each patient who received a heart transplant, their age are measured also. Now the observed data becomes $(X_i, Z_i, \delta_i), i = 1, \cdots, n$, where Z_i is the age of patient and the same definition of X_i and δ_i as previous section. A modified dataset is shown in Table 1.2.

Table 1.2: A presentation of survival data in a table

Patient	X	δ	Z
A	18	1	53
B	6	0	60
C	18	0	52
D	24	1	42
E	48	0	35

1.4 Extension of the Cox Proportional Hazards Model

In the proportional hazards model, the covariates are independent to time. For the case of time-dependent covariates, an extension of proportional hazards model is needed. The extension assumes the hazard function to be

$$\lambda(t \mid z) = \lambda_0(t) \exp(\sum_{i=1}^{p} \beta_i z_i g_i(t)).$$

$g(.)$ is assumed a known function but which is not always the case. Some special cases are discussed below. The first case is transformed back to proportional hazards model. The second case is our interest of this thesis which is the change-point hazard ratio model.

- If $g_i(t) = c, \forall i$, where c is a constant over t, then the extended model reduces to the proportional hazards model.

- If $p = 2, g_1(t) = 1_{[t<\tau]}, g_2(t) = 1_{t\geq\tau}, \beta_1 = \beta, \beta_2 = \gamma$, then the equation becomes

  $$\lambda(t \mid z) = \lambda_0(t) \exp(\beta z 1_{[t<\tau]} + \gamma z 1_{[t\geq\tau]}).$$

  For $t < \tau, \lambda(t \mid z) = \lambda_0(t) \exp(\beta z)$.
  For $t \geq \tau, \lambda(t \mid z) = \lambda_0(t) \exp(\gamma z)$.
  Different hazard function is contributing before and after the time $t = \tau$. The time $t = \tau$ is called change-point. Later, in chapter 2, we will call it a change-point hazard ratio model. Unlike the proportional hazards model, the hazard ratio between two individuals is no longer a constant over t and it does depend on the time, t.

1.5 Bootstrap

Bootstrap is a computer-intensive statistical method dealing with estimation and inference. For illustration, we focus on the situation that $X_1, \cdots, X_n$ follow a distribution F identically and independently. A statistics $T(F) = g(X_1, \cdots, X_n)$ is our interest. We write $T(F)$ instead of T because we want to emphasize it depends on F. Usually, statistical procedure requires to find the

- Variance of $T(F), V$,
- quantiles of $T(F), Q(p)$,
- distribution of $T(F)$, or specifically, $Pr(T(F) \leq s), P(s)$.

When the function g is complicated, deriving the distribution of $T(F)$ is difficult and sometime impossible . Even asymptotic approximate distribution of $T(F)$ is available, it may not work on small sample size.

Bootstrap can help in such a complicated case. Since F is normally not known, we use $\hat{F}$ to approximate it, where $\hat{F}$ is the empirical distribution function of $(X_1 = x_1, \cdots, X_n = x_n)$. In a parametric model with parameter β, we use $(X_1 = x_1, \cdots, X_n = x_n)$ to obtain an estimate of β and then use this estimate to construct $\hat{F}$. Thus, $T(F)$ is approximated by $T(\hat{F})$. With $\hat{F}$, we can simulate a dataset $(x_1, \cdots, x_n)^*$ from $\hat{F}$ and then calculate $t^* = g(X_1^* = x_1^*, \cdots, X_n^* = x_n^*)$. Now, we have a simulated values t^* from $T(\hat{F})$. Repeat this process, we get $(t^{*1}, t^{*2}, \cdots)$. Only a finite number of simulation, say B, can be made which is $(t^{*1}, \cdots, t^{*B})$. The simulation estimates can be done on the above problems:

- Variance estimate of $T(F), \hat{V}_B$:

$$\hat{V}_B = \frac{1}{B}\sum_{i=1}^{B}(t_i^* - \bar{t}^*)^2 \quad \text{where} \quad \bar{t}^* = \frac{1}{B}\sum_{i=1}^{B} t_i^*,$$

- quantiles estimate of $T(F), \hat{Q}_B(p)$:

  $\hat{Q}_B(p)$ = approximately the $p(B+1)^{th}$ ordered value of $(t^{*1}, \cdots, t^{*B})$,

- probability estimate of $Pr(T(F) \leq s), \hat{P}_B(s)$:

$$\hat{P}_B(s) = \frac{1}{B}\sum_{i=1}^{B} 1_{[t_i^* \leq s]}.$$

We can see from the above algorithm that there are 2 stages of approximation. Use the probability estimate as an example. The first is the approximation between F and $\hat{F}$ contributed from the data variability. The second is the approximation between $\hat{P}_B$ and $\hat{P}$ contributed from the finite simulation. As $B \to \infty$, the error between $\hat{P}_B$ and $\hat{P}$ is removed.

Chapter 2

A New Method

2.1 Introduction

It is increasingly popular to use the proportional hazards regression model (Cox, 1972) to analyse life time data. This is partly due to the semiparametric nature of the model and it is often unrealistic to make full parametric assumptions in biomedical statistical inference. However, the assumptions of the model may fail and experts have long warned us of the possible detrimental effects of the misspecification of the model. See, for example, Lagakos & Schoenfeld (1984), Struthers & Kalbfleisch (1986), Lagakos (1988), Lin & Wei (1989).

To test against model misspecification, a series of goodness-of-fit type of statistics and model checking techniques have been proposed, namely, Schoenfeld (1980), Wei (1984) and more recently Therneau, Grambsch & Fleming(1990), Liang, Self & Liu (1990), Lin, Wei & Ying (1993). All of the methods so far proposed are of kolmogrov-Smirnov type, namely, they consider the maximum difference between the observed and the expected under the model and reject the model when the difference gets too large.

Inspired by the work in change-point problems, the present paper proposes a test against any model misspecification based on the (partial) likelihood principle. This test can also be viewed as a test against a change in the proportionality of the hazards in Cox regression model. Even though the final form of our test is also of Kolmogrov-Smirnov type, our test is a likelihood ratio test and therefore should be efficient in some sense. The behavior of the test is very similar to those in the change-point problems (Siegmund, 1988) and the asymptotic distribution of the test is well known in the change-point problem literature.

Another close relative of our test is the change-point hazard rate model pro-

posed by Matthew and Farewell (1982). We shall discuss this model, its relationship to ours and subsequent developments in Section 2. Simulations show an inadequacy of large sample theory. We propose to use bootstrap distribution to implement the test.

This chapter and the coming chapter are organized as follows. In Section 2.2, we will define the test and present some basic properties. Alternative tests given by Liang, Self & Liu (1990) and Wei (1984) are also presented in this section. In Section 2.3, some variations of the test are discussed. In Section 2.4, bootstrap method is applied to the test and compared with the simulation in Section 2.2. We illustrate our test in Section 2.2 and 2.4 by applying them to some real data sets in Section 2.5.

2.2 Definition of the test

2.2.1 Our test statistic

Let T represent some failure time and Z be the corresponding covariate. We assume that the conditional hazard function of T at t given $Z = z$ as

$$\lambda(t \mid z) = \lambda_0(t) \exp[(\beta z I_{[t<\tau]} + \gamma z I_{[t\geq\tau]})],$$

where λ_0 is an unknown baseline hazard function, τ is an unknown parameter. If $H_0 : \beta = \gamma$, it corresponds to the proportional hazards model. If $H_1 : \beta \neq \gamma$, we call it a change-point hazard ratio model.

Denote the actual observations by $(Z_i, X_i, \delta_i), i = 1, \cdots, n$, where $X_i = min(T_i, C_i)$ and C_i is the censoring variable for the ith life time T_i. The test statistic we propose is

$$\Lambda = -2\bigg(\log L(\hat{\beta}, \hat{\beta}, 0) - \sup_{\tau} \log L(\hat{\beta}_\tau, \hat{\gamma}_\tau, \tau) \bigg),$$

where $L(\beta, \gamma, \tau)$ is the partial likelihood function:

$$L(\beta, \gamma, \tau) = \prod_{i=1}^{n} \left[\frac{\exp(\beta Z_i I_{[X_i<\tau]} + \gamma Z_i I_{[X_i\geq\tau]})}{\sum_{j=1}^{n} \exp(\beta Z_j I_{[X_i<\tau]} + \gamma Z_j I_{[X_i\geq\tau]}) I_{[X_j\geq X_i]}} \right]^{\delta_i},$$

$\hat{\beta}$ is the MLE of β under H_0, $\hat{\beta}_\tau$ and $\hat{\gamma}_\tau$ are the MLEs of β, γ under H_1 conditional on fixed τ.

The preceding test statistic originates in the change-point problems. Specifically, it is related to the work by Matthew and Farewell (1982). They proposed a

change-point hazard rate model for the time from remission induction to relapse of patients with leukemia and assumed that the failure rate to be λ_1 before τ and λ_2 after τ. The problem has later been discussed in detail by Mathews, Farewell & Pyke (1985), Worsley (1988), and more recently, Loader (1991), among others. The change-point hazard ratio model can be viewed as a regression generalization to the change-point hazard rate model.

It is interesting to note that Λ is a function of rank statistics when there is no censoring and a function of generalized rank statistics (Prentice, 1978) when censoring is present. This is quite different from the likelihood ratio statistic generated from the change-point hazard rate model, which is sensitive to the actual magnitude of the observation, see Worsley (1988). As a function of τ, $L(\hat{\beta}_\tau, \hat{\gamma}_\tau, \tau)$ changes its value only at an uncensored observation. Consequently, the sup in the definition is taken over the uncensored failure times only.

In practice, the sampling distribution of Λ under H_0 can be obtained by simulation. Since the exact baseline hazard has relatively small effect on the statistic and its sampling distribution, we can always assume the baseline hazard to be constant. Using simulation, some percentage points of the statistics under H_0 with $\beta = \gamma = 0$ and one half of the Z's to be 0 and the other half to be 1 are given in Table 2.1. (The results for the sample size ∞ is derived from asymptotic distribution , see section 3.1) No censoring is assumed and 20,000 simulations were generated for each sample size in Table 2.1.

Compared with the likelihood ratio statistic in the change-point hazard rate model, the percentage points in Table 2.1 is much smaller than those in Table 1 of Worsley (1988). This is because our test is a rank procedure while in change-point hazard rate model, it is not. Same differences exist for the modified tests appearing in the next section.

From Table 2.1, we see that even if the sample size is large than 200, the distribution of the statistic does not stabilize even though it does converge to a limiting distribution (see section 3.1). Blind use of large sample theory could be very misleading.

In general, the distribution under H_0 can be simulated as long as the baseline hazard function is known and the range of the values of the covariates is not too large. If censoring is present, some simulations not presented here show that the simulated table with no censoring can still be used to approximate the sampling distribution, where the sample size corresponds to the number of deaths.

Table 2.1: Percentage points of Λ

Sample	Percent				
size	70	80	90	95	99
10	3.86	4.26	5.84	6.44	10.59
20	4.65	5.45	6.76	8.16	11.16
30	5.03	5.86	7.30	8.69	11.66
40	5.27	6.14	7.61	9.01	12.34
50	5.42	6.30	7.77	9.23	12.46
70	5.63	6.54	8.03	9.52	12.72
100	5.84	6.75	8.28	9.73	13.28
200	6.15	7.12	8.69	10.16	13.42
500	6.67	7.63	9.21	10.70	14.12
∞	8.16	9.14	10.78	12.40	15.93

2.2.2 The alternative test statistic I

Liang, Self & Liu (1990) proposed a different statistic to the change-point problem. The notation for the test is different from our test but the hypothesis are equivalent. With the same meaning of $T, Z, X, \delta, \tau, \lambda_0$ as before, the conditional hazard function of T at t given $Z = z$ is

$$\lambda(t \mid z) = \lambda_0(t) \exp[(\beta + \gamma I_{[t \leq \tau]})z], \tag{2.2.1}$$

When $H_0 : \gamma = 0$, it corresponds to the proportional hazards model. When $H_1 : \gamma \neq 0$, it corresponds to a change-point hazard ratio model. The statistic they proposed is, under H_0

$$M = \sup_{\tau \in [a,b]} S(\tau), \tag{2.2.2}$$

where,

$$S(\tau) = \left. \frac{\frac{\partial l}{\partial \gamma}}{[-\frac{\partial^2 l}{\partial \gamma^2} - (-\frac{\partial^2 l}{\partial \gamma \beta})(-\frac{\partial^2 l}{\partial \beta^2})(-\frac{\partial^2 l}{\partial \gamma \beta})]^{1/2}} \right|_{\beta = \hat{\beta}, \gamma = 0}, \tag{2.2.3}$$

and $\exp(l)$ is the partial likelihood function:

$$\exp(l) = \prod_{i=1}^{n} \left[\frac{\exp[(\beta + \gamma I_{[X_i \leq \tau]})Z_i]}{\sum_{j=1}^{n} \exp[(\beta + \gamma I_{[X_i \leq \tau]})Z_j] I_{[X_j \geq X_i]}} \right]^{\delta_i}, \tag{2.2.4}$$

$\hat{\beta}$ is the MLE of β under H_0.

This statistic requires a known range for τ, say $\in [a, b]$. In the case of unknown range, $\tilde{M} = \sup_{X_{(1)} \leq \tau < X_{(n)}} S(\tau)$, (except the sample size ∞, which were obtained from the asymptotic distribution) we did a 20,000 simulations, similar to previous subsection, for each sample size in Table 2.2.

Table 2.2: Percentage points of $\tilde{M}$

Sample	Percent				
size	70	80	90	95	99
10	1.40	1.63	1.94	2.19	2.88
20	1.60	1.81	2.12	2.39	2.88
30	1.71	1.92	2.24	2.50	3.05
40	1.77	1.98	2.30	2.55	3.05
50	1.81	2.04	2.35	2.61	3.12
70	1.88	2.10	2.42	2.69	3.18
100	1.93	2.16	2.48	2.74	3.26
200	2.03	2.26	2.58	2.86	3.38
∞	2.86	3.02	3.28	3.52	3.99

Again, from Table 2.2, the distribution of the statistic, $\tilde{M}$ does not stabilize even when sample size is large than 200.

2.2.3 The alternative test statistic II

Wei (1984) proposed a statistic to test a fit of proportional hazards model. With the same meaning of $T, Z, X, \delta, \lambda_0$ as before, we present his test in the regression form. The null hypothesis of the test is describing the proportional hazards model, i.e. H_0 is

$$\lambda(t \mid z) = \lambda_0(t) \exp(\beta z).$$

The statistic he proposed is, under H_0

$$T_n = \sup_{0<\tau<\infty} \frac{\mid \frac{\partial}{\partial \beta} C_n(\beta, \tau) \mid}{[-\frac{\partial^2}{\partial \beta^2} C_n(\beta, \infty)]^{1/2}} \Bigg|_{\beta=\hat{\beta}},$$

where $C_n(\beta, \infty)$ is the logarithm of the partial likelihood function with

$$C_n(\beta,\tau)=\sum_{i=n}^{n}\delta_i\left[\beta z_i-\log\left[\sum_{j=1}^{n}\exp(\beta z_j)1_{[X_j\geq X_i]}\right]\right]1_{[X_i\leq\tau]},$$

and $\hat{\beta}$ is the MLE of β under H_0 satisfying

$$C_n(\hat{\beta},\infty)=\max_{\beta}C_n(\beta,\infty).$$

It can be proved that T_n converges to the maximum of a Brownian bridge process. H_0 is rejected if T_n is too large.

2.3 Variations of the test

2.3.1 Restricted test

As in change-point problems, the likelihood ratio test statistic tends to put an undue weight on the tail where there is little information to actually detect a change-point. In practice a change-point may be thought to lie within a restricted range of the observations. It is a common practice to restrict the change-point to lie between the pth-quantile and the $(1-p)$th-quantile of the sample, where $0<p<.5$.

In the Cox proportional hazards model, because of censoring, it seems more appropriate to restrict the change in the following way: let $I(\tau,\beta)$ be the information under H_0 for observations before τ, i.e.

$$I(\tau,\beta)=\sum_{i:X_i<\tau}\delta_i\left(\frac{S_2(X_i,\beta)}{S_0(X_i,\beta)}-\frac{S_1^2(X_i,\beta)}{S_0^2(X_i,\beta)}\right),$$

where for $k=0,1,2$

$$S_k(t,\beta)=\sum_{i=1}^{n}Z_i^k\exp(\beta Z_i)I_{[X_i\geq t]}/n,$$

and let $t_\tau=I(\tau,\hat{\beta})/I(\infty,\hat{\beta})$. Then a restricted statistic can be defined by

$$\Lambda_p=-2\left(\log L(\hat{\beta},\hat{\beta},0)-\sup_{p\leq t_\tau\leq 1-p}\log L(\hat{\beta}_\tau,\hat{\gamma}_\tau,\tau)\right).$$

Some percentage points for $p=.05$ and $p=.1$ under H_0 and same conditions as in Table 2.1 are presented in Table 2.3 and Table 2.4. (The results for the sample size ∞ is derived from asymptotic distribution, see section 3.1) We see that the distribution stabilizes to some degree as the sample size gets beyond 100.

Table 2.3: Percentage points of Λ_p

(a) $p = 0.05$

Sample	Percent				
size	70	80	90	95	99
10	3.86	4.26	5.84	6.43	10.59
20	4.65	5.45	6.72	8.16	11.11
30	4.96	5.79	7.21	8.58	11.58
40	5.24	6.11	7.57	9.01	12.23
50	5.42	6.30	7.77	9.23	12.46
70	5.55	6.54	8.04	9.52	12.72
100	5.33	6.42	8.11	9.64	13.28
200	5.28	6.29	7.99	9.58	13.27
500	5.40	6.35	8.14	9.81	13.71
∞	5.65	6.67	8.27	9.78	13.47

Table 2.4: Percentage points of Λ_p

(b) $p = 0.1$

Sample	Percent				
size	70	80	90	95	99
10	3.86	4.25	5.84	6.43	10.59
20	4.65	5.49	6.72	8.16	11.09
30	4.94	5.79	7.21	8.58	11.58
40	4.81	5.98	7.57	9.01	12.21
50	4.67	5.75	7.66	9.23	12.46
70	4.67	5.63	7.20	8.87	12.48
100	4.67	5.63	7.23	8.91	12.56
200	4.79	5.82	7.39	8.95	12.56
500	4.83	5.76	7.32	8.89	12.57
∞	5.11	6.09	7.69	9.25	13.00

Table 2.5: Percentage points of M_p

(a) $p = 0.05$

Sample	Percent				
size	70	80	90	95	99
10	1.34	1.62	1.86	2.15	2.70
20	1.59	1.81	2.11	2.38	2.88
30	1.69	1.91	2.23	2.49	3.05
40	1.76	2.98	2.30	2.55	3.05
50	1.78	2.03	2.35	2.61	3.12
70	1.82	2.07	2.41	2.68	3.18
100	1.85	2.10	2.45	2.73	3.26
200	1.89	2.15	2.49	2.80	3.34
∞	2.38	2.58	2.88	3.13	3.67

For the statistic $M = \sup_{\tau\in[a,b]} S(\tau)$, if the range for τ is not known, let define

$$M_p = \sup_{p\le t_\tau\le 1-p} S(\tau), \tag{2.3.5}$$

with the same definition of t_τ.

Similar simulations for the percentage points of M_p are presented in Table 2.5 and Table 2.6 and we can see an improvement as compared with Table 2.2.

2.3.2 Adjusting for other covariates

It is rare that a clinical data consists of only one covariate. If we have more than one covariate, it is desirable to check the assumption of proportional hazards in the main covariate adjusting for the other covariate(s). The omnibus test of Wei (1984) works only for one covariate or when the off diagonal element of the covariance matrix is zero, see Therneau et al. (1990).

Assume that W is the other covariate, and write $\tilde{L}(\beta, \gamma, \tau, \theta)$ as the partial likelihood adding θW_i to the corresponding terms inside the exponent in the definition of $L(\beta, \gamma, \tau)$. There are two ways of generalizing the test of the last section adjusting covariate W. The first is to construct conditional partial likelihood ratio statistic by

Table 2.6: Percentage points of M_p

(b) $p = 0.1$

Sample	Percent				
size	70	80	90	95	99
10	1.34	1.62	1.86	2.15	2.70
20	1.54	1.78	2.09	2.37	2.87
30	1.65	1.89	2.22	2.49	3.05
40	1.70	1.93	2.27	2.54	3.05
50	1.69	1.96	2.31	2.60	3.11
70	1.71	1.97	2.33	2.63	3.17
100	1.73	1.98	2.36	2.67	3.22
200	1.77	2.03	2.39	2.70	3.26
∞	2.26	2.47	2.77	3.04	3.61

$$\tilde{\Lambda}^c = -2\Big(\log \tilde{L}(\hat{\beta}, \hat{\beta}, 0, \hat{\theta}) - \sup_{\tau} \log \tilde{L}(\hat{\beta}_\tau, \hat{\gamma}_\tau, \tau, \hat{\theta}) \Big),$$

where $\hat{\theta}$ is the MLE of θ under the null hypothesis of no change. The second way is based on the likelihood ratio principle:

$$\tilde{\Lambda} = -2\Big(\log \tilde{L}(\hat{\beta}, \hat{\beta}, 0, \hat{\theta}) - \sup_{\tau} \log \tilde{L}(\hat{\beta}_\tau, \hat{\gamma}_\tau, \tau, \hat{\theta}_\tau) \Big),$$

where $\hat{\theta}_\tau$ is the MLE of θ under H_1 that there is a change.

It seems that $\tilde{\Lambda}^c$ is simpler to implement than $\tilde{\Lambda}$ but the latter should be more powerful than $\tilde{\Lambda}^c$ under the alternative hypothesis. Under the null hypothesis and any particular pattern of covariates distribution, the simulated percentage points of these two statistics can be obtained by modifying the program which gives Table 2.1. The large sample properties of these two statistics are discussed in Section 3.1. The restricted version of these two statistics can be constructed in an obvious way and their percentage points under the null hypothesis can also be simulated easily if the baseline hazard function is known.

2.4 Apply with bootstrap

In practice, the percentage points of the statistic Λ_p (or M_p) are not available if we do not know the baseline hazard function, $\lambda_0(t)$, and the actual value of β. However, with the bootstrap algorithm introduced in section 1.5, an approximate percentage points of Λ_p (or M_p) can be approximated. Recall , given Z_i, C_i is the censoring variable for the *ith* life time T_i and observe $X_i = \min(T_i, C_i)$ with $\delta_i = I_{[T_i \leq C_i]}$. Suppose $T_i \sim F$ and $C_i \sim G$, then the distribution of the statistic Λ_p (or M_p), can be expressed as a function of F and G, i.e. $\Lambda_p(F, G)$ (or $M_p(F, G)$). With appropriate estimates of F and G, say $\hat{F}$ and $\hat{G}$, we can approximate the corresponding distribution by $\Lambda_p(\hat{F}, \hat{G})$ (or $M_p(\hat{F}, \hat{G})$). Assuming the censoring time also satisfies the proportional hazards model:

$$\lambda_C(c \mid z) = \lambda_{0,C}(c) \exp(\beta_C z)$$

Let $A_{0,T}(t)$ and $A_{0,C}(c)$ be the baseline cumulative hazard function of T_i and C_i respectively under H_0 and which can be estimated by the Breslow estimator

$$\hat{A}_{0,T}(t) = \sum_{i=1}^{n} I_{[X_i \leq t]} \frac{\delta_i}{\sum_{j=1}^{n} \exp(\hat{\beta}_T z_j) I_{[X_j \geq X_i]}}, \tag{2.4.6}$$

$$\hat{A}_{0,C}(c) = \sum_{i=1}^{n} I_{[X_i \leq c]} \frac{1 - \delta_i}{\sum_{j=1}^{n} \exp(\hat{\beta}_C z_j) I_{[X_j \geq X_i]}}, \tag{2.4.7}$$

where $\hat{\beta}_T$ is the MLE of β_T under H_0 with dataset (z_i, X_i, δ_i) and $\hat{\beta}_C$ is the MLE of β_C under H_0 with dataset $(z_i, X_i, 1 - \delta_i)$,

With these estimates, we use the following to estimate the survivor function of T_i and C_i given the covariate

$$1 - \hat{F}(t \mid z) = \left[\exp\left(- \hat{A}_{0,T}(t) \right) \right]^{\exp(\hat{\beta}_T z)}, \tag{2.4.8}$$

$$1 - \hat{G}(c \mid z) = \left[\exp\left(- \hat{A}_{0,C}(c) \right) \right]^{\exp(\hat{\beta}_C z)}. \tag{2.4.9}$$

In parallel to section 2.2, generate T_i^* from (2.4.8) and C_i^* from (2.4.9) given $z_i^* = z_i$. Denote the bootstrap observations by $(z_i^*, X_i^*, \delta_i^*), i = 1, \cdots, n$, where $X_i^* = min(T_i^*, C_i^*)$ with $\delta_i^* = I_{[T_i^* \leq C_i^*]}$. Again, the percentage points of Λ_p^* (or M_p^*) can be computed with 20,000 simulations. Then use these as an approximate percentage points of the actual percentage values of Λ_p (or M_p.)

Table 2.7: Percentage points of Λ_p^*

$p = 0$					
Sample	Percent				
size	70	80	90	95	99
10	3.81	4.26	5.84	6.42	10.60
20	4.33	5.09	6.49	7.87	10.68
30	4.71	5.55	6.98	8.31	11.42
40	4.82	5.71	7.15	8.58	11.72
50	4.90	5.80	7.29	8.71	12.05
70	5.07	5.97	7.48	8.87	12.08
100	5.29	6.25	7.81	9.35	12.53
200	5.62	6.57	8.14	9.71	13.22

Table 2.7, Table 2.8 and Table 2.9 show the approximated percentage points of Λ_p^* based on one simulated data set from the model as in Table 2.1, Table 2.3 and Table 2.4. Table 2.10, Table 2.11 and Table 2.12 show the approximated percentage points of M_p^* based on one simulated data set from the model as in Table 2.2, Table 2.5 and Table 2.6. We can see the approximated percentage points agree closely with the actual percentage points.

Table 2.13, Table 2.14 and Table 2.15 show the approximated percentage points of M_p^* based on several simulated dataset of simple size 100 from the model as in Table 2.2, Table 2.5 and Table 2.6 (the percentage points from true distribution and asymptotic distribution are also included). We can see the values of the approximated percentage points do not vary very much and give a better approximation than the asymptotic distribution. Thus, bootstrap should be employed to implement the test for small sample size.

2.5 Examples

In this section, we analyze some real data sets using our methods. The p-value reported in this section are all obtained by the bootstrap distribution with 20,000 simulations. The results are compared with the test of Wei (1984). The examples show that our test is more efficient than Wei's test in their particular settings. The reason why these happen is discussed with the representation of Λ in the

Table 2.8: Percentage points of Λ_p^*

$p = 0.05$

Sample	Percent				
size	70	80	90	95	99
10	3.76	4.25	5.82	6.42	10.60
20	4.33	5.09	6.49	7.86	10.68
30	4.65	5.52	6.97	8.31	11.42
40	4.81	5.70	7.15	8.58	11.72
50	4.77	5.73	7.27	8.71	12.05
70	4.90	5.87	7.43	8.85	12.08
100	5.06	6.06	7.69	9.32	12.53
200	5.28	6.25	7.87	9.48	13.03

Table 2.9: Percentage points of Λ_p^*

$p = 0.1$

Sample	Percent				
size	70	80	90	95	99
10	3.76	4.25	5.82	6.42	10.60
20	4.18	5.02	6.46	7.85	10.68
30	4.42	5.38	6.88	8.26	11.42
40	4.38	5.38	6.97	8.52	11.72
50	4.43	5.40	7.05	8.58	12.00
70	4.61	5.54	7.13	8.58	12.00
100	4.69	5.69	7.28	8.89	12.18
200	4.72	5.67	7.24	8.85	12.52

Table 2.10: Percentage points of M_p^*

$p = 0$

Sample	Percent				
size	70	80	90	95	99
10	1.30	1.53	1.80	2.15	2.70
20	1.55	1.77	2.07	2.33	2.86
30	1.58	1.80	2.11	2.37	2.89
40	1.73	1.94	2.24	2.50	3.04
50	1.91	2.13	2.43	2.69	3.23
70	1.99	2.21	2.52	2.78	3.27
100	1.89	2.12	2.45	2.73	3.25
150	1.97	2.19	2.51	2.79	3.29
200	2.03	2.26	2.58	2.86	3.43

Table 2.11: Percentage points of M_p^*

$p = 0.05$

Sample	Percent				
size	70	80	90	95	99
10	1.28	1.51	1.78	2.12	2.70
20	1.54	1.76	2.07	2.33	2.86
30	1.57	1.80	2.11	2.37	2.89
40	1.72	1.94	2.24	2.50	3.04
50	1.87	2.11	2.42	2.68	3.23
70	1.90	2.15	2.50	2.77	3.27
100	1.82	2.08	2.43	2.72	3.25
150	1.86	2.10	2.45	2.75	3.29
200	1.89	2.14	2.51	2.80	3.39

Table 2.12: Percentage points of M_p^*

$p = 0.1$					
Sample	Percent				
size	70	80	90	95	99
10	1.28	1.51	1.78	2.12	2.70
20	1.50	1.73	2.05	2.32	2.85
30	1.53	1.77	2.10	2.37	2.89
40	1.66	1.90	2.23	2.50	3.04
50	1.74	2.01	2.37	2.66	3.20
70	1.76	2.02	2.39	2.70	3.25
100	1.71	1.98	2.35	2.66	3.24
150	1.75	2.00	2.36	2.67	3.22
200	1.77	2.03	2.41	2.70	3.32

Table 2.13: Percentage points of M_p^* based on different datasets

$p = 0$					
Sample	Percent				
size	70	80	90	95	99
100	1.89	2.12	2.45	2.73	3.25
100	1.90	2.14	2.46	2.73	3.26
100	1.91	2.14	2.46	2.72	3.25
100	1.87	2.09	2.42	2.69	3.22
100	1.91	2.13	2.44	2.71	3.25
100	1.91	2.13	2.45	2.73	3.26
100	1.94	2.16	2.47	2.74	3.26
100	1.90	2.13	2.45	2.73	3.26
100	1.85	2.08	2.40	2.67	3.19
true	1.93	2.16	2.48	2.74	3.26
∞	2.86	3.02	3.28	3.52	3.99

Table 2.14: Percentage points of M_p^* based on different datasets

$p = 0.05$					
Sample	Percent				
size	70	80	90	95	99
100	1.82	2.08	2.43	2.72	3.25
100	1.83	2.09	2.43	2.72	3.25
100	1.83	2.08	2.42	2.71	3.24
100	1.81	2.06	2.40	2.69	3.22
100	1.83	2.09	2.42	2.70	3.25
100	1.83	2.08	2.43	2.72	3.26
100	1.85	2.11	2.45	2.73	3.26
100	1.82	2.08	2.43	2.72	3.25
100	1.79	2.05	2.39	2.66	3.19
true	1.85	2.10	2.45	2.73	3.26
∞	2.38	2.58	2.88	3.13	3.67

Table 2.15: Percentage points of M_p^* based on different datasets

$p = 0.1$					
Sample	Percent				
size	70	80	90	95	99
100	1.71	1.98	2.35	2.66	3.24
100	1.72	1.98	2.35	2.65	3.22
100	1.71	1.97	2.34	2.64	3.21
100	1.71	1.97	2.32	2.64	3.19
100	1.73	1.99	2.33	2.62	3.22
100	1.72	1.98	2.34	2.65	3.22
100	1.74	2.00	2.36	2.66	3.20
100	1.72	1.98	2.34	2.64	3.23
100	1.70	1.96	2.31	2.60	3.17
true	1.73	1.98	2.36	2.67	3.22
∞	2.26	2.47	2.77	3.04	3.61

next section.

2.5.1 Male mice data

This data set is given in Hoel (1972) and goodness-of-fit for proportional hazards model was tested by Wei (1984). Two groups of sizes 22 and 29 of male mice were given 300 rads of radiation and followed for cancer incidence. For more detailed descriptions, see the two papers mentioned above. In Wei (1984), the goodness-of-fit test yields a value 1.29 with an approximate-p value 0.06. In our case, $\Lambda_{.05} = 11.24$, which corresponds to an p-value 0.02 If the size of the test is 5 percent, Λ will reject the null hypothesis that the data set obeys the proportional hazards model while Wei's test won't.

2.5.2 Stanford heart transplant data

Stanford heart transplant data has been analyzed using different methods in many papers. See, for example, Miller (1981). The reason this data set has received so much attention is partly because the proportional hazard model does not fit it very well. We analyzed the data set as it is in Miller and Halpern (1982). We only included the patients who actually received a transplant. The total number of patients is 157, among whom 55 were censored. We assume the age covariate is of interest. It seems that the logarithm of the age covariate fits the data better than without the transformation. Using this transformation, we get $\hat{\beta} = .815$ with standard deviation .399. Our test with $p = .1$ yields a value $\Lambda_{.1} = 7.83$, which corresponds to an p-value 0.10. Wei's test yields a value 0.49 which corresponds to an approximate p-value 0.97 indicating a very good fit.

2.5.3 CGD data

The chronic granulotomous disease trial was a placebo controlled randomized trial. The data set is described and analyzed using Cox regression in Harrington and Fleming (1991). We only included the patients at the first infection. The total number of patients is 128, among whom 44 have died. Assuming the treatment covariate is of interest, a test of fit $\Lambda_{.1} = 7.32$ yields an p-value 0.067, which is not significant, but to a careful investigator, it suggests some degree of dissatisfaction with the proportional hazards model. In comparison, Wei's test gives 0.87 with approximate p-value 0.45, which indicates a very good fit.

Chapter 3

Large Sample Properties and Discussions

A brief discussion of the large sample properties of the test is given in the Section 3.1. Some final comments and discussions on possible further researches are given in Section 3.2.

3.1 Large sample properties and relationship to goodness of fit test

3.1.1 Large sample properties of Λ and Λ_p

For fixed τ, define

$$\Lambda(\tau) = 2[\log L(\hat{\beta}_\tau, \hat{\gamma}_\tau, \tau) - \log L(\hat{\beta}, \hat{\beta}, 0)]. \tag{3.1.1}$$

It is easy to see that $\Lambda = \sup_\tau \Lambda(\tau)$.

Using asymptotic representations of $\hat{\beta}, \hat{\beta}_\tau, \hat{\gamma}_\tau$ and two-term Taylor expansion, after some lengthy algebra, we can show that (details are presented in the Appendix)

$$\Lambda(\tau) = \{IU(\tau) - I(\tau)U\}^2/\{I(\tau)[I - I(\tau)]I\} + o_p(1), \tag{3.1.2}$$

where $U(\tau)$ and $I(\tau)$ are the first derivative and minus the second derivative of $\log L(\beta, \gamma, \tau)$ with respect to β respectively, and $U = U(\infty)$ and $I = I(\infty)$. Since $U(\tau)/I^{1/2}$ converges weakly to a time changed-Brownian motion process in τ (Anderson and Gill, 1982) and $I(\tau)$ is a consistent estimator for the variance of $U(\tau)$,

we see that $\Lambda(\tau)$ converge to $B^2(t)/t(1-t)$, the square of a Ornstein-Uhlenbeck process at $t = EI(\tau)/EI$ and Λ converges to $\sup_{0<t<1} B^2(t)/t(1-t)$. It is easy to see that for $0 < p < .5, \Lambda_p$ converges in distribution to $\sup_{p<t<1-p} B^2(t)/t(1-t)$.

The goodness-of-fit test of Wei (1984) is equivalent to

$$\sup_{\tau} \mid U(\tau) - t_\tau U \mid /I^{1/2}, \tag{3.1.3}$$

where $t_\tau = I(\tau)/I$. The square of the test is equivalent to $\sup_\tau \Lambda(\tau)t_\tau(1-t_\tau)$, a weighted version of our test, which is more powerful to test a change if the change occurs in the middle, or, if the change occurs at τ, where $I(\tau)/I$ is close to 1/2. In the three examples of Section 2.5, the maximums occur at τ where $T(\tau)/I$ is close to 0 or 1, and therefore our test is more powerful than the goodness-of-fit test of Wei. If the opposite happens, Wei's test is more powerful but the performance of our test is not bad. In practice, if there is prior information as to where the change will occur, we can choose the appropriate test to achieve the maximum efficiency. Because of the structure of the Wei's test, there is very little power to detect a change if the change-point is close to 0.1 or 0.9 while our procedure still preserves some power to detect a change if the change-point is in the middle.

3.1.2 Large sample properties of $\tilde{\Lambda}^c$ and $\tilde{\Lambda}$

The asymptotic distributions of both statistics $\tilde{\Lambda}^c$ and $\tilde{\Lambda}$ under the null hypothesis do not converge to any well known probability distribution in general, which is, at first sight, rather surprising. On the other hand, this is a possible reason why people have been refrained from discussing the covariate adjustment in this situation other than in the special case when Z and W are independent. From the asymptotic expressions of these two statistics (see Appendix), we can see that if $EI_{zw}(\tau)/EI_{zz}(\tau)$ is a constant in τ, then both of them have the same asymptotic distribution as that of Λ, where $I_{zz}(\tau)$ is minus the second partial derivative of $\log L(\beta, \gamma, \tau, \theta)$ with respect to β and $I_{zw}(\tau)$ is minus the cross derivative with respect to β and θ. Note that if Z and W are independent, then $EI_{zw}(\tau) = 0$. Obviously, restricted versions of $\tilde{\Lambda}^c$ and $\tilde{\Lambda}$ converge to that of Λ under the same condition.

3.2 Discussions

In contrast to recent work in this field, this paper puts emphasis on the change in the hazards ratio alternative rather than trying to find the alternative functional form when the simple proportional hazards model does not fit. It is natural in our approach that if the initial test rejects the null hypothesis, a change-point hazard ratio model can be tried, where the change occurs at the point which maximized the likelihood function. There are certainly advantages and disadvantages. One advantage is that it is simple to introduce one more parameter (as opposed to introducing $z^2, \cdots$) and it is easy to interpret(hazard ratio change from $\exp(\beta)$ to $\exp(\gamma)$ for life time larger that τ with 1 unit of covariate increase). Another advantage is that the alternative is within the framework of Cox regression model (as oppose to generalized risk model) and the computation for the change-point hazard ratio model can easily be carried out with available statistical packages. Two disadvantages are that it introduces one more parameter and the model represents a non-smooth change in the hazard ratio.

Since our procedure is based on the likelihood ratio principle, the efficiency of the procedure is expected to be good. We emphasize here that the procedure developed in this paper in only a first step in linking the model checking problems of survival analysis to the change-point problems. Under the change-point hazard ratio model, some other procedures for the model checking problems of survival data can be readily found from their counterparts developed for the change-point problems. For example, the Bayesian procedures originated by Chernoff and Zacks (1964) can be generalized to the model checking problems. See also Zacks (1991). Similar tests based on the alternative of more than one change-point in change-point problems.

In multivariate covariate situations, we might have more than one change, with different changes corresponding to different covariates. Likelihood ratio tests certainly can be constructed based on these models.

In recent work of Lin, Wei & Ying (1993), new omnibus tests for the fit of proportional hazards model have been proposed and it is demonstrated that it is very powerful. Their test is still of the Kolmogrov-Smirnov type while the test proposed here is of likelihood ratio type. It would be interesting to see what will happen if the two ideas blend together.

Exact distribution can be simulated only when we know the baseline hazard function. The large sample percentage points cannot be used because of inade-

quacy. Bootstrap is thus introduced to find better percentage points for the test statistics. Similar approach may be applied to other statistics having a limiting distribution if bootstrap distribution is better.

Appendix

The asymptotic expression for $\Lambda(\tau)$:

Under the null hypothesis, we have

$$\begin{aligned}
\hat{\beta} - \beta &= U/I + o_p(n^{-1/2}), \\
\hat{\beta}_\tau - \beta &= U(\tau)/I(\tau) + o_p(n^{-1/2}), \\
\hat{\gamma}_\tau - \gamma &= [U - U(\tau)]/[I - I(\tau)] + o_p(n^{-1/2}).
\end{aligned}$$

Define $U(\beta, \tau)$ to be the first derivative of $\log L(\beta, \gamma, \tau)$ with respect to β. Using a two-term Taylor expansion we have

$$\begin{aligned}
\Lambda(\tau) &= 2\Big[(\hat{\beta}_\tau - \hat{\beta})U(\hat{\beta}, \tau) - (\hat{\beta}_\tau - \hat{\beta})^2 I(\tau)/2 \\
&\quad + (\hat{\gamma}_\tau - \hat{\beta})[-U(\hat{\beta}, \tau)] - (\hat{\gamma}_\tau - \hat{\beta})^2 [I - I(\tau)]/2\Big] + o_p(1),
\end{aligned}$$

Substitute $\hat{\beta}_\tau$, $\hat{\gamma}_\tau$, $\hat{\beta}$ and $U(\hat{\beta}, \tau)$ by $U(\tau) - (\hat{\beta} - \beta)I(\tau) + o(n^{1/2})$, and after some algebraic simplification(using Mathematica), we get the proof.

The asymptotic expressions for $\tilde{\Lambda}^c$ and $\tilde{\Lambda}$:

In addition to the notation used above, we denote W as the first derivative of $\log(\beta, \gamma, \tau, \theta)$ with respect to θ and $\bar{U}(\tau) = U - U(\tau)$ and $\bar{I}(\tau) = I - I(\tau)$. Under the null hypothesis, we have

$$\begin{aligned}
\hat{\beta} - \beta &= U/I_{zz} - I_{zw}(\hat{\theta} - \theta)/I_{zz} + o_p(n^{-1/2}), \\
\hat{\theta} - \theta &= (-I_{wz}U/I_{zz} + W)/F + o_p(n^{-1/2}), \\
\hat{\beta}_\tau - \beta &= U(\tau)/I_{zz}(\tau) - I_{zw}(\tau)(\hat{\theta}_\tau - \theta)/I_{zz}(\tau) + o_p(n^{-1/2}), \\
\hat{\gamma}_\tau - \gamma &= \bar{U}(\tau)/\bar{I}_{zz}(\tau) - \bar{I}_{zw}(\tau)(\hat{\theta}_\tau - \theta)/\bar{I}_{zz}(\tau) + o_p(n^{-1/2}), \\
\hat{\theta}_\tau - \theta &= (-I_{wz}(\tau)U(\tau)/I_{zz}(\tau) - \bar{I}_{wz}(\tau)\bar{U}(\tau)/\bar{I}_{zz}(\tau) + W)/F_\tau + o_p(n^{-1/2}),
\end{aligned}$$

where

$$
\begin{aligned}
F &= I_{ww} - I_{wz}I_{zw}/I_{zz} \\
F_\tau &= I_{ww} - I_{wz}(\tau)I_{zw}(\tau)/I_{zz}(\tau) - \bar{I}_{wz}(\tau)\bar{I}_{zw}(\tau)/\bar{I}_{zz}(\tau).
\end{aligned}
$$

Then, similar to the expansion of $\Lambda(\tau)$, we have

$$
\tilde{\Lambda}(\tau) = X^2(\tau)/Var(X(\tau)) + o_p(1), \tag{1}
$$

where

$$
\begin{aligned}
X(t) &= -[I_{ww}I_{zz} - I_{wz}^2]U(t) + [I_{ww}I_{zz}(t) - I_{wz}(t)I_{zw}]U \\
&\quad +[I_{zw}(t)I_{zz} - I_{zw}I_{zz}(t)]W, \\
Var(X(t)) &= [I_{ww}I_{zz}(t) - I_{zw}(t)I_{zw}]\bar{I}_{zz}(t) + [I_{zw}(t)I_{zz} - I_{zz}(t)I_{zw}]\bar{I}_{zw}(t).
\end{aligned}
$$

If $EI_{zw}(t)/EI_{zz}(t)$ is constant in t, then in (1), X(t) can be replaced by

$$
\tilde{X}(t) = I_{zz}F[-U(t) + UI_{zz}(t)/I_{zz}],
$$

and $Var(X(t))$ by

$$
Var(\tilde{X}(t)) = I_{zz}F^2I_{zz}(t)\bar{I}_{zz}(t),
$$

so the right side of (1) converges to square of a Ornstein-Uhlenbeck process.

For the proof of $\tilde{\Lambda}^c$, we observe theat if $EI_{zw}(t)/EI_{zz}(t)$ is constant in t, then $\hat{\theta}$ has the same expression as $\hat{\theta}_\tau$ up to $O(n^{-1/2})$, so $\tilde{\Lambda}^c$ is equivalent to $\tilde{\Lambda}$ in this case.

REFERENCES

ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-20.

BHAT RESEARCH GROUP (1982). A Randomized Trial of Propranolol in Patients with Acute Myocardial Infarction. *The Journal of American Medical Association,* **247**, 1707-14.

CHERNOFF, H. & ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Ann. Math. Statist.* **35**, 999-1028.

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

COX, D. R (1975). Partial likelihood. *Biometrika.* **62**, 269-76.

DEMETS, D. L., HARDY, H., FRIEDMAN, L. M. & LAN, K.K.G. (1984). Statistical Aspects of Early Termination in the Beta-Block Heart Attack Trial. *Controlled Clinical Trials.* **5**, 362-72.

DICKSON, E. R., GRAMBSCH, P. M., FLEMING, T. R., FISHER, L. D. & LANGWORTHY, A. (1989). Prognosis in primary biliary cirrhosis: model for decision making. *Hepatology* **10**, 1-7.

HARRINGTON, D. P. & FLEMING, T. R. (1991). *Counting processes and survival analysis.* New York : Wiley.

HOEL, D. G. (1972). A representation of mortality data by competing risk. *Biometrics* **28**, 475-88.

LAGAKOS, S. W. (1988). The loss in efficiency from misspecifying covariates in proportional hazard regression models. *Biometrika* **75**, 156-60.

LAGAKOS, S. W. & SCHOENFELD, D.A. (1984). Properties of proportional-hazards score tests under misspecified regression models. *Biometrics* **40**, 1037-48.

LIANG, K. E., SELF, S. G. & LIU, X. (1990). The Cox proportional hazards model with change point: An epidemiologic application. *Biometrics* **46**, 783-93.

LIN, D. Y. & WEI, L.J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **84**, 1074-8.

LIN, D. Y., WEI, L.J. & YING, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557-72.

LOADER, C. R. (1991). Inference for a hazard rate change point. *Biometrika* **78**, 749-57.

MATHEWS, D. E. & FAREWELL, V.T. (1982). On testing for a constant hazard rate against a change-point alternative. *Biometrics* **38**, 463-8.

MATHEWS, D. E., FAREWELL, V.T. & PYKE, R. (1985). Asymptotic score-statistics processes and tests for constant hazard against a change-point alternative. *Ann. Statist.* **13**, 583-91.

MILLER, R. G. (1981). *Survival analysis.* New York : Wiley.

MILLER, R. G. & HALPERN, J. (1982). Regression with Censored Data. *Biometrika* **69**, 521-31.

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167-79.

SCHOENFELD, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika* **69**, 239-41.

SIEGMUND, D. (1988). Confidence sets in change-point problems. *International Statistical Review* **56**, 31-48.

STRUTHERS, C. A. & KALBFLEISCH, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73**, 363-9.

THERNEAU, T. M., GRAMBSCH, P. M. & FLEMING, T. R. (1990). Martigale-based residuals for survival models. *Biometrika* **77**, 147-60.

WEI, L. J. (1984). Testing goodness of fit for proportional hazards model with censored observations. *J. Am. Statist. Assoc.* **79**, 649-52.

WORSLEY, K. J. (1988). Exact percentage points of the likelihood ratio test for a change-point hazard-rate model. *Biometrics* **44**, 259-63.

ZACKS, S. (1991). Handbook of Sequential Analysis. B.K. Ghosh and P.K. Sen(Ed). Nekker, New York