**CHECKING THE CENSORED TWO-SAMPLE ACCELERATED
LIFE MODEL USING INTEGRATED CUMULATIVE
HAZARD DIFFERENCE**

by

SEUNG-HWAN LEE, B.S., M.S.

A DISSERTATION

IN

MATHEMATICS

**Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of**

DOCTOR OF PHILOSOPHY

Approved

Co-Chairperson of the Committee

Co-Chairperson of the Committee

Accepted

Dean of the Graduate School

August, 2004

ACKNOWLEDGMENTS

I would like to thank my esteemed advisor, Dr. Song Yang for his support, encouragement, guidance, and especially his patience in the research for this dissertation. This dissertation would have been impossible without the expert guidance of Dr. Yang. Although he was not at Texas Tech University, Dr. Yang always pushed me hard and encouraged me through email, fax, and telephone. I would also like to thank Dr. Ruymgaart, Dr. Mansouri, and Dr. Paige for serving on my committee. In addition, many thanks go to the staff of the Department of Mathematics and Statistics for assisting me in various ways.

I would like to give my deepest thanks to my mom and dad for their encouragement and unconditional love. They have been a constant source of strength and support throughout my graduate studies at Texas Tech University. A special thanks also goes to my brother and sister-in-law. They encouraged and inspired me by telephone every week from Korea.

Finally, I give unmeasurable amounts of thanks to Eun-Joo Lee, my wife, who has experienced everything along with me.

CONTENTS

ACKNOWLEDGMENTS	ii
ABSTRACT	iv
LIST OF TABLES	v
LIST OF FIGURES	vi
I. INTRODUCTION	1
II. BACKGROUND AND PRELIMINARIES	4
2.1 Notation and the Model Specification	4
2.2 Counting Process and Martingale	6
III. YANG'S METHODS	11
3.1 Scale Change Estimators	11
3.2 Yang's Lack-of-Fit Test	17
IV. SOME STOCHASTIC PROCESSES AND LACK-OF-FIT TEST . .	21
4.1 Some Theoretical Results	21
4.2 Approximation Procedures of Some Stochastic Processes	24
4.3 Lack-of-Fit Test	32
V. SIMULATION STUDIES	37
VI. APPLICATIONS	42
VII. CONCLUSION	51
BIBLIOGRAPHY	53
APPENDIX	55

ABSTRACT

In this dissertation, some lack-of-fit tests will be discussed for the censored two-sample accelerated life model. Conventional scale estimators with two-sample censored data such as rank-based estimators and minimum distance estimators have difficulties to apply easily due to the fact that their asymptotic variances involve the unknown density, or they require some strict conditions. The object of this work is to provide an asymptotically equivalent martingale-based stochastic process of some estimating functions, which is easier to apply than existing methods from the literature.

An extreme value of the observed process compared with simulated realizations of the approximation process would indicate the model misspecifications. The approximation process involving the martingale structure can be achieved through some approximation procedures of the observed process under the two-sample scale model. The p -value applied to the approximation of the observed process leads to the construction of the lack-of-fit tests. Comparison of the processes enables one to get some information visually from the graph about how the model is misspecified.

LIST OF TABLES

3.1	Yang's point estimates and confidence intervals for the vaginal cancer data	17
5.1	Performance of the proposed lack-of-fit test under Cases 1-3 of censoring proportion 25%, 40% and 50% for sample size $m = 25, 50$ with several weights, and $\alpha=0.05$	40
5.2	Performance of Yang's lack-of-fit test under Cases 1-3 of censoring proportion 25%, 40% and 50% for sample size $m = 25, 50$ with several weights, and $\alpha=0.05$	41
6.1	Times to death from vaginal cancer or censorship following treatment of rats with the carcinogen 7,12-DMBA	43
6.2	Times for treatment to disease progression of patients with ovarian cancer	45

LIST OF FIGURES

2.1	Sample path of counting process	8
5.1	Comparison of Yang and the proposed tests on test rejection % of Case 3 with the pair $(\bar{F}^{2+\omega}(t)F^2(t)/t, \bar{F}^{2+\omega}(t)F(t)/t)$ for Yang, $\bar{F}^{2+\omega}(t)F(t)/t$ for New 1 and $\bar{F}^{2+\omega}(t)F^2(t)/t$ for New 2, sample size $m = 25, 50$ and $\alpha = 0.05$	39
6.1	Estimated cumulative hazard functions $\hat{\Lambda}_1$ (solid) and $\hat{\Lambda}_2$ (dotted) for the two groups in the vaginal cancer data	46
6.2	Estimated cumulative hazard functions $\hat{\Lambda}_1(t)$ (solid) and $\hat{\Lambda}_2(1.12t)$ (dotted) for the vaginal cancer data	47
6.3	Plot of the observed process(solid) and 50 approximation processes(dotted) for the vaginal cancer data	48
6.4	Estimated cumulative hazard functions $\hat{\Lambda}_1$ (solid) and $\hat{\Lambda}_2$ (dotted) for the two groups in the ovarian cancer data	49
6.5	Plot of the observed process(solid) and 50 approximation processes(dotted) for the ovarian cancer data	50

CHAPTER I

INTRODUCTION

Survival analysis is concerned with statistical analysis of time to event data. Data on a response variable, which is time until some event happens, are called lifetimes. The event may be the death of a patient in medical studies. Thus, the data recorded in the study are actual lifetimes. When the outcome of a study is the time to event, a number of problems can occur. A common feature of time-to-event data is censoring. Censored data occur when the time until the event happens cannot be fully observed. Censoring arises in various ways: right censoring, left censoring, double censoring, etc. Right censoring is very common in lifetime data. Right censoring occurs when the actual lifetime is longer than the observed lifetime. For instance, we cannot afford to wait until events have happened to all subjects, for example, until all are dead. Some patients might have left the study early: some patients may still be alive when the study is complete, may die from other causes during the study, or may be lost to follow up. Thus, the only information we have about some patients is that they were still alive at the last follow-up. Thus, the observed lifetime is less than the actual lifetime. These are called censored observations.

Comparison of two samples of possibly censored data is an important problem in survival analysis. Its application includes assessment of treatment effect when patients are randomized into two groups, where one group receives a new treatment and the other group receives a standard treatment. The proportional hazard model (Cox,

1972) has been the most used model in such two-group comparison. The validity of the proportional hazards model can be tested by several different approaches. Gill and Schumacher (1987) proposed a simple test of the proportional hazards assumption based on the generalized rank estimators of the relative risk. The Gill and Schumacher test does not require partition of the time axis that many other methods do.

When the assumption of proportional hazards is in doubt, one alternative to the Cox model is the two-sample accelerated life model. This model has a simple form and good interpretation of the parameter. It states that the lifetime is accelerated or decelerated by a scale factor. Extending the idea of Gill and Schumacher (1987), Yang (1998) introduced some lack-of-fit tests for the accelerated life model using certain weighted integrated empirical hazard difference of the two groups. The weights are based on Kaplan-Meier estimates of the cumulative distribution function of the pooled data. The weights are chosen such that the variance of the estimators can be easily and reliably estimated. Thus, confidence intervals of the parameter and lack-of-fit tests can be obtained with little extra computations. Although no asymptotic results are established there, it can be shown, similar to Yang (1997) for the general regression setting, that those tests are asymptotically consistent against certain monotonicity departures from the two-sample accelerated life model.

In this work, we introduce some new model checking procedures for the two-sample accelerated life model. These procedures are based on certain stochastic processes that make use of the estimating functions proposed in Yang (1998). The stochastic processes can be represented in terms of the martingale residuals, and their conditional

distributions given the data can be approximated by zero-mean Gaussian processes similar to Lin, Wei and Ying (1993) who investigated model checking issues regarding the Cox regression model. The p -value used for the proposed method can be estimated empirically through simulations. The observed process can be compared with simulated realizations of the approximation process for a visual inspection of the model adequacy. Based on the conditional distribution of the observed process, we derive an omnibus test that is asymptotically consistent against a general departure from the two-sample accelerated life model.

We organize this dissertation as follows. The notation and the model, and some martingale theory related to counting processes will be discussed in Chapter II. Chapter III will describe the scale change estimator of Yang (1998) along with its asymptotic properties and related estimating function. Also, some lack-of-fit test he proposed will be discussed. In Chapter IV, based on the class of estimating functions in Yang (1998), certain stochastic processes will be introduced. These processes are the foundation of the test procedures proposed in this research. Some martingale-based approximation procedures and construction of the lack-of-fit tests will be provided also in this chapter. Extensive numerical simulation results for several weight functions will be presented in Chapter V. Chapter VI will give applications to two data sets, one on time to death of vaginal cancer for rats and another on time to progression for patients with ovarian carcinoma. Chapter VII will conclude this dissertation.

CHAPTER II

BACKGROUND AND PRELIMINARIES

In Section 2.1, we will define some notation and functions commonly used in distributions for censored data, and describe the statistical model we will discuss. Accelerated life time model, especially, will be considered in terms of cumulative hazard function for the two-sample case. For the stochastic integral for the proposed method to be specified in Chapter IV, counting process and some martingale theory will be introduced in Section 2.2. Martingale methods provide very useful ways of studying the distributions in survival analysis. We will also see how martingale can be obtained for a counting process. This will play a central role in the development of approximation procedures that will be discussed in Chapter IV.

2.1 Notation and the Model Specification

Two-sample scale model will be considered in this research. Let $X_{ij}, i = 1, 2, j = 1, \dots, n_i$, be independent, positive failure times with an absolutely continuous distribution F_i , and C_{ij} be the corresponding censoring variables with a distribution function G_i . We assume that the variables C_{ij} are also independent of the variables X_{ij} . Let the censoring indicator Δ_{ij} be 1 when $X_{ij} \leq C_{ij}$ and 0 otherwise, i.e., the random variables X_{ij} are censored by the random variables C_{ij} . We observe $T_{ij} = X_{ij} \wedge C_{ij}$ and Δ_{ij} , where $\wedge$ denotes the minimum, $i = 1, 2, j = 1, \dots, n_i$. Also, let $\Lambda_i(t) = \int_0^t \lambda_i(u) du$ be the cumulative hazard function for X_{ij} , where $\lambda_i(t)$ is the

hazard function defined as

$$\begin{aligned}\lambda_i(t) &= \lim_{\delta \rightarrow 0} \frac{1}{\delta} P(X_{ij} \in (t, t + \delta) | X_{ij} > t) \\ &= \frac{f_i(t)}{1 - F_i(t)}.\end{aligned}$$

Note that $F_i(t)$ is the distribution function corresponding to $\Lambda_i(t)$ with the density function $f_i(t) = \frac{dF_i(t)}{dt}$ and the survival function $S_i(t) = 1 - F_i(t)$. Survival function describes the probability of surviving at least until time t .

Define

$$N_{ij}(t) = I(T_{ij} \leq t, \Delta_{ij} = 1) = \Delta_{ij} I(T_{ij} \leq t),$$

and

$$Y_{ij}(t) = I(T_{ij} \geq t),$$

where $I(\cdot)$ is the indicator function for the event specified in the parenthesis. Note that the sample paths of N_{ij} are step functions with jumps of size +1 and with $N_{ij}(0) = 0$. Define

$$N_i(t) = \sum N_{ij}(t),$$

$$Y_i(t) = \sum Y_{ij}(t),$$

where the sum is over $j = 1, \dots, n_i$. Then $N_i(t)$ can be interpreted as the number of deaths or failures in group i before or at t , and $Y_i(t)$ indicates the number of at risk in group i just before time t , $i = 1, 2$.

For the survival function $S_i(t) = 1 - F_i(t)$, it is commonly estimated by

$$\hat{S}_i(t) = \prod_{s \leq t} \left\{ 1 - \frac{\Delta N_i(s)}{Y_i(s)} \right\},$$

where $\Delta N_i(s) = N_i(s) - N_i(s-)$, $i = 1, 2$. This is called the product-limit estimator (Kaplan and Meier, 1958) which plays a central role in distributions for censored data. The distribution function $F_i(t)$ can then be estimated with $\hat{F}_i(t) = 1 - \hat{S}_i(t)$. $\hat{F}_i(t)$ will be used for some data-dependent weight function to be specified in Chapter III.

We will investigate the two-sample accelerated life model for the random variable X_{ij} , $i = 1, 2, j = 1, \dots, n_i$, that are i.i.d. with F_i and Λ_i . The model postulates that, for some $\theta \in (0, \infty)$, the random variable X_{2j}/θ , $j = 1, \dots, n_2$, have the same distribution as the random variable X_{1j} , $j = 1, \dots, n_1$. Thus θ is the scale factor for acceleration or deceleration depending on whether $\theta > 1$. For an alternative formulation, this model is equivalent to assuming that Λ_1 and Λ_2 are related by

$$\Lambda_2(t) = \Lambda_1\left(\frac{t}{\theta}\right). \quad (2.1)$$

Notice that, in (2.1), $\theta > 1$ implies that X_2 is stochastically larger than X_1 . Using the hazard-based model which is often convenient in the analysis of censored data, we investigate that any individual survival time t in the second group given by a new treatment would have survived time t/θ for some positive θ in the first group given by a standard treatment. Our goal in this research is to test the validity of the two-sample scale model (2.1) based on the available data $\{T_{ij}, \Delta_{ij}\}$.

2.2 Counting Process and Martingale

Counting process and martingale provide useful tools in studying estimators and tests in survival analysis. A counting process is a stochastic point process describing the cumulative number of events up to time t . We can then derive martingale structure

associated with the counting process. For the simplicity of illustration, we consider one-sample case.

Definition 2.2.1. (Fleming and Harrington, 1991) A counting process is a stochastic process $\{N(t) : t \geq 0\}$ adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$ with

$$N(0) = 0, N(t) \leq \infty \text{ a.s.},$$

and whose sample paths are nondecreasing, right-continuous, and have only jump discontinuities, with jumps of size +1.

It must satisfy

1. $N(t) \geq 0$.
2. $N(t)$ is integer valued.
3. If $s < t$, then $N(s) \leq N(t)$.
4. For $s < t$, $N(t) - N(s)$ equals the number of events occurred in $(s, t]$.

The realizations of N , seen as function of t , are integer-valued step functions, zero at time zero, with jumps of size +1 only. The sample paths of a counting process N are shown in Figure 2.1 using the data set from the literature (D. Collett, 1994, IUD, Chapter 1).

Let X and C be survival times and censoring times that are independent of each

Figure 2.1: Sample path of counting process

other. Let

$$T = \min\{X, C\},$$

$$\Delta = I(X \leq C),$$

$$N(t) = I(T \leq t, \Delta = 1),$$

$$Y(t) = I(T \geq t).$$

Then $N(t)$ is a counting process, and the censored data (T, Δ) can be represented by $(N(t), Y(t), t \geq 0)$. Much of the modern survival analysis is based on such representation. Define

$$A(t) = \int_0^t Y(s) d\Lambda(s). \quad (2.2)$$

Under independent censorship, the process given by

$$M(t) = N(t) - A(t) \tag{2.3}$$

is a martingale. Note that the probability that $N(t)$ makes a jump in $[t, t + dt)$ is $Y(t)d\Lambda(t)$. It follows that the process A given in (2.2) approximates the number of jumps by N over $(0, t]$. Note that the process $M = N - A$ has mean zero, since

$$\begin{aligned} E\{N(t)\} &= P\{T \leq t, T \leq C\} \\ &= E \int_0^t Y(s)d\Lambda(s) \\ &= E\{A(t)\}. \end{aligned}$$

Denote the complete history (everything that has happened) until just before time t (not including t) by $\mathcal{F}_{t-}$. It is written as

$$\mathcal{F}_{t-} = \sigma\{N(s), I(T \leq s, \Delta = 0) : 0 \leq s < t\}.$$

Similarly, $\mathcal{F}_t = \sigma\{N(s), I(T \leq s, \Delta = 0) : 0 \leq s \leq t\}$.

Proposition 2.2.2. *(Fleming, T.R. and Harrington, D.P., 1991) Under some conditions and notation above, the process M is a martingale with respect to a filtration $\{\mathcal{F}_t\}$.*

Moreover, the sum of martingales is also a martingale. Specifically, suppose that we have several counting process $N_i(t)$, $i = 1, \dots, n$, and determine $M_i(t) = N_i(t) - A_i(t)$. Then

$$\sum_{i=1}^n c_i M_i(t)$$

is also a martingale for constants c_i .

In summary, a martingale is a special stochastic process whose increment over an interval $(t, t + s]$, given the past up to and including time t , has expectation zero, i.e.,

$$E\{M_{t+s} - M_t | \mathcal{F}_t\} = 0$$

for all t , $0 < s < \infty$. Note that given $\mathcal{F}_t$, M_t is fixed.

We need one more concept, that of the predictable variation process of a martingale M . As $Y(t)d\Lambda(t)$ is estimated by $dN(t)$, the estimator's variance can be estimated by the conditional variance. It is called predictable variation process. That is a process $\langle M \rangle = \{\langle M \rangle(t) : t \geq 0\}$ defined by

$$d\langle M \rangle(t) = E\{(dM(t))^2 | \mathcal{F}_{t-}\} = \text{Var}\{dM(t) | \mathcal{F}_{t-}\}.$$

And for a martingale, $M(t) = N(t) - A(t)$,

$$\langle M \rangle(t) = \int_0^t Y(s)d\Lambda(s) = A(t)$$

when $A(t)$ is continuous. Therefore, by Corollary 2.3.2 and Lemma 2.3.2 of Fleming and Harrington (1991),

$$E\{M^2(t)\} = E\{A(t)\}, \quad t \geq 0$$

if $E\{A(t)\} < \infty$ (or equivalently if $E\{N(t)\} < \infty$) for any t . So we have

$$\text{Var}\{M(t)\} = E\{N(t)\}$$

which is central in some approximation procedure in Chapter IV.

CHAPTER III

YANG'S METHODS

Some scale estimators for the censored two-sample accelerated life model were introduced by Yang (1998). They are zeros of some integrated weighted difference between the two cumulative hazard estimators. These estimating functions produce estimators that are asymptotically normal with a density free asymptotic variance that can be easily estimated. In relation to Gill and Schumacher (1987), and Wei and Gail (1983), some lack-of-fit tests for the accelerated life model are also derived from the scale estimators.

3.1 Scale Change Estimators

Recall that $N_i(t)$ denotes the number of deaths or failures in group i before or at t , and $Y_i(t)$ the number at risk in group i at t — for $i = 1, 2$. Then Nelson estimators (Nelson, 1969) of the cumulative hazard functions Λ_i are given by

$$\hat{\Lambda}_i(t) = \int_0^{t \wedge T_{i,max}} \frac{dN_i(s)}{Y_i(s)} \quad i = 1, 2,$$

where $T_{i,max} = \max_j T_{ij}$. Let $n = n_1 + n_2$. Since the accelerated life model assumes $\Lambda_2(t) = \Lambda_1\left(\frac{t}{\theta}\right)$ for some θ , both $\hat{\Lambda}_1(t)$ and $\hat{\Lambda}_2(\theta t)$ estimate $\Lambda_1(t)$ under the assumed model. Therefore, an estimator $\hat{\theta}$ can be defined as a zeros of

$$n^{1/2} \int_0^\infty B_n(t; c) d(\hat{\Lambda}_1(t) - \hat{\Lambda}_2(ct)), \tag{3.1}$$

where $B_n(t; c) = \int_t^\infty W_n(s; c) ds$, and $W_n(s; c)$ is a data dependent weight function to be specified later. To avoid technical difficulties in analyzing the asymptotic behavior of $\hat{\theta}$, the integration range is truncated. Suppose

$$u > 0, v < \tau_1 \wedge (\tau_2/c) \quad (3.2)$$

for c in a neighborhood $N(\theta)$, where $\tau_i = \sup\{t; F_i(t) < 1\}$. Then we consider (3.1) with truncated integration range $[u, v]$. The assumption $v < \tau_1 \wedge (\tau_2/c)$, in particular, is used in order to avoid the explosive behavior of $\hat{\Lambda}_1(t)$ and $\hat{\Lambda}_2(ct)$ at the upper tail. Then the estimating function can be written as

$$D_n(c) = n^{1/2} \int_u^v W_n(t; c) \{\hat{\Lambda}_1(t) - \hat{\Lambda}(ct)\} dt \quad (3.3)$$

Note that when $u = 0, v = \infty$, $D_n(c)$ in (3.3) can be rewritten in the form of (3.1) from integration of parts. In practice, the estimator can be chosen as the minimizer of $\{D_n(c)\}^2$. Since the estimating function $D_n(c)$ is not smooth, the ordinary Newton method can not be used for finding $\hat{\theta}$. Moreover, $D_n(c)$ may not be a monotone function due to the presence of right censoring. To overcome these problems, we can use the optimization algorithms such as grid search or random search for general non-smooth functions to locate the true θ . Yang (1998) shows that, under regularity conditions, $\hat{\theta}$ is consistent, and the order of asymptotic convergence is the usual $n^{1/2}$. These results follow from the next proposition. Note that the estimating function in (3.3) is well-defined when the weight $W_n(t; c)$ is bounded and vanished outside of $[0, T_{1,max} \wedge (T_{2,max}/c)]$.

Proposition 3.1.1. (Yang, 1998) Let $\Gamma_{ni} = n_i^{-1} \sum \bar{G}_{ij}$, where G_{ij} is the distribution function of the censoring variable C_{ij} and $\bar{D} = 1 - D$ for a distribution D . Suppose that the following conditions hold:

1. As $n_1, n_2 \rightarrow \infty$, $\Gamma_{n1}(t), \Gamma_{n2}(ct)$, and $W_n(t; c)$ converge uniformly to $\Gamma_1(t), \Gamma_2(ct)$, and $W(t; c)$, respectively for $t \in [u, v]$ and $c \in N(\theta)$.
2. $\Gamma_1(u), \Gamma_2(v) > 0$ for $c \in N(\theta)$.

Then, with probability 1 and uniformly for $c \in N(\theta)$, $n^{-1/2} D_n(c)$ converges to

$$L(c) = \int_u^v W(t; c) \{ \Lambda_1(t) - \Lambda_2(ct) \} dt.$$

If, in addition, $W(t; c) > 0$, then $L(c)$ has a unique zero $c = \theta$.

To see the asymptotic behavior of $\hat{\theta}$, we assume that

$$\frac{n_i}{n} \rightarrow \psi_i \quad i = 1, 2,$$

where $n = n_1 + n_2$. Suppose $W_n(t; c_n)$ converges in probability to $W(t; \theta)$ as $c_n \rightarrow \theta$ uniformly for $t \in [u, v]$. Rewrite $\Lambda_1(t)$ as $\Lambda(t)$, and let $\lambda(t) = \frac{d\Lambda(t)}{dt}$. Proposition 3.1.1 shows that by large sample approximation and under some regularity conditions, $\hat{\theta}$ is a consistent estimator of θ , and $D_n(c)$ is asymptotically linear near θ such that

$$D_n(c_n) = n^{1/2} \int_u^v W_n(t; \theta) \{ \hat{\Lambda}_1(t) - \hat{\Lambda}_2(\theta t) - \left(\frac{c_n}{\theta} - 1 \right) t \lambda(t) \} dt + o_p(1) \quad (3.4)$$

for c_n such that $n^{1/2} \left(\frac{c_n}{\theta} - 1 \right) = O_p(1)$. Then, by the standard martingale central limit theorem, $n^{1/2} \left(\frac{\hat{\theta}}{\theta} - 1 \right)$ is asymptotically normal with mean zero and variance

$\sigma^2 = \mu^{-2} \rho^2$, where

$$\begin{aligned}\mu &= \int_u^v W(t; \theta) t \lambda(t) dt, \\ \rho^2 &= \lim_n \int_u^v \int_u^v W(s; \theta) W(t; \theta) \int_0^{s \wedge t} \left\{ \frac{1}{\psi_1 \Gamma_{n_1}(x)} + \frac{1}{\psi_2 \Gamma_{n_2}(\theta x)} \right\} \frac{d\Lambda(x)}{\bar{F}(x)} ds dt \quad (3.5)\end{aligned}$$

with $F(t)$ being the distribution function corresponding to $\Lambda(t)$, and $\bar{F} = 1 - F$.

Let $f(t) = \frac{dF(t)}{dt}$ be the density of $F(t)$. Then $\lambda(t) = \frac{f(t)}{1-F(t)}$. Looking at expression (3.5), we can see that for a general weight function W , the asymptotic slope, μ , of the asymptotic variance usually depends on the unknown density function $f(t)$ through the hazard function $\lambda(t)$. The density function and hazard function are not simply estimated for censored data. However, observe that if the weight $W(t; \theta)$ is chosen to be $\zeta\{F(t)\}/t$ for some function ζ , then

$$\mu = \int_u^v \zeta(F) \bar{F}^{-1} dF.$$

Hence, the corresponding asymptotic variance does not involve the density function.

Thus, the asymptotic variance can be easily estimated.

In particular, Yang looks at weight functions of form,

$$W(t; \theta) = \{1 - F(t)\}^\epsilon \{F(t)\}^\delta / t$$

for some $\epsilon, \delta > 0$. The weights have a factor of $\frac{1}{t}$ due to an invariant argument.

Then the slope becomes, $\mu = \int_u^v (1 - F)^{\epsilon-1} F^\delta dF$. So, by taking $W_n(t; c) = (1 - \hat{F}_n(t; c))^\epsilon \hat{F}_n^\delta(t; c)/t$, where $\hat{F}(t; c)$ is the product-limit estimator (Kaplan and Meier, 1958) based on Δ_{ij} and $T_{ij}, T_{2j}/c$ through the survival function from the combined

sample, σ^2 can be consistently estimated by $\hat{\sigma}^2 = \mu_n^{-2} \rho_n^2$, where

$$\mu_n = \int_u^v \{1 - \hat{F}(t; \hat{\theta})\}^{\epsilon-1} \hat{F}^\delta(t; \hat{\theta}) \hat{F}(dt; \hat{\theta}),$$

$$\rho_n^2 = 2 \int_u^v \int_u^t W_n(t; \hat{\theta}) W_n(s; \hat{\theta}) \int_0^s \left\{ \frac{n}{Y_1(x)} + \frac{n}{Y_2(\hat{\theta}x)} \right\} d\hat{\Lambda}(x; \hat{\theta}) ds dt,$$

where $\hat{\Lambda}(t; c)$ the cumulative hazard function corresponding to $\hat{F}(t; c)$.

With the consistent estimator $\hat{\sigma}^2 = \mu_n^{-2} \rho_n^2$ and the asymptotic normality of $n^{1/2} \left(\frac{\hat{\theta}}{\theta} - 1 \right)$, asymptotic confidence interval for θ can be easily obtained. One possible $(1 - \alpha)100\%$ confidence interval of θ is

$$\left[\frac{\hat{\theta}}{1+w}, \frac{\hat{\theta}}{1-w} \right],$$

where $w = n^{-1/2} \hat{\sigma} Z_{\alpha/2}$, with $Z_{\alpha/2}$ the $(1 - \alpha/2)$ -percentile of the standard normal distribution. However, such confidence intervals may contain negative values. A logarithm transformation can be used as a remedy. Since $n^{1/2}(\log(\hat{\theta}) - \log(\theta))$ can be shown to be asymptotically equivalent to $n^{1/2} \left(\frac{\hat{\theta}}{\theta} - 1 \right)$, $n^{1/2}(\log(\hat{\theta}) - \log(\theta))$ is asymptotically normal with mean zero and variance that can be consistently estimated by $\hat{\sigma}^2$. Therefore, another possible $(1 - \alpha)100\%$ confidence interval of θ is

$$[\hat{\theta} \exp(-w), \hat{\theta} \exp(w)],$$

with w as before. This interval lies in the positive half axis as long as $\hat{\theta}$ is positive.

It can be shown by Taylor expansion on $\exp(w)$, etc., that this interval is always shorter than the interval $\left[\frac{\hat{\theta}}{1+w}, \frac{\hat{\theta}}{1-w} \right]$. The asymptotic normality of $n^{1/2}(\log(\hat{\theta}) - \log(\theta))$ provides the basis for an asymptotic $(1 - \alpha)100\%$ tests on the scale parameter.

The test would be of the form, $H_0 : \theta = 1$ (no difference between the two groups)

versus $H_1 : \theta > 1$ (longer life for group two). In this test, H_0 is rejected if $\hat{\theta} \geq \exp(n^{-1/2}\hat{\sigma}Z_\alpha)$.

The key to the easy calculations of the confidence intervals and hypothesis test mentioned above is the choice of the weights. The weights of the form $W(t) = \{1 - F(t)\}^\epsilon \{F(t)\}^\delta / t$ eliminate the problem of an unknown density function in the expression of the asymptotic variance. Yang looks at many different values of ϵ and δ to find the weight function which gives the best results for the estimate of the scale change. He concludes that the two best weight functions are $\bar{F}^{i+w}(t)F(t)^i/t, i = 1, 2$, where w is the with probability 1 limit of

$$\hat{w} = \frac{\hat{\pi}}{1 - \hat{\pi}}$$

with the sample censoring proportion $\hat{\pi} = 1 - \frac{1}{n} \sum_{ij} \Delta_{ij}$. Thus, the sample version of the weight $\bar{F}^{i+w}(t)F(t)^i/t$ is $(1 - \hat{F}(t; c))^{i+\hat{w}}\hat{F}^i(t; c)/t, i = 1, 2$. We denote the corresponding estimates obtained from the two weight functions by $\hat{\theta}_i, i = 1, 2$.

$\hat{\theta}_1, \hat{\theta}_2$ and logrank estimator were compared by Yang. The confidence intervals for the logrank estimator were obtained from the methods of Wei and Gail (1983). Three cases were considered: (a) both the survival and the censoring distributions belong to the Log-logistic power-scale family, (b) the survival distributions are Weibull and the censoring distributions are Lognormal, and (c) the survival distributions are logarithms of uniform variables. In each case, $\theta = 2, n_1 = n_2 = m$, and the two samples have the same censoring pattern that does not depend on the scale parameter.

From simulations and other numerical studies, Yang (1998) concludes that infer-

Table 3.1: Yang's point estimates and confidence intervals for the vaginal cancer data

Data	Estimator	Point Estimate	95% Confidence Interval
Original Data	$\hat{\theta}_1$	1.12	(.99, 1.27)
	$\hat{\theta}_2$	1.12	(.99, 1.26)
	logrank	1.13	(.97, 1.30)
Original Data Minus 100	$\hat{\theta}_1$	1.21	(.98, 1.50)
	$\hat{\theta}_2$	1.21	(.98, 1.49)
	logrank	1.21	(.96, 1.54)

ence procedures are not only computationally simple and fast but also behave well. Table 3 shows that Yang's results for the estimators $\hat{\theta}_i$ with weights $\bar{F}^{i+\omega}(t)F^i(t)/t, i = 1, 2$, and logrank estimator, to the original data and to the original data minus 100.

3.2 Yang's Lack-of-Fit Test

Consider the test problem of $H_0 : \Lambda_2(t) = \Lambda_1(t/\theta)$ for some θ versus $H_1 : \Lambda_2(t) \neq \Lambda_1(t/\theta)$ for any θ . In Section 3.1, under H_0 , the scale estimators from different weight functions are asymptotically normal and close to each other. Looking at (3.4), we notice that an estimating function evaluated at another scale estimator has asymptotically mean zero. Under H_1 , this quantity may approach infinity as the sample sizes increases. This is the idea for Yang's test of H_0 versus H_1 .

Let $W_{ni}(t; c) = \{1 - \hat{F}(t; c)\}^{\epsilon_i} \hat{F}^{\delta_i}(t; c)/t, i = 1, 2$, with corresponding estimating functions $D_{ni}(c)$ and estimators $\hat{\theta}_i$. Then by (3.4), under H_0 , $D_{n_2}(\hat{\theta}_1)$ is asymptotically

equivalent to

$$\begin{aligned} & n^{1/2} \int_u^v W_2(t; \theta) \{ \hat{\Lambda}_1(t) - \hat{\Lambda}_2(\theta t) \} dt - n^{1/2} \left(\frac{\hat{\theta}_1}{\theta} - 1 \right) \int_u^v W_2(t; \theta) \lambda(t) dt \\ &= n^{1/2} \int_u^v \left\{ W_2(t; \theta) - \frac{\gamma_2}{\gamma_1} W_1(t; \theta) \right\} \{ \hat{\Lambda}_1(t) - \hat{\Lambda}_2(\theta t) \} dt + o_p(1), \end{aligned}$$

where

$$\begin{aligned} \gamma_i &= \int_u^v W_i(t; \theta) \lambda(t) t dt \\ &= \int_u^v (1 - F)^{\epsilon_i - 1} F^{\delta_i} dF. \end{aligned}$$

It follows that, under H_0 , $D_{n2}(\hat{\theta}_1)$ is asymptotically normal with mean zero and a variance that can be consistently estimated by

$$\begin{aligned} V_n^2 &= 2 \int_u^v \int_u^t \left\{ W_2(s; \hat{\theta}_1) - \frac{\gamma_{n2}}{\gamma_{n1}} W_1(s; \hat{\theta}_1) \right\} \left\{ W_2(t; \hat{\theta}_1) - \frac{\gamma_{n2}}{\gamma_{n1}} W_1(t; \hat{\theta}_1) \right\} \\ &\quad \times \int_0^s \left\{ \frac{n}{Y_1(x)} + \frac{n}{Y_2(\hat{\theta}_1 x)} \right\} d\hat{\Lambda}(x; \hat{\theta}_1) ds dt, \end{aligned}$$

where

$$\gamma_{nj} = \int_u^v \{ 1 - \hat{F}(t; \hat{\theta}_1) \}^{\epsilon_j - 1} \hat{F}^{\delta_j}(t; \hat{\theta}_1) \hat{F}'(dt; \hat{\theta}_1).$$

Thus $D_{n2}(\hat{\theta}_1)$ can be used as a test statistic to test the lack of fit of the assumed model. The expression of the estimated variance V_n^2 indicates that the corrected but asymptotically equivalent statistic $D_{n2}(\hat{\theta}_1) - \frac{\gamma_{n2}}{\gamma_{n1}} D_{n1}(\hat{\theta}_1)$ should have better finite sample behavior. This leads to an asymptotic level α test that rejects H_0 when

$$\left| D_{n2}(\hat{\theta}_1) - \frac{\gamma_{n2}}{\gamma_{n1}} D_{n1}(\hat{\theta}_1) \right| V_n^{-1} > Z_{\alpha/2}.$$

Note that this test is invariant when $\hat{\theta}_1$ and $\hat{\theta}_2$ are exchanged. This lack-of-fit test can be related to the idea of Gill and Schumacher (1987) for testing proportionality of hazard functions. Under H_0 , the scale estimators $\hat{\theta}_i$ should be close to each other. If $\epsilon_1 = \epsilon_2$ or $\delta_1 = \delta_2$ in the weights $W_{ni}(t; c)$, $i = 1, 2$, then the ratio of the two weight functions is monotone and one of two assigns more weight to the early deaths than the other. Under a monotone alternative H_1 , the two estimators should differ. Thus, similar to Gill and Schumacher (1987), a lack-of-fit test can be constructed by comparing the two estimators. The asymptotic representations of $n^{1/2}(\hat{\theta}_1/\hat{\theta} - 1)$ and $n^{1/2}(\hat{\theta}_2/\hat{\theta} - 1)$ provide that $n^{1/2}(\hat{\theta}_1/\hat{\theta}_2 - 1)$ is asymptotically equivalent to

$$n^{1/2} \int_u^v \left\{ \frac{W_2(t; \theta)}{\gamma_2} - \frac{W_1(t; \theta)}{\gamma_1} \right\} \{ \hat{\Lambda}_1(t) - \hat{\Lambda}_2(\theta t) \} dt.$$

This is an asymptotically equivalent form of the test statistic $D_{n2}(\hat{\theta}_1) - \frac{\gamma_{n2}}{\gamma_{n1}} D_{n1}(\hat{\theta}_1)$. However, $\hat{\theta}_2/\hat{\theta}_1$ requires finding $\hat{\theta}_1$ and $\hat{\theta}_2$ or using minimization twice. Since minimization usually dominates the computation time, it is computationally much more efficient to work with $D_{n2}(\hat{\theta}_1) - \frac{\gamma_{n2}}{\gamma_{n1}} D_{n1}(\hat{\theta}_1)$.

For the consistency of the proposed tests, we expect that, under regularity conditions and under H_1 , $n^{-1/2} D_{ni}(c)$ converges in probability to

$$L_i(c) = \int_u^v W_i(t; c) \{ \Lambda_1(t) - \Lambda_2(ct) \} dt.$$

We need this convergence to be uniform for c in a compact interval so that $\hat{\theta}_i$ converges in probability to one of the zeros of L_i . Suppose $\Lambda_2(t) = \Lambda_1(t/\theta(t))$ for a bounded nonconstant monotone function $\theta(t)$. By the monotonicity of Λ_1 , L_i should have at least one zero in the range of $\theta(t)$ and no zero outside. Let c_1 be a root of L_1 . If

$r(t; c) = W_2(t; c)/W_1(t; c)$ is monotone, then

$$L_1(c_1) = \int_u^v W_1(t; c_1) [\Lambda_1(t) - \Lambda_1\{c_1 t / \theta(t)\}] dt = 0,$$

$$L_2(c_1) = \int_u^v r(t; c_1) W_1(t; c_1) [\Lambda_1(t) - \Lambda_1\{c_1 t / \theta(t)\}] dt.$$

Thus, $L_2(c_1)$ cannot be zero except in degenerate case. If $\theta(t)$ and $r(t; c)$ are both monotone in t but in opposite directions, $L_2(c_1)$ is negative. Thus, under H_1 with monotone $\theta(t)$, although $\hat{\theta}_1$ may not be asymptotically unique, the test statistic $D_{n2}(\hat{\theta}_1) - \gamma_{n2}/\gamma_{n1} D_{n1}(\hat{\theta}_1)$ diverges to infinity for any $\hat{\theta}_1$. Since the variance estimator V_n^2 can be shown to converge in probability to a positive number, the standardized test $|D_{n2}(\hat{\theta}_1) - \gamma_{n2}/\gamma_{n1} D_{n1}(\hat{\theta}_1)| V_n^{-1}$ is consistent against a monotone alternative H_1 in the sense described above. Similarly, if $\theta(t)$ is convex or concave, the test is consistent if $r(t; c) = \theta(t)$.

CHAPTER IV

SOME STOCHASTIC PROCESSES AND LACK-OF-FIT TEST

Based on the class of estimating functions by Yang (1998), some new lack-of-fit tests for the accelerated life model can be derived using integrated cumulative hazard difference. In this chapter, we develop such a test based on martingale. The new test is asymptotically consistent against a general departure from the two-sample accelerated life model. The procedures are based on certain stochastic processes that can be represented in terms of martingale, and their conditional distributions given the data can be approximated by zero-mean Gaussian processes. Comparison of the processes enables one to get some information visually from the graph about how the model is misspecified. This graphical method is very useful during the stage of model building.

4.1 Some Theoretical Results

Motivated by the class of estimating functions of Yang (1998) described in (3.3), we consider the process in x ,

$$D_n(x, c) = n^{1/2} \int_u^x W_n(t; c) \{ \hat{\Lambda}_1(t) - \hat{\Lambda}_2(ct) \} dt \quad \text{for } u < x \leq v, \quad (4.1)$$

Note that when $x = v$, the process $D_n(x, c)$ in (4.1) becomes the estimating function described in (3.3) so that $\hat{\theta}$ is the zero of $D_n(v, c)$. Taylor expansion of $\Lambda_1(ct/\theta)$

around $c = \theta$ yields

$$\Lambda_1\left(\frac{ct}{\theta}\right) = \Lambda_1(t) + \left(\frac{c}{\theta} - 1\right)t\lambda_1(t) + R\left(\frac{ct}{\theta}\right), \quad (4.2)$$

where $\lambda_1 = \frac{d\Lambda_1}{dt}$ and $R\left(\frac{ct}{\theta}\right)$ is a Lagrange remainder, i.e., $\frac{\lambda'(\xi)}{2}(ct/\theta - t)^2$ for some point ξ between ct/θ and t . Let

$$W(t) = \{1 - F(t)\}^{\epsilon+\omega}\{F(t)\}^\delta/t,$$

$$W_n(t; c) = \{1 - \hat{F}(t; c)\}^\epsilon\{\hat{F}(t; c)\}^\delta/t,$$

where $\epsilon, \delta > 0$, and $\hat{F}(t; c)$ is the product-limit estimator (Kaplan and Meier, 1958) on Δ_{ij} and $T_{ij}, T_{ij}/c$, i.e., the pooled estimator from the combined sample. More generally, as in Yang (1998), we can use $W_n(t; c) = \xi(\hat{F}(t; c))$ for a suitable function ξ . By the uniform convergence in probability of W_n to W through the product-limit estimator to the survivor function, Yang (1998) has shown that

$$n^{1/2} \int_u^v W_n(t; \theta) \{\hat{\Lambda}_1(t) - \hat{\Lambda}_2(\theta t)\} = n^{1/2} \int_u^v W(t) \{\hat{\Lambda}_1(t) - \hat{\Lambda}_2(\theta t)\} + o_p(1). \quad (4.3)$$

Theorem 4.1.1. *In addition to the assumptions of Proposition 3.1.1, assume*

1. u, v satisfy (3.2).
2. F and G_i have Lebesgue densities f and g_i , respectively.
3. f'_0 is uniformly continuous and integrable, where $f_o(t) = \exp(t)f(\exp(t))$.
 $\sup_t |f'_0(t)| \leq C, \sup_t |g_{0i}(t)| \leq C, i = 1, \dots, n$, for some C , where $g_{0i}(t) = \exp(t)g_i(\exp(t))$.

4. $\int_{-\infty}^{\infty} W(t; a_n) dt = o_p(1)$ as $a_n \rightarrow 0$, where

$$W(t; a_n) = \sup_{|s| \leq a_n} (|f'_0(t+s) - f'_0(t)| + |f_0(t+s) - f_0(t)|).$$

5. $\sup_i E|\ln(Y_i) \wedge \ln(C_i)|^r < \infty$ for some $r > 0$.

Then

$$D_n(x, \hat{\theta}) = \tilde{D}_n(x, \theta) + o_p(1), \quad (4.4)$$

where

$$\tilde{D}_n(x, \theta) = D_n(x, \theta) - \frac{\gamma_x}{\gamma_v} D_n(v, \theta),$$

and

$$\gamma_x = \int_u^x W_n(t; \theta) \lambda_1(t) dt.$$

Proof. From the Appendix of Yang (1998) and its correction (2000), under the above assumptions, we have

$$n^{1/2} \left(\frac{\hat{\theta}}{\theta} - 1 \right) = -n^{1/2} \frac{1}{\gamma_v} \int_u^v W_n(t; \theta) \{ \hat{\Lambda}_1(t) - \hat{\Lambda}_2(\theta t) \} dt + o_p(1),$$

and

$$D_n(v, \hat{\theta}) = D_n(v, \theta) + n^{1/2} \left(\frac{\hat{\theta}}{\theta} - 1 \right) \gamma_v + o_p(1), \quad (4.5)$$

where $\gamma_v = \int_u^v W_n(t; \theta) \lambda_1(t) dt$. Now, similar to Yang (1998), we can obtain, uniformly for $x \in (u, v)$,

$$D_n(x, \hat{\theta}) = D_n(x, \theta) + n^{1/2} \left(\frac{\hat{\theta}}{\theta} - 1 \right) \gamma_x + o_p(1), \quad (4.6)$$

where $\gamma_x = \int_u^x W_n(t; \theta) \lambda_1(t) dt$. Hence, from (4.5) and (4.6), it follows that

$$D_n(x, \hat{\theta}) = D_n(x, \theta) - \frac{\gamma_x}{\gamma_v} D_n(v, \theta) + o_p(1)$$

uniformly for $x \in (u, v)$. □

4.2 Approximation Procedures of Some Stochastic Processes

Consider $N_{1j}(t)$, $N_{2j}(t; c)$ and $Y_{1j}(t)$, $Y_{2j}(t; c)$ at $c = \theta$.

$$\begin{aligned} N_{1j}(t) &= I(T_{1j} \leq t, \Delta_{1j} = 1) \quad j = 1, \dots, n_1, \\ N_{2j}(t; \theta) &= I\left(\frac{T_{2j}}{\theta} \leq t, \Delta_{2j} = 1\right) \quad j = 1, \dots, n_2, \end{aligned}$$

and

$$\begin{aligned} Y_{1j}(t) &= I(T_{1j} \geq t) \quad j = 1, \dots, n_1, \\ Y_{2j}(t; \theta) &= I\left(\frac{T_{2j}}{\theta} \geq t\right) \quad j = 1, \dots, n_2, \end{aligned}$$

where $T_{2j}/\theta = (X_{2j}/\theta) \wedge (C_{2j}/\theta)$ for $j = 1, \dots, n_2$. From Proposition 2.2.2, the processes given by

$$\begin{aligned} M_{1j}(t) &= I(T_{1j} \leq t, \Delta_{1j} = 1) - \int_0^t I(T_{1j} \geq s) d\Lambda_1(s) \\ &= N_{1j}(t) - \int_0^t Y_{1j}(s) d\Lambda_1(s) \end{aligned} \tag{4.7}$$

for $j = 1, \dots, n_1$, and

$$\begin{aligned} M_{2j}(t; \theta) &= I\left(\frac{T_{2j}}{\theta} \leq t, \Delta_{2j} = 1\right) - \int_0^t I\left(\frac{T_{2j}}{\theta} \geq s\right) d\Lambda_1(s) \\ &= N_{2j}(t; \theta) - \int_0^t Y_{2j}(s; \theta) d\Lambda_2(\theta s) \end{aligned} \tag{4.8}$$

for $j = 1, \dots, n_2$, are zero-mean martingales with respect to the filtrations

$$\mathcal{F}_{1t} = \sigma\{N_{1j}(s), I(T_{1j} \leq s, \Delta_j = 0) : 0 \leq s \leq t, j = 1, \dots, n_1\},$$

and

$$\mathcal{F}_{2t} = \sigma\{N_{2j}(s), I\left(\frac{T_{2j}}{\theta} \leq s, \Delta_j = 0\right) : 0 \leq s \leq t, j = 1, \dots, n_2\}.$$

Note that given the past up to (but not including) time t , Y_{ij} and Λ_i are fixed or nonrandom. We say in such a case that Y_{ij} and Λ_i are predictable. Martingale residuals can then be obtained by replacing the unknown quantities in (4.7) and (4.8) with consistent estimators. We now define the martingale residuals,

$$\begin{aligned}\hat{M}_{1j}(t) &= N_{1j}(t) - \int_0^t Y_{1j}(s) d\hat{\Lambda}_1(s) \quad j = 1, \dots, n_1 \\ \hat{M}_{2j}(t; \hat{\theta}) &= N_{2j}(t; \hat{\theta}) - \int_0^t Y_{2j}(s; \hat{\theta}) d\hat{\Lambda}_2(s\hat{\theta}) \quad j = 1, \dots, n_2,\end{aligned}$$

where

$$\begin{aligned}\hat{\Lambda}_1(t) &= \int_0^t \frac{d(\sum_{j=1}^{n_1} \Delta_{1j} I(T_j \leq s))}{\sum_{j=1}^{n_1} I(T_{1j} \geq s)}, \\ \hat{\Lambda}_2(ct) &= \int_0^t \frac{d(\sum_{j=1}^{n_2} \Delta_{2j} I(T_{2j}/c \leq s))}{\sum_{j=1}^{n_2} I(T_{2j}/c \geq s)}.\end{aligned}$$

The martingale residuals can be interpreted as the difference at time t between the observed and expected numbers of events for the j^{th} subject. Martingale methods provide direct and useful ways of studying estimators and tests for the censored life time data. In this research, we will see how the statistics can be obtained as stochastic integrals with respect to counting processes and martingales.

We now examine the process $D_n(x, c)$ at $c = \theta$ under the two-sample scale model.

Consider the process $D_n(x, \theta)$ from (4.1).

$$\begin{aligned} D_n(x, \theta) &= n^{1/2} \int_u^x W_n(t; \theta) \{ \hat{\Lambda}_1(t) - \hat{\Lambda}_2(\theta t) \} dt \\ &= n^{1/2} \int_u^x W_n(t; \theta) \{ \hat{\Lambda}_1(t) - \Lambda_1(t) - \hat{\Lambda}_2(\theta t) + \Lambda_2(\theta t) + \Lambda_1(t) - \Lambda_2(\theta t) \} dt. \end{aligned}$$

If the two-sample scale model is true, i.e., $\Lambda_1(t) = \Lambda_2(\theta t)$, we have

$$D_n(x, \theta) = n^{1/2} \int_u^x W_n(t; \theta) \{ \hat{\Lambda}_1(t) - \Lambda_1(t) - (\hat{\Lambda}_2(\theta t) - \Lambda_2(\theta t)) \}, \quad (4.9)$$

where

$$\begin{aligned} \hat{\Lambda}_1(t) - \Lambda_1(t) &= \int_0^t \frac{d(\sum_{j=1}^{n_1} \Delta_{1j} I(T_{1j} \leq s))}{\sum_{j=1}^{n_1} I(T_{1j} \geq s)} - \Lambda_1(t) \\ &= \int_0^t \frac{d(\sum_{j=1}^{n_1} I(T_{1j} \leq s, \Delta_{1j} = 1))}{\sum_{j=1}^{n_1} I(T_{1j} \geq s)} - \Lambda_1(t) \\ &= \int_0^t \frac{d(\sum_{j=1}^{n_1} N_{1j}(s))}{\sum_{j=1}^{n_1} Y_{1j}(s)} - \Lambda_1(t), \end{aligned}$$

and

$$\begin{aligned} \hat{\Lambda}_2(\theta t) - \Lambda_2(\theta t) &= \int_0^{\theta t} \frac{d(\sum_{j=1}^{n_2} \Delta_{2j} I(T_{2j} \leq s))}{\sum_{j=1}^{n_2} I(T_{2j} \geq s)} - \Lambda_2(\theta t) \\ &= \int_0^{\theta t} \frac{d(\sum_{j=1}^{n_2} I(T_{2j} \leq s, \Delta_{2j} = 1))}{\sum_{j=1}^{n_2} I(T_{2j} \geq s)} - \Lambda_2(\theta t) \\ &= \int_0^t \frac{d(\sum_{j=1}^{n_2} I(T_{2j}/\theta \leq s, \Delta_{2j} = 1))}{\sum_{j=1}^{n_2} I(T_{2j}/\theta \geq s)} - \Lambda_2(\theta t) \\ &= \int_0^t \frac{d(\sum_{j=1}^{n_2} N_{2j}(s; \theta))}{\sum_{j=1}^{n_2} Y_{2j}(s; \theta)} - \Lambda_2(\theta t). \end{aligned}$$

Note that $T_{2j}/c = (X_{2j}/c) \wedge (C_{2j}/c)$, so $T_{2j}/\theta = (X_{2j}/\theta) \wedge (\tilde{C}_{2j})$, where $\tilde{C}_{2j} = C_{2j}/\theta$

at $c = \theta$. These results associated with (4.7) and (4.8) imply

$$\begin{aligned}
\hat{\Lambda}_1(t) - \Lambda_1(t) &= \int_0^t \frac{d(\sum_{j=1}^{n_1} N_{1j}(s))}{\sum_{j=1}^{n_1} Y_{1j}(s)} - \Lambda_1(t) \\
&= \int_0^t \frac{d(\sum_{j=1}^{n_1} N_{1j}(s) - \int_0^s \sum_{j=1}^{n_1} Y_{1j}(u) d\Lambda_1(u) + \int_0^s \sum_{j=1}^{n_1} Y_{1j}(u) d\Lambda_1(u))}{\sum_{j=1}^{n_1} Y_{1j}(s)} \\
&\quad - \Lambda_1(t) \\
&= \int_0^t \frac{d(\sum_{j=1}^{n_1} N_{1j}(s) - \int_0^s \sum_{j=1}^{n_1} Y_{1j}(u) d\Lambda_1(u))}{\sum_{j=1}^{n_1} Y_{1j}(s)} \\
&= \int_0^t \frac{d(\sum_{j=1}^{n_1} M_{1j}(s))}{\sum_{j=1}^{n_1} Y_{1j}(s)}, \tag{4.10}
\end{aligned}$$

and

$$\begin{aligned}
\hat{\Lambda}_2(\theta t) - \Lambda_2(\theta t) &= \int_0^t \frac{d(\sum_{j=1}^{n_2} N_{2j}(s; \theta))}{\sum_{j=1}^{n_2} Y_{2j}(s; \theta)} - \Lambda_2(\theta t) \\
&= \int_0^t \frac{d(\sum_{j=1}^{n_2} N_{2j}(s; \theta) - \int_0^s \sum_{j=1}^{n_2} Y_{2j}(u; \theta) d\Lambda_2(\theta u))}{\sum_{j=1}^{n_2} Y_{2j}(s; \theta)} \\
&\quad + \int_0^t \frac{d(\int_0^s \sum_{j=1}^{n_2} Y_{2j}(u; \theta) d\Lambda_2(\theta u))}{\sum_{j=1}^{n_2} Y_{2j}(s; \theta)} - \Lambda_2(\theta t) \\
&= \int_0^t \frac{d(\sum_{j=1}^{n_2} N_{2j}(s; \theta) - \int_0^s \sum_{j=1}^{n_2} Y_{2j}(u; \theta) d\Lambda_2(\theta u))}{\sum_{j=1}^{n_2} Y_{2j}(s; \theta)} \\
&= \int_0^t \frac{d(\sum_{j=1}^{n_2} M_{2j}(s; \theta))}{\sum_{j=1}^{n_2} Y_{2j}(s; \theta)} \tag{4.11}
\end{aligned}$$

for $0 \leq s \leq t$. Write (4.10) and (4.11) as

$$\begin{aligned}
\hat{\Lambda}_1(t) - \Lambda_1(t) &= \sum_{j=1}^{n_1} \int_0^t \frac{1}{\sum_{j=1}^{n_1} Y_{1j}(s)} dM_{1j}(s) \\
&= \int_0^t \frac{1}{Y^1(s)} dM_1(s), \tag{4.12}
\end{aligned}$$

and

$$\begin{aligned}\hat{\Lambda}_2(\theta t) - \Lambda_2(\theta t) &= \sum_{j=1}^{n_2} \int_0^t \frac{1}{\sum_{j=1}^{n_2} Y_{2j}(s; \theta)} dM_{2j}(s; \theta) \\ &= \int_0^t \frac{1}{Y^2(s; \theta)} dM_2(s; \theta),\end{aligned}\tag{4.13}$$

respectively, where

$$\begin{aligned}Y^1(s) &= \sum_{j=1}^{n_1} Y_{1j}(s), \\ Y^2(s; \theta) &= \sum_{j=1}^{n_2} Y_{2j}(s; \theta),\end{aligned}$$

and

$$\begin{aligned}M_1(t) &= \sum_{j=1}^{n_1} \left\{ \Delta_{1j} I(T_{1j} \leq t) - \int_0^t I(T_{1j} \geq s) d\Lambda_1(s) \right\} \\ &= \sum_{j=1}^{n_1} \left\{ N_{1j}(t) - \int_0^t Y_{1j}(s) d\Lambda_1(s) \right\} \\ &= \sum_{j=1}^{n_1} M_{1j}(t),\end{aligned}$$

and

$$\begin{aligned}M_2(t; \theta) &= \sum_{j=1}^{n_2} \left\{ \Delta_{2j} I(T_{2j}/\theta \leq t) - \int_0^t I(T_{2j}/\theta \geq s) d\Lambda_2(\theta s) \right\} \\ &= \sum_{j=1}^{n_2} \left\{ N_{2j}(t; \theta) - \int_0^t Y_{2j}(s; \theta) d\Lambda_2(\theta s) \right\} \\ &= \sum_{j=1}^{n_2} M_{2j}(t; \theta).\end{aligned}$$

As illustrated early in this section, $M_{1j}(t) = N_{1j}(t) - \int_0^t Y_{1j}(s) d\Lambda_1(s)$ and $M_{2j}(t; \theta) =$

$N_{2j}(t; \theta) - \int_0^t Y_{2j}(s; \theta) d\Lambda_2(s\theta)$ are martingales. Moreover,

$$M_1(t) = \sum_{j=1}^{n_1} M_{1j}(t),$$

$$M_2(t; \theta) = \sum_{j=1}^{n_2} M_{2j}(t; \theta)$$

are also martingales. Besides, since the process given at time t by

$$\frac{I(Y^1(s) > 0)}{Y^1(s)} = \begin{cases} \frac{1}{Y^1(s)} & \text{if } Y^1(s) > 0, \\ 0 & \text{if } Y^1(s) = 0, \end{cases}$$

is a left-continuous adapted process with right-hand limits, $\{\mathcal{M} : t \geq 0\}$ given by

$$\begin{aligned} \mathcal{M} &= \int_0^t \frac{I(Y^1(s) > 0)}{Y^1(s)} dM_1(s) \\ &= \int_0^t \frac{d(\sum_{j=1}^{n_1} N_{1j}(s))}{Y^1(s)} - \int_0^t I(Y^1(s) > 0) d\Lambda_1(s) \end{aligned}$$

is a martingale. It follows, since $\mathcal{M}(0) = 0$, that

$$E \int_0^t \frac{d(\sum_{j=1}^{n_1} N_{1j}(s))}{Y^1(s)} = E \int_0^t I(Y^1(s) > 0) d\Lambda_1(s), \quad (4.14)$$

i.e., $E(\mathcal{M}) = 0$. We will use these results for the approximation procedures.

The process $D_n(x, \theta)$ in (4.9) associated with (4.12) and (4.13) yields

$$\begin{aligned} D_n(x, \theta) &= n^{1/2} \int_u^x W_n(t; \theta) \left\{ \sum_{j=1}^{n_1} \int_0^t \frac{1}{Y^1(s)} dM_{1j}(s) \right\} dt \\ &\quad - n^{1/2} \int_u^x W_n(t; \theta) \left\{ \sum_{j=1}^{n_2} \int_0^t \frac{1}{Y^2(s; \theta)} dM_{2j}(s; \theta) \right\} dt \\ &= n^{1/2} \sum_{j=1}^{n_1} \int_u^x \left\{ \int_t^\infty W_n(s; \theta) ds \right\} \frac{1}{Y^1(t)} dM_{1j}(t) \\ &\quad - n^{1/2} \sum_{j=1}^{n_2} \int_u^x \left\{ \int_t^\infty W_n(s; \theta) ds \right\} \frac{1}{Y^2(t; \theta)} dM_{2j}(t; \theta). \end{aligned} \quad (4.15)$$

So by Theorem 4.1.1,

$$\begin{aligned}
D_n(x, \hat{\theta}) &= D_n(x, \theta) - \frac{\gamma_x}{\gamma_v} D_n(v, \theta) + o_p(1) \\
&= n^{1/2} \sum_{j=1}^{n_1} \int_u^x \left\{ \int_t^\infty W_n(s; \theta) ds \right\} \frac{1}{Y^1(t)} dM_{1j}(t) \\
&\quad - n^{1/2} \sum_{j=1}^{n_2} \int_u^x \left\{ \int_t^\infty W_n(s; \theta) ds \right\} \frac{1}{Y^2(t; \theta)} dM_{2j}(t; \theta) \\
&\quad - \frac{\gamma_x}{\gamma_v} \left[n^{1/2} \sum_{j=1}^{n_1} \int_u^v \left\{ \int_t^\infty W_n(s; \theta) ds \right\} \frac{1}{Y^1(t)} dM_{1j}(t) \right. \\
&\quad \left. - n^{1/2} \sum_{j=1}^{n_2} \int_u^v \left\{ \int_t^\infty W_n(s; \theta) ds \right\} \frac{1}{Y^2(t; \theta)} dM_{2j}(t; \theta) \right] + o_p(1),
\end{aligned}$$

where $M_{1j}(t)$, $M_{2j}(t; \theta)$ are martingales.

Consequently, the distribution of the process $D_n(x, \hat{\theta})$ is determined by martingale. The distribution of M_{ij} is, however, unknown. More specifically, if we knew the stochastic structure of the martingale process M_{ij} , we could easily simulate the process $D_n(x, \theta)$ after replacing the unknown quantity θ by the consistent estimator $\hat{\theta}$. To overcome this problem is to replace M_{ij} in (4.15) by a similar process, say M_{ij}^* which has a known distribution.

By Fleming and Harrington (1991, Lemma 2.3.2, Theorem 2.5.3), we have

$$E\{M_{1j}(t)\} = 0,$$

$$E\{M_{2j}(t; \theta)\} = 0,$$

so

$$\text{var}\{M_{1j}(t)\} = E\{N_{1j}(t)\},$$

$$\text{var}\{M_{2j}(t; \theta)\} = E\{N_{2j}(t; \theta)\}$$

for $j = 1, \dots, n_i$. Based on the above, as in Lin et al. (1993), natural candidates for $M_{1j}^*(t)$, $M_{2j}^*(t; \theta)$ are $N_{1j}(t)G_j$ and $N_{2j}(t; \theta)G_j$, respectively, where $N_{1j}(t)$ and $N_{2j}(t; \theta)$ are observed counting processes, and $\{G_j : j = 1, \dots, n_i\}$ is a random sample of standard normal variables. Therefore, the distribution of the process $D_n(x, \theta)$ in (4.15) can be approximated by $D^*(x)$ to be specified in (4.17). Let

$$A(x) = D^*(x) - \frac{\hat{\gamma}_x}{\hat{\gamma}_v} D^*(v), \quad (4.16)$$

where

$$\hat{\gamma}_x = \int_u^x W_n(t; \hat{\theta}) t d\hat{\Lambda}(t),$$

and

$$\begin{aligned} D^*(x) &= n^{1/2} \sum_{j=1}^{n_1} \int_u^x \left\{ \int_t^\infty W_n(s; \hat{\theta}) ds \right\} \frac{1}{Y^1(t)} d\{G_j N_{1j}(t)\} \\ &\quad - n^{1/2} \sum_{j=1}^{n_2} \int_u^x \left\{ \int_t^\infty W_n(s; \hat{\theta}) ds \right\} \frac{1}{Y^2(t; \hat{\theta})} d\{G_j N_{2j}(t; \hat{\theta})\} \\ &= n^{1/2} \sum_{j=1}^{n_1} G_j \Delta_j I(u \leq T_{1j} \leq x) \left\{ \int_{T_{1j}}^\infty W_n(s; \hat{\theta}) ds \right\} \frac{1}{Y^1(T_{1j})} \\ &\quad - n^{1/2} \sum_{j=1}^{n_2} G_j \Delta_j I(u \leq T_{2j}/\hat{\theta} \leq x) \left\{ \int_{T_{2j}/\hat{\theta}}^\infty W_n(s; \hat{\theta}) ds \right\} \frac{1}{Y^2(T_{2j}/\hat{\theta}; \hat{\theta})}. \end{aligned} \quad (4.17)$$

Note that $A(x)$ can be obtained after replacing θ , M_{ij} by $\hat{\theta}$, $G_j N_{ij}$, respectively.

To understand (4.17), note that

$$\begin{aligned} \int_u^v f(t) dN_{1j}(t) &= \int_u^v f(t) d\{\Delta_{1j} I(T_{1j} \leq t)\} \\ &= \Delta_{1j} I(u \leq T_{1j} \leq v) f(T_{1j}), \end{aligned}$$

and

$$\begin{aligned}\int_u^v f(t) dN_{2j}(t; \hat{\theta}) &= \int_u^v f(t) d\{\Delta_{2j} I(T_{2j}/\hat{\theta} \leq t)\} \\ &= \Delta_{2j} I\left(u \leq \frac{T_{2j}}{\hat{\theta}} \leq v\right) f\left(\frac{T_{1j}}{\hat{\theta}}\right).\end{aligned}$$

By the arguments in Lin et al. (1993), the conditional distribution of the process $A(x)$ given the data is the same in the limit as the unconditional distribution of the process $D_n(x, \hat{\theta})$. More specifically, approximation of the null distribution of the observed process $D_n(x, \hat{\theta})$ can be obtained from the process $A(x)$ by generating a number of G_j 's while holding the observed data fixed after $\hat{\theta}$ is obtained. Note that, in the process $A(x)$, the only random components are G_j 's conditional on the data. Also note that, as in Lin et al. (1993), the process $A(x)$, which consists of independent standard normal samples at each fixed point, converges weakly to the zero-mean Gaussian process. Again, the null distribution of the observed process $D_n(x, \hat{\theta})$ can be approximated by $A(x)$. These results along with graphical methods lead to an idea to construct a lack-of-fit test, which will be described in Section 4.3.

4.3 Lack-of-Fit Test

The accelerated life model assumes that Λ_1 and Λ_2 are related by

$$\Lambda_2(t) = \Lambda_1\left(\frac{t}{\hat{\theta}}\right)$$

for some $\theta > 0$, but otherwise remains unknown. Consider the test problem of

$$H_0 : \Lambda_2(t) = \Lambda_1\left(\frac{t}{\theta}\right) \quad \text{for some } \theta$$

versus

$$H_1 : \Lambda_2(t) \neq \Lambda_1\left(\frac{t}{\theta}\right) \quad \text{for any } \theta.$$

The process $D^*(x)$ in (4.17) can be written as

$$\sum_{\substack{(i=1, 1 \leq j \leq n_1) \\ (i=2, 1 \leq j \leq n_2)}} G_j \Delta_{ij} I(u \leq T_{ij}/\hat{\theta}^{i-1} \leq x) h^*(T_{ij}),$$

where

$$h^*(T_{ij}) = I(i=1) \frac{\{\int_{T_{1j}}^{\infty} W_n(s; \hat{\theta}) ds\}}{Y^1(T_{1j})} - I(i=2) \frac{\{\int_{T_{2j}/\hat{\theta}}^{\infty} W_n(s; \hat{\theta}) ds\}}{Y^2(T_{2j}; \hat{\theta})}$$

with

$$Y^1(s) = \sum_{j=1}^{n_1} Y_{1j}(s) = \sum_{j=1}^{n_1} I(T_{1j} \geq s),$$

$$Y^2(s; \hat{\theta}) = \sum_{j=1}^{n_2} Y_{2j}(s; \hat{\theta}) = \sum_{j=1}^{n_2} I\left(\frac{T_{2j}}{\hat{\theta}} \geq s\right),$$

where $(T_j/\hat{\theta}) = (X_{2j}/\hat{\theta}) \wedge (C_{2j}/\hat{\theta})$ for $j = 1, \dots, n_2$.

So, the process $\hat{D}_n(x, \hat{\theta})$ in (4.16) can be seen as a special case of

$$\sum_{\substack{(i=1, 1 \leq j \leq n_1) \\ (i=2, 1 \leq j \leq n_2)}} G_j \Delta_{ij} I(u \leq T_{ij}/\hat{\theta}^{i-1} \leq x) h(T_{ij}) \quad (4.18)$$

for some suitable function $h(T_{ij})$. Note that, in (4.18), only G_j 's are random conditional on the data.

Recall the observed process,

$$D_n(x, \hat{\theta}) = n^{1/2} \int_u^x W_n(t; \hat{\theta}) \{ \hat{\Lambda}_1(t) - \hat{\Lambda}_2(\hat{\theta}t) \} dt$$

for $u < x \leq v$. The null distribution of this process can be approximated by the zero-mean Gaussian process $A(x)$. Note that the process $A(x)$ is conditional on the observed data $\{T_{ij}, \Delta_{ij}\}$ whereas the process $D_n(x, \hat{\theta})$ is unconditional. To approximate the null distribution of the process $D_n(x, \hat{\theta})$, we simulate a number of realizations from $A(x)$ by repeatedly generating normal random samples G_j 's while fixing the observed data T_{ij}, Δ_{ij} . Note that the observed process $D_n(x, \hat{\theta})$ randomly fluctuates around zero under the null hypothesis. The observed process $D_n(x, \hat{\theta})$ can be compared with simulated realizations of the approximation process $A(x)$ for a visual inspection of the model adequacy, i.e., plotting the observed process $D_n(x, \hat{\theta})$ along with the approximation process $A(x)$ in large quantities by generating independent standard normal samples G_j 's can visually present the validity of the model. Remarkable pattern of the observed process $D_n(x, \hat{\theta})$ compared to the simulated realizations of the approximation process $A(x)$ would indicate the departure from the null hypothesis.

The following notation will be used: a capital letter (S) refers to statistic, a small-case letter (s) to its observed value, and the corresponding quantities under the Gaussian approximations are indicated by ' $\wedge$ ' (e.g. $\hat{S}$). Since the process $D_n(x, \hat{\theta})$ randomly fluctuates around zero under the null hypothesis, the statistic

$$S = \sup_x |D_n(x, \hat{\theta})|$$

can be used as a numerical measure for checking the adequacy of the model. Define

$$\hat{S} = \sup_x |A(x)|.$$

The distribution of $S = \sup_x |D_n(x, \hat{\theta})|$ can be approximated by the conditional distribution of $\hat{S} = \sup_x |A(x)|$ given the data under the null hypothesis. So an unusually large value of $S = \sup_x |D_n(x, \hat{\theta})|$ would suggest that the model is misspecified. By arguments of Lin et al. (1993), $P(\hat{S} \geq s)$ converges almost surely to $P(S \geq s)$ as $n \rightarrow \infty$. So the p -value $P(S \geq s)$, which serves as a measure of the strength of evidence against H_1 , can be approximated by $P(\hat{S} \geq s)$. Note that a small p -value means that the null hypothesis is strongly rejected. $P(\hat{S} \geq s)$ can be empirically estimated by generating a large number of normal samples G_j 's. Note that $P(\hat{S} \geq s)$ is conditional on the observed data $\{T_{ij}, \Delta_{ij}\}$. Theorem 4.3.1 shows that this test is asymptotically consistent against a general departure from the two-sample accelerated life model.

Theorem 4.3.1. *Under the assumptions of Proposition 3.1.1 and Theorem 4.1.1, the test based on $\sup_x |D_n(x, \hat{\theta})|$ is consistent against H_1 .*

Proof. Assume that

$$\hat{\theta} \rightarrow \theta^*$$

and

$$\int_u^x W_n(t; c) \{\hat{\Lambda}_1(t) - \hat{\Lambda}_2(ct)\} dt \rightarrow \int_u^x W(t; \theta^*) \{\Lambda_1(t) - \Lambda_2(\theta^* t)\} dt$$

as $n \rightarrow \infty$ under the alternative hypothesis. It suffices to show that $\sup_x |D_n(x, \hat{\theta})|$ is

nonzero under the alternative hypothesis. Under the alternative hypothesis,

$$\sup_x \left| \int_u^x W(t; \theta^*) \{ \Lambda_1(t) - \Lambda_2(\theta^* t) \} dt \right| > 0.$$

It follows that the asymptotic equivalent form of $\sup_x |D_n(x, \hat{\theta})|$,

$$n^{1/2} \sup_x \left| \int_u^x W(t; \theta^*) \{ \Lambda_1(t) - \Lambda_2(\theta^* t) \} dt \right| \rightarrow \infty$$

as $n \rightarrow \infty$.

□

CHAPTER V

SIMULATION STUDIES

In this chapter, we illustrate simulation studies for the proposed methods in this research. In simulation studies, we look at the common situations in statistical analysis for censored data: both the survival and the censoring distributions belong to the Log-logistic power-scale family, the survival distributions are Weibull and the censoring distributions are Lognormal, and the survival distributions are Lognormal and the censoring distributions are logarithms of uniform variables. Specifically, for $i = 1, 2, j = 1, \dots, m$, we consider the following three cases:

- Case 1. $2^{1-i}X_{ij}$ has density $2t(1+t^2)^{-2}, t > 0$, and C_{ij} has density $2h^2t(1+h^2t^2)^{-2}, t > 0$, for some constant h ;
- Case 2. $2^{1-i}X_{ij}$ has density $2t\exp(-t^2), t > 0$, and $\log(C_{ij})$ is normal with mean h and standard deviation 1.
- Case 3. $\log(2^{1-i}X_{ij})$ is standard normal and $\log(C_{ij})$ is uniform($h, 1+h$) for some constant h .

The constant h is chosen to obtain the censoring proportion of interest. In each case, $\theta = 2$, $n_1 = n_2 = m$, and the two-sample has the same censoring pattern that does not depend on the scale parameter.

The simulations on the tests were based on 500 repetitions for censored portions of 25%, 40%, 50% for sample size of $m = 25, 50$, and 1000 realizations for the p -value.

The results of the proposed tests are presented in Table 5.1. As seen in Table 5.1, the proposed method behaves well on the whole, even when it undergoes heavy censoring. Denote $1 - F$ by $\bar{F}$. The weight functions used for the proposed methods are $\bar{F}^{1+\omega}(t)F(t)/t$, $\bar{F}^{2+\omega}(t)F^2(t)/t$, $\bar{F}^{2+\omega}(t)F(t)/t$ and $\bar{F}^{1+\omega}(t)F^2(t)/t$. Sample version of the weight $\bar{F}^{2+\omega}(t)F^2(t)/t$, e.g., is $(1 - \hat{F}(t; c))^{2+\hat{\omega}}\hat{F}^2(t; c)/t$, where

$$\hat{\omega} = \frac{\hat{\pi}}{1 - \hat{\pi}}$$

with the sampling proportion $\hat{\pi} = 1 - \frac{1}{n} \sum_{ij} \Delta_{ij}$. Note that $\hat{\omega}$ is the almost sure limit of ω . For comparison, Yang's method was also implemented using the same data sets. Table 5.2 shows the simulation results of Yang's method for $\theta = 2$, censored portions of 25%, 40%, and sample sizes $m = 25, 50$ with weight pairs $(\bar{F}^{1+\omega}(t)F^2(t)/t, \bar{F}^{1+\omega}(t)F(t)/t)$, $(\bar{F}^{2+\omega}(t)F^2(t)/t, \bar{F}^{1+\omega}(t)F^2(t)/t)$, $(\bar{F}^{2+\omega}(t)F(t)/t, \bar{F}^{2+\omega}(t)F^2(t)/t)$, $(\bar{F}^{1+\omega}(t)F(t)/t, \bar{F}^{2+\omega}(t)F(t)/t)$ and $(\bar{F}^{2+\omega}(t)F^2(t)/t, \bar{F}^{2+\omega}(t)F(t)/t)$. Both the proposed method in this research and Yang's method provide similar results in cases 1 and 2. However, behaviors of the proposed method on case 3 are remarkably better than Yang's method, especially when it is exposed to heavy censoring. Figure 5.1 visually shows the comparisons of Yang and the proposed method on case 3. From various simulations with the weight function $\bar{F}^\omega(t)F^\delta(t)$ of Yang's method, $(\bar{F}^{2+\omega}(t)F^2(t)/t, \bar{F}^{2+\omega}(t)F(t)/t)$ was used as the best weights pair for comparison in Figure 5.1. To compare with this, weight functions, $\bar{F}^{2+\omega}(t)F^2(t)/t$ and $\bar{F}^{2+\omega}(t)F(t)/t$, were used for the proposed methods.

Figure 5.1: Comparison of Yang and the proposed tests on test rejection % of Case 3 with the pair $(\bar{F}^{2+\omega}(t)F^2(t)/t, \bar{F}^{2+\omega}(t)F(t)/t)$ for Yang, $\bar{F}^{2+\omega}(t)F(t)/t$ for New 1 and $\bar{F}^{2+\omega}(t)F^2(t)/t$ for New 2, sample size $m = 25, 50$ and $\alpha = 0.05$.

For easier references, each method with some weight functions is denoted in Figure 5.1 as follows: Yang's method using the weight pair $(\bar{F}^{2+\omega}(t)F^2(t)/t, \bar{F}^{2+\omega}(t)F(t)/t)$ is denoted by Yang (dotted), and New 1 (dashdot) and New 2 (dashed) are the proposed methods using $\bar{F}^{2+\omega}(t)F^2(t)/t$ and $\bar{F}^{2+\omega}(t)F(t)/t$, respectively. We can notice that New 1 and New 2 still perform well even though censoring percentages are relatively high. Note that, in Table 5.2, wp 1 is $(\bar{F}^{1+\omega}(t)F^2(t)/t, \bar{F}^{1+\omega}(t)F(t)/t)$, $(\bar{F}^{2+\omega}(t)F^2(t)/t, \bar{F}^{1+\omega}(t)F^2(t)/t)$ is wp 2, $(\bar{F}^{2+\omega}(t)F(t)/t, \bar{F}^{2+\omega}(t)F^2(t)/t)$ is wp 3, wp 4 is $(\bar{F}^{1+\omega}(t)F(t)/t, \bar{F}^{2+\omega}(t)F(t)/t)$, and wp 5 is $(\bar{F}^{2+\omega}(t)F^2(t)/t, \bar{F}^{2+\omega}(t)F(t)/t)$.

Table 5.1: Performance of the proposed lack-of-fit test under Cases 1-3 of censoring proportion 25%, 40% and 50% for sample size $m = 25, 50$ with several weights, and $\alpha=0.05$.

Case	Censored Portion	m	Test Rejection %			
			$\bar{F}^{1+\omega}F/t$	$\bar{F}^{2+\omega}F^2/t$	$\bar{F}^{2+\omega}F/t$	$\bar{F}^{1+\omega}F^2/t$
1	25%	25	3.4	5.2	4.6	4
		50	3	3	2.4	3.2
	40%	25	5	5.4	5.4	5.4
		50	3.6	3.2	2.8	3
	50%	25	6	6.8	5.8	7
		50	5.4	3.6	3	4.4
2	25%	25	3	2.6	2.4	2
		50	3	2.6	3.2	1.6
	40%	25	3.2	3	3.6	2.2
		50	3	2.8	3.4	1.6
	50%	25	4.4	4.6	5	4.2
		50	4	3.8	3.6	2.6
3	25%	25	5.2	4.8	5.4	5.4
		50	4	3.2	4.2	3.6
	40%	25	7	5.6	5.6	5.8
		50	4.2	3.4	4.6	3.8
	50%	25	6.8	8.4	6.8	6.6
		50	5.8	4.6	5.2	5.2

Table 5.2: Performance of Yang's lack-of-fit test under Cases 1-3 of censoring proportion 25%, 40% and 50% for sample size $m = 25, 50$ with several weights, and $\alpha=0.05$.

Case	Censored Portion	m	Test Rejection %				
			wp 1	wp 2	wp 3	wp 4	wp 5
1	25%	25	5	2	7	7	2.4
		50	2.4	1	4	6.4	2
	40%	25	9	5	10	8.6	6
		50	5	2	6	7.4	4
2	25%	25	2	2	4	5	1.4
		50	1	1	2	2	1
	40%	25	6	2	8	6.4	1.8
		50	1.2	1	7	3	1
3	25%	25	6	7.2	10	10	7.2
		50	4.4	4.2	5	7	4.2
	40%	25	9	10	32	12.6	10
		50	7	6	22	12	6

CHAPTER VI

APPLICATIONS

In this Chapter, we apply the proposed methods for checking the two-sample scale model, or $\Lambda_2(t) = \Lambda_1(t/\theta)$ as an alternative way to two data sets from the literature. The data sets are presented in Tables 6.1 and 6.2. Note that * indicates the censored observations. Figures 6.1 and 6.4 show the estimated cumulative hazards for Λ_1 and Λ_2 of each data set.

The first example we discuss is concerned with the time to death from vaginal cancer or the censorship following treatment of rats with carcinogen 7,12- dimethylbenz(a) anthracene (abbreviated DMBA). Pike(1966), Kalbfleisch and Prentice (1980), Wei and Gail (1983), and Yang (1998) analyzed this data using their methods. Let X_i be the original group i survival time. Kalbfleish and Prentice (1980) assumed that $X_i^* = X_i - g_0$, $i = 1, 2$, obey a two-sample scale model, where g_0 is a guarantee time that must pass before the carcinogen begins to cause cancer, i.e., for the original data, $\Lambda_2(t) = \Lambda_1(t/\theta + g_0(1 - 1/\theta))$, Kalbfleish and Prentice (1980) used $g_0 = 100$. From the quantile-quantile plot, Wei and Gail (1983) argued that the original data could also be reasonably fit by a two-sample scale model. Using some lack-of-fit tests based on certain weighted integrated empirical hazard difference of two groups, Yang (1998) concluded that the original data could be well-fit by the two-sample accelerated life model.

In the vaginal cancer data, the rats are divided into two groups according to

Table 6.1: Times to death from vaginal cancer or censorship following treatment of rats with the carcinogen 7,12-DMBA

Group 1	143	164	188	188	190	190	206	209	213	216	216*
($n_1=19$)	220	227	230	234	244*	246	265	304			
Group 2	142	156	163	198	204*	205	232	232	233	233	233
($n_2=21$)	233	239	240	261	280	280	296	296	323	344*	

pretreatment regime. The effect of this regime was also found to be well described by the two-sample accelerated life model using the proposed methods in this research. The sample censoring proportion $\hat{\pi}$ is $\frac{4}{40} = 0.1$. Thus, the fractional power is $\hat{\omega} = \frac{0.1}{1-0.1} \approx 0.11$. To check the adequacy of the scale models, the weights, $\bar{F}^{1+\omega}(t)F(t)/t$ and $\bar{F}^{2+\omega}(t)F^2(t)/t$, are used. Note that the sample versions of these weights are $(1 - \hat{F}(t; c))^{1+\hat{\omega}}\hat{F}(t; c)/t$ and $(1 - \hat{F}(t; c))^{2+\hat{\omega}}\hat{F}^2(t; c)/t$, respectively. Both $\bar{F}^{1+\omega}(t)F(t)/t$ and $\bar{F}^{2+\omega}(t)F^2(t)/t$ give the point estimate of the scale change 1.12 from the estimating function described in (3.3) in Chapter III.

The overall test $\sup_x |A(x)|$ yields p -values of 0.2770 and 0.5610 with respect to two weights $\bar{F}^{1+\omega}(t)F(t)/t$ and $\bar{F}^{2+\omega}(t)F^2(t)/t$. In our illustrations, $P(\hat{S} \geq s)$ that approximates the p -value $P(S \geq s)$ is empirically estimated by generating 1000 standard normal samples. The results of p -values indicate that the original data can be reasonably fit by the two-sample scale model, i.e., $\Lambda_2(t) = \Lambda_1(t/\theta)$. Figure 6.2 shows $\hat{\Lambda}_1(t)$ and $\hat{\Lambda}_2(1.12t)$. The plot of the observed process $D_n(x, \hat{\theta})$ in a solid line along with 50 simulated processes $A(x)$ in dotted lines in Figure 6.3 supports this conclusion graphically.

The second example is about a record of 35 patients with limited Stage II or IIIA ovarian carcinoma at the Mayo clinic. In the ovarian cancer data, there are 35 patients who are divided into two groups, 15 with low-grade cancer and 20 with well-differentiated cancer. Fleming et al. (1980) discussed that grade of disease was associated with time to progression of disease. Gill and Schumacher (1987) argued that the proportional hazards assumption was rejected using this data. Yang (1998) concluded that the original data could not be reasonably fit by the two-sample accelerated life model from some lack-of-fit tests based on certain weighted integrated empirical hazard difference of the groups.

Using the proposed method in this research, it was also found that the association between the grade of disease and progression of disease could not be described by the two-sample accelerated life model. The sample censoring proportion $\hat{\pi}$ is $\frac{13}{35} = 0.37$. Thus, the fractional power is $\hat{\omega} = \frac{0.37}{1-0.37} \approx 0.59$. As we did in case of vaginal cancer data, to check the validity of the scale model, the proposed lack-of-fit tests with weights, $\bar{F}^{1+\omega}(t)F(t)/t$ and $\bar{F}^{2+\omega}(t)F^2(t)/t$, are applied for this data set. Sample versions of these weights are $(1-\hat{F}(t; c))^{1+\hat{\omega}}\hat{F}(t; c)/t$ and $(1-\hat{F}(t; c))^{2+\hat{\omega}}\hat{F}^2(t; c)/t$, respectively. From the estimating function, the weights, $\bar{F}^{1+\omega}(t)F(t)/t$ and $\bar{F}^{2+\omega}(t)F^2(t)/t$, yield 0.75 and 0.69, respectively for the point estimate of the scale change. The overall test $\sup_x |A(x)|$ yields p -values of 0.0130 and 0.0360 from two weights $\bar{F}^{1+\omega}(t)F(t)/t$ and $\bar{F}^{2+\omega}(t)F^2(t)/t$, respectively. Note that the p -value for the test is based on 1000 realizations. The results of p -values indicate a lack of fit of the accelerated life model, or $\Lambda_2(t) \neq \Lambda_1(t/\theta)$. That is, two groups are not related by scale factor.

Table 6.2: Times for treatment to disease progression of patients with ovarian cancer

Stage II patients	28	89	175	195	309	377*	393*	421*	447*
($n_1=15$)	462	709*	744*	770*	1106*	1206*			
Stage IIA patients	34	88	137	199	280	291	299*	300*	309
($n_2=20$)	351	358	369	369	370	375	382	392	429*
	451	1119*							

The estimated cumulative hazard plot in Figure 6.4 shows that, for the two groups, the estimated cumulative hazards appear quite the same when the time to disease is before 350 days. After that, there is a distinguished difference between the two groups. This pattern is impossible for the two-sample accelerated life model. From the patterns of the observed process $D_n(x, \hat{\theta})$ in a solid line and 50 simulated processes $A(x)$ in dotted lines in Figure 6.5, we visually conclude the same result.

Figure 6.1: Estimated cumulative hazard functions $\hat{\Lambda}_1$ (solid) and $\hat{\Lambda}_2$ (dotted) for the two groups in the vaginal cancer data

Figure 6.2: Estimated cumulative hazard functions $\hat{\Lambda}_1(t)$ (solid) and $\hat{\Lambda}_2(1.12t)$ (dotted) for the vaginal cancer data

Figure 6.3: Plot of the observed process(solid) and 50 approximation processes(dotted) for the vaginal cancer data

Figure 6.4: Estimated cumulative hazard functions $\hat{\Lambda}_1$ (solid) and $\hat{\Lambda}_2$ (dotted) for the two groups in the ovarian cancer data

Figure 6.5: Plot of the observed process(solid) and 50 approximation processes(dotted) for the ovarian cancer data

CHAPTER VII

CONCLUSION

Motivated by the class of estimating functions in Yang (1998), some stochastic processes can be constructed, which are based on integrated cumulative hazard difference. These stochastic processes can be represented by the martingale residuals under the two-sample accelerated life model. Their conditional distributions given the data can be approximated by zero-mean Gaussian processes similar to Lin, Wei & Ying (1993) who investigated model checking issues regarding the Cox regression model.

Due to the fact that the observed process fluctuates around zero under the two-sample accelerated life model, once the estimate of the scale change is obtained from the estimating function, the observed process can be compared with the simulated realizations of the approximation process for a visual inspection in determining the adequacy of the model. Based on the conditional distribution of the observed process, an omnibus test is derived with minimum conditions. Unlike Yang's method, this is asymptotically consistent against a general departure from the two-sample accelerated life model. Thus, the proposed methods provide a very simple but reliable techniques in checking the model without the complicated conditions used in most methods, and also give the visual way for checking the model adequacy, which is very effective during the stage of the model building.

The proposed methods are illustrated with two real examples from the literature.

The first example is about the times to death from vaginal cancer or the censorship following treatment of rats with carcinogen DMBA. Using the proposed methods, this data can be reasonably fit by the two-sample accelerated life model. The second example is concerned with a record of 35 patients with limited stage II or IIIA ovarian carcinoma at the Mayo clinic. In the ovarian cancer data, patients are in two groups with low-grade or well-differentiated cancer, and it is found that the association between the grade of disease and progression of disease could not be described by the two-sample accelerated life model.

BIBLIOGRAPHY

1. Collet, D. (1994). *Modeling Survival Data in Medical Research*. London: Chapman & Hall.
2. Cox, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Sec B.* 34, 187-220.
3. Fleming, T.R. and Harrington, D.P. (1991). *Counting processes and survival analysis*. New York: John Wiley & Sons.
4. Fleming, T.R., O'Fallon, J.R., O'Brien, P.C. and Harrington, D.P. (1980). Modified Kolmogorov-Smirnov test procedures with application to arbitrarily censored data. *Biometrics* 36, 607-625.
5. Gill, R.D. (1980). *Censoring and Stochastic Integrals*. Mathematical Center Tract 124, Amsterdam: Mathematische Centrum.
6. Gill, R.D. (1984). Understanding cox's regression model: a martingale approach. *Journal of the American Statistical Association* 79, 441-447.
7. Gill, R.D., and Schumacher, M. (1987). A simple test of the proportional hazards assumption. *Biometrika* 74, 289-230.
8. Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons.
9. Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457-481.
10. Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons.
11. Lin, D.Y., Wei, L.J. and Ying, Z. (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika* 80, 557-572.
12. Lin, D.Y., Fleming, T.R. and Wei, L.J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* 81, 73-81.
13. Padgett, W.J. and Wei, L.J. (1982). Estimation of the ratio of scale parameters in the two sample problem with arbitrary right censorship. *Biometrika* 69, 252-256.
14. Pike, M.C. (1966). A method of analysis of a certain class of experiments in carcinogenesis. *Biometrika* 18, 303-328.
15. Shorack, G. R. and Wellner, J.A. (1986). *Empirical processes with applications to statistics*. New York: John Wiley & Sons.

16. Therneau, T.M., Grambsch, P.M. and Fleming, T.R. (1990). Martingale-based residuals for survival models. *Biometrika* 77,147-160.
17. Wei, L.J. and Gail, M.H. (1983). Nonparametric estimation for a scale-change with censored observations. *Journal of the American Statistical Association* 78, 312-318.
18. Wei, L.J., Ying, Z. and Lin, D.Y. (1990). Linear regression analysis censored survival data based on rank tests. *Biometrika* 19, 845-851.
19. William, M.J. (1998). *Estimating equations for two-sample scale estimation with censored data*. M.S. Thesis, Texas Tech University, Lubbock, TX, USA.
20. Yang, S. (1997). Extended weighted log-rank estimating functions in censored regression. *Journal of American Statistical Association* 92, 977-984.
21. Yang, S. (1998). Scale estimators and lack-of-fit test for life models. *Biometrics* 54, 1040-1052. Correction (2000) in *Biometrics* 56, 324.

APPENDIX

Matlab codes for the simulation results of lack-of-fit tests in Table 5.1

```
clear all

rep=500;

n1=50;

n2=50;

Ma = zeros(rep,3);

for k=1:rep

% survival dist.(T) = Log-logistic, censoring dist.(L) = Log-logistic
% c=.31 for 25% censoring proportion, .52 for 40%, and .72 for 50%
c=.31;

T1=sqrt(1./((1-rand(n1,1)) - 1));

T2=2*sqrt(1./((1-rand(n2,1)) - 1));

T=[T1;T2];

L=sqrt((1./((1-rand((n1+n2),1)) - 1)./c^2));

% survival dist.(T) = Weibull, censoring dist.(L) = Lognormal
% c=.09 for 25% censoring proportion, .7 for 40%, and .08 for 50%
% c=0.9;

% T1 = (-log(1-rand(n1,1))).^(1/2);

% T2 = 2*(-log(1-rand(n1,1))).^(1/2);
```

```

% T = [T1;T2];

% L = exp(randn((n1+n2),1)+c);

% survival dist.(T) = Lognormal, censoring dist.(L) = log(Uniform)
% c=.59 for 25% censoring proportion, .59 for 40%, and -.17 for 50%
% c=.59;

% T1=exp(randn(n1,1));
% T2=2*exp(randn(n2,1));
% T=[T1;T2];
% L=exp(rand(n1+n2,1)+c);

t = min(T,L);

delta = (T<=L);

X = [t delta];

perc(k) = 100*(1-sum(delta)/(n1+n2));

n=length(X);

pc = length(find(X(:,2)==0))/length(X);

om = pc/(1-pc);

X1 = X(1:n1,1); d1 = X(1:n1,2);

X2 = X(n1+1:n,1); d2 = X(n1+1:n,2);

I1 = find(d1); I2 = find(d2);

x1 = X1(I1); x2 = X2(I2);

```

```

[ox1 J1] = sort(x1); [ox2 J2] = sort(x2);

[n11 m11] = size(ox1); [n22 m22] = size(ox2);

diff1 = x1 - [0;x1(1:(n11-1))]; diff2 = x2 - [0;x2(1:(n22-1))];

k1 = find(diff1); k2 = find(diff2);

oxx1 = ox1(k1); oxx2 = ox2(k2);


oxx = [oxx1 ; oxx2];

oxxdif = oxx - [ 0 ; oxx(1:(length(oxx)-1),1)];

oxx = sort(oxx(find(oxxdif)));


cc = 50;

r = 1;

ccc = 2*cc + 1;

c = 2+ r*[(1:ccc)' - (cc+1)]/cc;

m = length(c);


y1 = zeros(length(oxx),m);

y2 = zeros(length(oxx),m);

cF = zeros(length(oxx),m);

L1h = zeros(length(oxx),m);

cL2h = zeros(length(oxx),m);

oyyy = zeros(length(oxx),m);

```



```

oyy = zeros(length(oxx),m);

for j=1:m

    oyy = [ oxx1 ; oxx2/c(j)];

    moxy = min(max(oxx1),max(oxx2/c(j)));

    oyydif = oyy - [ 0 ; oyy(1:(length(oyy)-1),1)];

    oyy = sort(oyy(find(oyydif)));

    loxy = find(oyy == moxy);

    lxy(j)=loxy;

    oyyy(:,j) = oyy;

    cnum1 = zeros(length(oyy),1);

    cnum2 = zeros(length(oyy),1);

    cden1 = zeros(length(oyy),1);

    cden2 = zeros(length(oyy),1);

    cnd = zeros(length(oyy),1);

    cr1 = zeros(length(oyy),1);

    cr2 = zeros(length(oyy),1);

    for i=1:loxy

        cnum1(i) = sum(x1 == oyy(i));

        cnum2(i) = sum(x2/c(j) == oyy(i));

        cden1(i) = sum(X1 >= oyy(i));

```

```

        cden2(i) = sum(X2/c(j) >= oyy(i));

        cnd(i) = (cnum1(i)+cnum2(i))/(cden1(i) + cden2(i));

        cr1(i) = cnum1(i)/cden1(i);

        cr2(i) = cnum2(i)/cden2(i);

    end

```

```

    y1(:,j) = cden1;

    y2(:,j) = cden2;

    cF(:,j)=1-cumprod(1-cr1);

    cF(:,j) = 1 - cumprod(1 - cnd);

    L1h(:,j) = cumsum(cr1);

    cL2h(:,j) = cumsum(cr2);

```

```

end

```

```

W = (1 - cF).^ (1+om).*(cF.^1)./oyyy;

```

```

D = zeros(length(oxx),m);

```

```

for j=1:m

```

```

    for i=1:lxj(j)-1

```

```

        D(i,j) = W(i,j).*(L1h(i,j) - cL2h(i,j)).*[oyyy(i+1,j) - oyyy(i,j)];

```

```

    end

```

```

end

```

```

D(isnan(D))=0;

Dnc = sqrt(n)*sum(D);

[minDnc, loc] = min(Dnc.^2);

th = c(loc);

q = lxy(loc);

Dl = D(:,loc);

obD = sqrt(n)*cumsum(Dl(1:q));

supobD = max(abs(obD));

B=1000;

siD = zeros(q,B);

oxt = [oxx1 ; oxx2/th];

oxtdif = oxt - [ 0 ; oxt(1:(length(oxt)-1),1)];

oxt = sort(oxt(find(oxtdif)));

l1=length(find(oxx1<=min(max(oxx1),max(oxx2/th))));

l2=length(find(oxx2/th<=min(max(oxx1),max(oxx2/th))));

for j=1:B

    pden1 = zeros(q,1);

    pden2 = zeros(q,1);

    pnum1 = zeros(q,1);

    pnum2 = zeros(q,1);

```

```

pnd = zeros(q,1);

R1 = zeros(q,1);

R2 = zeros(q,1);

r1 = zeros(q,1);

r2 = zeros(q,1);

N1 = zeros(q,1);

N2 = zeros(q,1);

G = randn(length(oxt),1);

for i=1:q

    N1(i) = sum(G(i).*(x1 == oxt(i)));

    N2(i) = sum(G(i).*(x2/th == oxt(i)));

    pnum1(i) = sum(x1 == oxt(i));

    pnum2(i) = sum(x2/th == oxt(i));

    pden1(i) = sum(X1 >= oxt(i));

    pden2(i) = sum(X2/th >= oxt(i));

    pnd(i) = (pnum1(i)+pnum2(i))/(pden1(i)+pden2(i));

    R1(i) = N1(i)/pden1(i);

    R2(i) = N2(i)/pden2(i);

    r1(i) = pnum1(i)/pden1(i);

    r2(i) = pnum2(i)/pden2(i);

end

```

```

F = 1- cumprod(1 - pnd);

W1 = (1 - F).^ (1+om).*(F.^1)./oxt(1:q);

La = cumsum(pnd);

gax = cumsum(W1(1:q).*oxt(1:q).*([La(2:q);La(q)]-La(1:q)));

gav = sum(W1(1:q).*oxt(1:q).*([La(2:q);La(q)]-La(1:q)));

D11 = sqrt(n)*cumsum(W1.*cumsum(R1).*([oxt(2:q);oxt(q)]-oxt(1:q)));

D21 = sqrt(n)*cumsum(W1.*cumsum(R2).*([oxt(2:q);oxt(q)]-oxt(1:q)));

D12 = sqrt(n)*sum(W1.*cumsum(R1).*([oxt(2:q);oxt(q)]-oxt(1:q)));

D22 = sqrt(n)*sum(W1.*cumsum(R2).*([oxt(2:q);oxt(q)]-oxt(1:q)));

siD(1:q,j) = D11(1:q) - D21(1:q) - (gax./gav).*(D12 - D22);

end

supsiD = sort(max(abs(siD)));

csiD = supsiD(round(B*95/100));

Ma(k,1) = supobD;

Ma(k,2) = csiD;

Ma(k,3) = (supobD >= csiD);

end

```