

# TREE-STRUCTURED EXTREME VALUE MODEL REGRESSION

Hongshik Ahn

Division of Biometry and Risk Assessment  
National Center for Toxicological Research  
Jefferson, Arkansas 72079

*Key Words and Phrases: Bootstrap; Recursive partitioning; Survival analysis; Exponential regression; Weibull regression.*

## ABSTRACT

This paper gives a method for fitting piecewise Weibull regression models to censored survival data. Recursive stratification is performed using statistical tests for variances and residual analysis. The average squared error of the residuals is used as a splitting criterion in cross-validation. The bootstrap resampling is used to keep the probability of a type I error of the method (the error of finding two or more strata when there is one). Real and simulated data are used for illustration.

## 1. INTRODUCTION

Many regression techniques are available for use with censored survival data. One way to examine the relationship of covariates to survival time is through a regression model in which the latter has a distribution that depends on the covariates. Parametric regression models would be appropriate for this situation. Among the parametric models, the Weibull distribution is an important life distribution and many statistical methods are based on it. However, as in the normal regression, the model is often difficult to interpret, especially when there are many correlated

covariates. One way to avoid this disadvantage is to stratify the data according to particular covariate values and fit separate Weibull regression models to each stratum. In this paper, we explore tree-structured Weibull regression modeling for survival data.

## 2. WEIBULL AND EXTREME VALUE REGRESSION MODELS

The Weibull distribution includes the exponential if the value of the shape parameter equals 1. The exponential regression model requires that individuals have constant hazard functions. In contrast, the Weibull model does not require a constant hazard function and hence is useful in many situations. Pike (1966), Peto and Lee (1973) and Nelson (1972) discussed this model.

Let  $T_1, \dots, T_n$  and  $C_1, \dots, C_n$  be independent random variables, where  $C_i$  is the censoring time associated with the survival time  $T_i$ ,  $i = 1, \dots, n$ . We observe  $(W_1, \delta_1), \dots, (W_n, \delta_n)$ , where  $W_i = \min\{T_i, C_i\}$ ,  $\delta_i = I(T_i \leq C_i)$  and  $I(\cdot)$  is the indicator function. Assume that for each  $i$ , a  $p$ -dimensional covariate vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  independent of  $T_i$  is available.

### 2.1 Model

The probability density function of Weibull random variable  $T$  given  $\mathbf{x}$  is

$$f(t|\mathbf{x}) = (\sigma\alpha_x)^{-1}(t/\alpha_x)^{\sigma^{-1}-1} \exp\{-(t/\alpha_x)^{\sigma^{-1}}\}, \quad t > 0. \quad (1)$$

Let  $\mu_x = \mathbf{x}\boldsymbol{\beta} = \ln \alpha_x$ . Then we can rewrite (1) as

$$f(t|\mathbf{x}) = \sigma^{-1} e^{-\mu_x} (te^{-\mu_x})^{\sigma^{-1}-1} \exp\{-(te^{-\mu_x})^{\sigma^{-1}}\}, \quad t > 0.$$

The probability density function of  $Y = \ln T$  given  $\mathbf{x}$  is

$$f(y|\mathbf{x}) = \sigma^{-1} \exp[\{(y - \mathbf{x}\boldsymbol{\beta})/\sigma\} - \exp\{(y - \mathbf{x}\boldsymbol{\beta})/\sigma\}], \quad -\infty < y < \infty \quad (2)$$

and the survival function of  $Y$  given  $\mathbf{x}$  is

$$S(y|\mathbf{x}) = \exp[-\exp\{(y - \mathbf{x}\boldsymbol{\beta})/\sigma\}], \quad \infty < y < \infty. \quad (3)$$

If we define  $Z = \sigma^{-1}(Y - \mu\mathbf{x})$ , then  $Z$  has the extreme value distribution with probability density function

$$f(z|\mathbf{x}) = \exp(z - e^z), \quad -\infty < z < \infty.$$

This is a location-scale regression model with error  $Z$ . Maximum likelihood estimation is used to estimate  $\boldsymbol{\beta}$  and  $\sigma$ .

## 2.2 Maximum likelihood methods

Since we work with log times,  $y_i = \ln t_i$  represents a log lifetime or log censoring time. From the probability density function (2) and survival function (3) of  $Y$ , the likelihood function for a censored sample based on  $n$  observations is

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma) &= \prod_{i=1}^n [\sigma^{-1} \exp\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\} - \exp\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\}]^{\delta_i} \\ &\quad \times [\exp[-\exp\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\}]]^{1-\delta_i}, \end{aligned}$$

where  $\delta_i$  is 0 if the  $i$ th individual is censored and 1 otherwise. The log-likelihood function is thus,

$$\ln L(\boldsymbol{\beta}, \sigma) = \sum_{i=1}^n [-\exp\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\} + \delta_i \{ \exp\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\} - \ln \sigma \}].$$

If we let  $z_i = (y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma$ , then the first and second derivatives of  $\ln L$  with respect to  $\boldsymbol{\beta}$  and  $\sigma$  are

$$\partial \ln L / \partial \beta_r = \sigma^{-1} \sum_{i=1}^n x_{ir} (e^{z_i} - \delta_i), \quad r = 1, \dots, p,$$

$$\partial \ln L / \partial \sigma = \sigma^{-1} \sum_{i=1}^n \{-\delta_i + z_i(e^{z_i} - \delta_i)\}$$

and

$$\partial^2 \ln L / (\partial \beta_r \partial \beta_s) = -\sigma^{-2} \sum_{i=1}^n x_{ir} x_{is} e^{z_i}, \quad r, s = 1, \dots, p,$$

$$\partial^2 \ln L / \partial \sigma^2 = \sigma^{-2} \sum_{i=1}^n \{\delta_i(1 + 2z_i) - e^{z_i}(2z_i + z_i^2)\},$$

$$\partial^2 \ln L / (\partial \beta_r \partial \sigma) = \sigma^{-2} \sum_{i=1}^n x_{ir} \{(\delta_i - e^{z_i}) - z_i e^{z_i}\}, \quad r = 1, \dots, p.$$

The maximum likelihood equations

$$\partial \ln L / \partial \beta_r = 0, \quad r = 1, \dots, p, \quad \text{and} \quad \partial \ln L / \partial \sigma = 0$$

are solved by the Newton-Raphson method to get the m.l.e.'s  $\hat{\boldsymbol{\beta}}$  and  $\hat{\sigma}^2$  of  $\boldsymbol{\beta}$  and  $\sigma^2$ , respectively. Negative values of  $\sigma$  can arise during the iteration. This is avoided by replacing each negative value with one-half the  $\sigma$  value in the previous iteration.

### 3. TREE-STRUCTURED MODELS

Let  $X$  denote the  $n \times p$  matrix of covariates,  $\boldsymbol{x}^k = (x_{1k}, \dots, x_{nk})$  be the  $n$ -dimensional vector for the  $k$ th covariate, and  $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})$  be the  $p$ -dimensional vector of covariates for the  $i$ th case. We give a brief explanation of the procedure to build a tree-structure in this section. Full details of the procedure and the algorithm are given in Ahn (1992).

#### 3.1 Splitting rules

A binary tree is constructed by splitting the data in each node into two subnodes. Each split is based on a question of the form: Is  $x_{ik} \leq c$ ? Cases satisfying the inequality are sent to the left subnode and otherwise to the right subnode. To choose  $k$ , we study the patterns of the residuals along each covariate axis and select the one for which the residuals appear most non-random.

In each node, we fit Weibull regression model using the sample at the node and get the residuals  $Z$ , where  $Z = (\ln W - \mathbf{x}\hat{\boldsymbol{\beta}})/\hat{\sigma}$ . The data values may be divided into two groups according to the size of the residuals. The two splitting methods (the R and M methods) in Loh (1991) are modified and implemented in the Weibull regression trees for classifying the data into two classes.

In the M method, a covariate vector is considered as a class 1 vector if its corresponding residual is larger than the median of the residuals for the sample and as class 2 otherwise. In the R method, the data values are divided into two groups corresponding to non-negative and negative residuals. In both methods, we compute the  $t$ -statistic for means and Levene's statistic for variances (Levene's, 1960, test), of the data values in the two groups. The  $P$ -value from the larger of the  $t$  and Levene's statistics is computed for each predictor. Suppose the  $i$ th covariate yields the smallest  $P$ -value. The data in the node are split into two parts, with one subset containing the remaining cases, where  $c$  is the average of the two sample means.

The above process is repeated at each subsequent node until either the smallest  $P$ -value is less than the significance level determined by cross-validation (see Section 3.2) or the node contains too few cases.

### 3.2 Stopping rules

A measure of goodness-of-fit is needed to determine whether a node should be split or declared terminal. We use average squared error as a loss function for cross-validation of the Weibull regression trees. The cases in the node are randomly divided into  $V$  subsets. Construct a nested sequence of trees from the data from  $(V - 1)$  subsets and the remaining subset is used as test sample. The procedure is applied  $V$  times, each time leaving out a different subset as test sample. If a cross-validation tree has an estimate of average squared error that is at least  $(1 - f)$  times smaller than that for the trivial tree, then it is considered superior to the trivial tree because of variability due to cross-validation. If the proportion of times (out of  $V$ ) that a superior cross-validation tree is found exceeds another pre-specified number  $\eta$ , the node is split.

### 3.3 Bootstrap selection of parameter values

The bootstrap method introduced in Ahn and Loh (1994) is modified for Weibull regression trees and used to estimate  $f$  and  $\eta$ . We use the hypothesis testing approach of bounding the probability  $\alpha$ , say, that a non-trivial tree results when a single Weibull regression model satisfies all the data. Let  $\hat{\alpha}(f, \eta)$  be the bootstrap estimate of  $\alpha$ , under the hypothesis that no splits are needed. Let  $\hat{f}$  and  $\hat{\eta}$  be the values such that  $\hat{\alpha}(\hat{f}, \hat{\eta})$  is closest to  $\alpha$ . Three methods of searching for  $f$  and  $\eta$  may be used.

1. Fixing  $f = \eta$ , choose the value of  $f$  for which  $\hat{\alpha}(\hat{f}, \hat{f})$  is closest to  $\alpha$ .
2. Fixing  $f = 0$ , select the value of  $\eta$  for which  $\hat{\alpha}(0, \hat{\eta})$  is closest to  $\alpha$ .
3. Fixing  $\eta = 0.5$ , choose the value of  $f$  for which  $\hat{\alpha}(\hat{f}, 0.5)$  is closest to  $\alpha$ .

In each case, a grid with increments of 0.1 are used.

## 4. EXAMPLES

The proposed methods were tested on real and artificial data sets. We report the results in this section. The value of  $\alpha$  was chosen to be .05 in this section.

### 4.1 Simulated data

#### 4.1.1 One Weibull regression model

In the first simulation experiment, survival times were generated from a Weibull distribution with mean  $\mu_i = e^{\mathbf{x}_i \boldsymbol{\beta}}$ , where  $\boldsymbol{\beta} = (3, 3)'$ ,  $\mathbf{x}_i = (1, x_{i1})$  and  $x_{i1} \in \{\pm 1, \pm 2, \pm 3, \pm 4\}$ . Each design point was replicated eight times, giving a total of 64 cases per trial. The value of the shape parameter is  $\sigma = 3$ . Censoring times were independently generated from a Weibull distribution with mean 2000 so that about 20% of the observations were censored. Fifty simulation trials were performed for each of the R and M methods and each of the three bootstrap methods for choosing  $f$  and  $\eta$ .

Table 1: Simulation results for one Weibull regression model using the bootstrap to choose  $f$  and/or  $\eta$ . Nominal significance level is  $\alpha = 0.05$ ; 20% censoring; 50 simulations.

Bootstrap method	M method		R method	
	#splits	freq.	#splits	freq.
1st ( $f = \eta$ )	0	49	0	46
	1	1	1	4
2nd ( $f = 0$ )	0	42	0	45
	1	7	1	4
	2	1	2	1
3rd ( $\eta = .5$ )	0	50	0	46
			1	4

The results are given in Table I. The number of splits and the number of times they were observed are shown in the second and third columns of the table. Since the data were generated from a single Weibull regression model, the correct trees are those with no splits. The probability of a type I error appears to be quite satisfactory, especially in the first and third bootstrap methods.

#### 4.1.2 Two Weibull regression models

In the second experiment, data were generated from two Weibull regression models. The purpose of this experiment is to compare the power of the individual methods in detecting the need to partition the data. Table II gives the simulation results. The powers are larger for the first and third bootstrap estimation methods in both the M and R methods.

#### 4.2 Stanford heart transplant data

The Stanford heart transplant data were previously analyzed by Miller and Halpern (1982) using the Cox (1972), Buckley-James (1979) and Miller (1976) regression methods. Fitting  $\log_{10}(\text{survival time})$  on age at the time of the first transplant and T5 mismatch score, they concluded that T5 mismatch score was not significant and that a quadratic model in age was satisfactory. Miller and Halpern's analysis excluded 5 patients with survival times less than 10 days. The results of their

Table 2: Simulation results for two Weibull regression models, using the bootstrap to find  $f$  and/or  $\eta$ . Significance level is  $\alpha = 0.05$ , 20% censoring, 50 trials.

Bootstrap method	M method		R method	
	#splits	freq.	#splits	freq.
1st ( $f = \eta$ )	0	2	0	1
	1	43	1	41
	2	5	2	8
2nd ( $f = 0$ )	0	32	0	35
	1	18	1	13
			2	2
3rd ( $\eta = .5$ )	0	9	0	3
	1	38	1	42
	2	3	2	5

Table 3: Coefficient estimates and standard errors for Cox regression of log survival time on age and mismatch for the Stanford heart transplant data, with and without five patients with survival times less than 10 days.

	Age		Mismatch	
	Estimate	S.E.	Estimate	S.E.
157 cases	0.030	0.011	0.167	0.183
	Age		Age-squared	
	Estimate	S.E.	Estimate	S.E.
152 cases	-0.146	0.055	0.0023	0.0007

analysis using Cox regression are shown in Table III. Wei, Ying and Lin (1990) re-analyzed the data using linear regression based on rank tests and also concluded that a quadratic model in age was better than a linear one.

FIG. 1(a) shows a scatterplot of  $\log(\text{survival time})$  (to base 10) versus age at transplant, with a smooth curve superimposed, for the 152 patients who survived at least 10 days. While a quadratic in age is preferable to a linear one, it is seen that the smooth curve is quite different from a parabola. If the data are divided at some point of age between 40 and 45 years, then a linear fit might be adequate in each group. FIG. 1(b) shows the same plot without taking logarithms of survival times.

Table 4: Regression estimates of the coefficients, Wald test statistics and  $P$ -values on the covariates for the Stanford heart transplant data with the Weibull regression model. One case with zero survival time was deleted.

Variable	Estimate	S.E.	$\hat{\beta}/\text{S.E.}$	$\chi^2$ (1 d.f.)	$P$ -value (Wald test)
intercept	9.869	0.966	10.22	104.40	< 0.0001
age	-0.056	0.020	-2.79	7.77	0.0053
mismatch	-0.328	0.322	-1.02	1.04	0.3088
shape	1.761	0.148			

On the other hand, a Weibull regression model is fitted to the entire sample. One case with zero survival time was deleted before fitting the model since the log-survival times are used in parametric regression models. Table IV gives the regression estimates, Wald test statistics and the  $P$ -values for the coefficients. As in Cox regression, survival times were shorter for older patients ( $P$ -value is 0.005), but mismatch score was not significant at level 0.05.

The log-likelihoods of the exponential regression and the Weibull regression models for the data are  $-303.37$  and  $-274.23$ , respectively. Hence, the likelihood ratio statistic to test

$$H_0 : \sigma = 1 \text{ versus } H_1 : \sigma \neq 1 \quad (4)$$

or equivalently,

$$H_0 : \text{Exponential regression model} \text{ versus } H_1 : \text{Weibull regression model}$$

is  $\Lambda = -2 \ln[L(\tilde{\beta}, \tilde{\sigma})/L(\hat{\beta}, \hat{\sigma})] = 2(303.37 - 274.23) = 58.28$ , where  $\tilde{\beta}$  and  $\tilde{\sigma}$  are the m.l.e.'s of  $\beta$  and  $\sigma$  under  $H_0$ . The statistic  $\Lambda$  is approximately  $\chi^2$  with degree of freedom 1, when  $H_0$  is true. The  $P$ -value of the test is less than 0.001. Therefore, there is a significant difference between the two models and the Weibull model is more adequate than the exponential regression model for the data.

Next, we fit the data using Weibull regression trees. Before we report our results of the Weibull regression trees, we define  $node(i, j)$  as the  $j$ th node from the left at the  $i$ th level. The root node is  $node(0, 1)$ .

Table 5: Regression estimates and  $P$ -values of Wald tests at the terminal nodes of the tree in FIG. 2.

Node	Parameter	Estimate	S.E.	$\hat{\beta}/\text{S.E.}$	$P$ -value (Wald test)
age $\leq$ 31	intercept	6.830	3.186	2.14	0.0321
	age	0.046	0.124	0.37	0.7105
	mismatch	0.079	1.432	0.06	0.9559
	shape	2.468	0.613		
31 < age $\leq$ 41	intercept	-1.961	4.974	-0.39	0.6934
	age	0.248	0.137	1.81	0.0709
	mismatch	0.918	0.708	1.30	0.1948
	shape	1.708	0.348		
age > 41	intercept	14.489	2.432	5.96	< 0.0001
	age	-0.144	0.049	-2.95	0.0032
	mismatch	-0.671	0.337	-1.99	0.0465
	shape	1.581	0.153		

#### 4.2.1 Using the M method

Applying our method with the second and third bootstrap estimation methods gave a tree with one split, on age at 41.7 years. The first bootstrap method (with  $f = \eta$ ) produced the tree in FIG. 2 which has one more split. The second split was on age again, at 31.2 years. FIG. 3 shows the Kaplan-Meier estimates of the survival distributions for the data in the three terminal nodes. Regression estimates and  $P$ -values for the coefficients are given in Table V. Age and mismatch score were significant at level 0.05 only for the group of patients whose ages were greater than 41 years. Survival times were seen to be shorter for the patients with larger mismatch scores. Neither covariate was significant in the other two groups.

Table VI gives the likelihood ratio test statistics and  $P$ -values to test (4) at each terminal node of the tree in FIG. 2. There was a significant difference between the two models at level 0.01 in the terminal nodes. Therefore, the Weibull regression models were more adequate than the exponential regression models in the nodes.

#### 4.2.2 Using the R method

The R method gave trivial trees for the first and third bootstrap methods. The

Table 6: The likelihood ratio test statistics and  $P$ -values to test (4) at each terminal node of the Weibull regression tree in FIG. 2. One case with zero survival time was deleted.

Node	Log-likelihood for the exponential regression model	Log-likelihood for the Weibull regression model	Likelihood ratio test statistic ( $\chi^2$ with d.f. 1)	$P$ -value
(2, 1)	-54.40	-44.79	19.22	< 0.0001
(2, 2)	-55.07	-50.78	8.58	0.0034
(1, 2)	-181.76	-167.93	27.66	< 0.0001

Table 7: Regression estimates and  $P$ -values of Wald test for the two terminal nodes of the tree in FIG. 4.

Node	Parameter	Estimate	S.E.	$\hat{\beta}/S.E.$	$P$ -value (Wald test)
age $\leq$ 43	intercept	6.785	1.363	4.98	< 0.00001
	age	0.013	0.037	0.35	0.7233
	mismatch	0.481	0.594	0.81	0.4182
	shape	1.930	0.270		
age $>$ 43	intercept	17.462	2.608	6.70	< 0.0001
	age	-0.201	0.051	-3.94	< 0.0001
	mismatch	-0.670	0.342	-1.96	0.0501
	shape	1.548	0.160		

second bootstrap method with  $f = 0$  gave a tree with one split, on age at 43.3 years. FIG. 4 shows the tree obtained from this method, FIG. 5 shows the Kaplan-Meier estimates of the survival distributions for the terminal nodes and Table VII gives the regression estimates. Only for the group of patients whose ages were greater than 43 years was age significant. The  $P$ -value of mismatch score was 0.0501 in the old age group. Neither covariate was significant in the other group. The median survival times were 631 and 323 days in the two nodes.

The likelihood ratio test statistics and  $P$ -values to test (4) at the terminal nodes with the R method are given in Table VIII. As in the M method, the Weibull regression models were preferable to the exponential regression models in the nodes.

Table 8: The likelihood ratio test statistics and  $P$ -values to test (4) at the two terminal nodes of the Weibull regression tree in FIG. 4. One case with zero survival time was deleted.

Node	Log-likelihood for the exponential regression model	Log-likelihood for the Weibull regression model	Likelihood ratio test statistic ( $\chi^2$ with d.f. 1)	$P$ -value
(1, 1)	-134.80	-120.20	29.21	< 0.0001
(1, 2)	-157.25	-146.22	22.06	< 0.0001

Table 9: Cox regression estimates of the coefficients and the  $P$ -values of likelihood ratio, score and Wald tests for the liver transplant data.

Variable	$\hat{\beta}$	S.E.	$\hat{\beta}/\text{S.E.}$	$P$ -value		
				Likelihood ratio	Score	Wald
age	0.039	0.008	5.15	< 0.0001	< 0.0001	< 0.0001
log(albumin)	-2.533	0.648	-3.91	0.0002	0.0001	0.0001
log(bilirubin)	0.871	0.083	10.54	< 0.0001	< 0.0001	< 0.0001
edema	0.859	0.271	3.17	0.0022	0.0013	0.0015
log(protime)	2.380	0.767	3.10	0.0037	0.0020	0.0019

### 4.3 Mayo liver transplant data

The Mayo liver transplant data are published in Fleming and Harrington (1991, Appendix D.1). The survival times and censoring indicators were known for the 418 patients. Dickson *et al.* (1989) selected five variables: age in years, albumin (in mg/dl), serum bilirubin (in mg/dl), presence of edema (0 = no edema and no diuretic therapy for edema; 0.5 = edema present for which no diuretic therapy was given, or edema resolved with diuretic therapy; 1 = edema despite diuretic therapy), and prothrombin time (in seconds, abbreviated to “protime” here).

Dickson *et al.* (1989) fitted a Cox proportional hazards model with covariates age, log(albumin), log(bilirubin), edema and log(protime). The regression estimates of the coefficients and other summary statistics are given in Table IX. Using martingale residuals, Lin, Wei and Ying (1992) found that even though a log transformation of bilirubin was suggested, the resulting model was still not satisfactory.

Table 10: Regression estimates of coefficients and  $P$ -values of Wald test for the liver transplant data for the Weibull regression model.

Parameter	Estimate	S.E.	$\hat{\beta}/\text{S.E.}$	$\chi^2$ (1 d.f.)	$P$ -value (Wald test)
intercept	12.218	1.454	8.40	70.64	< 0.0001
age	-0.027	0.005	-5.03	25.35	< 0.0001
log(albumin)	1.669	0.428	3.90	15.24	< 0.0001
log(bilirubin)	-0.577	0.056	-10.37	107.54	< 0.0001
edema	-0.630	0.181	-3.49	12.18	0.0005
log(protime)	-1.758	0.525	-3.35	11.22	0.0008
shape	0.681	0.043			

In this paper, we applied the Weibull regression trees to the data. First, a Weibull regression model was fitted to the whole sample. The regression estimates of the parameters, Wald test statistics and  $P$ -values for the coefficients were given in Table X. As in the Cox model, all the covariates are significant at level 0.05.

The log-likelihoods of the exponential and Weibull regression models for the data were  $-329.89$  and  $-314.38$ , respectively. The  $P$ -value of test (4) was less than 0.01. Therefore, there was a significant difference between the two models and the Weibull model was more adequate than the exponential regression model.

Next, we analyzed the data using Weibull regression trees. Trivial trees were obtained in the M method. Hence, we report only the result of applying our procedure with the R method. The first and second bootstrap methods gave one split, on log(bilirubin) at 1.15. The third bootstrap method (with  $\eta = 0.5$ ) produced the tree in FIG. 6 which has one more split. The second split was on edema at 0.2. FIG. 7 shows the Kaplan-Meier estimates of the survival times for the three terminal nodes. Regression estimates and  $P$ -values are given in Table XI. Since the covariate edema has constant value 0, it was not included in the model at *node*(2, 3) ( $\log(\text{bilirubin}) > 1.15$  and  $\text{edema} \leq 0.2$ ).

The covariates log(bilirubin) and log(protime) were significant in two nodes of the tree at level 0.05. Survival times tended to be shorter for the patients with larger log(bilirubin) and log(protime) values. Age affected the survival time significantly in *node*(1, 1) and *node*(2, 4). Survival times were shorter for the patients with smaller

Table 11: Regression estimates of the parameters and  $P$ -values of Wald test at the terminal nodes of the tree in FIG. 6.

Node	Parameter	Estimate	S.E.	$\hat{\beta}/\text{S.E.}$	$P$ -value (Wald test)
log(bilirubin) $\leq 1.15$	intercept	12.189	2.187	5.57	< 0.0001
	age	-0.032	0.008	-3.85	0.0001
	log(albumin)	0.690	0.730	0.95	0.3445
	log(bilirubin)	-0.746	0.165	-4.52	< 0.0001
	edema	-0.992	0.353	-2.81	0.0050
	log(protime)	-1.065	0.728	-1.46	0.1433
log(bilirubin) $> 1.15$ & edema $\leq 0.2$	intercept	13.535	2.422	5.59	< 0.0001
	age	-0.011	0.007	-1.53	0.1249
	log(albumin)	2.043	0.496	4.12	< 0.0001
	log(bilirubin)	-0.290	0.145	-2.00	0.0459
	log(protime)	-3.093	0.974	-3.18	0.0015
	shape	0.603	0.062		
log(bilirubin) $> 1.15$ & edema $> 0.2$	intercept	24.334	6.323	3.85	0.0001
	age	-0.074	0.023	-3.24	0.0012
	log(albumin)	0.906	1.260	0.72	0.4721
	log(bilirubin)	-0.365	0.259	-1.41	0.1585
	edema	0.607	0.718	0.84	0.3982
	log(protime)	-5.919	2.361	-2.51	0.0122
	shape	0.738	0.112		

Table 12: The likelihood ratio test statistics and  $P$ -values to test (4) at each terminal node of the tree in FIG. 6.

Node	Log-likelihood for the exponential regression model	Log-likelihood for the Weibull regression model	Likelihood ratio test statistic ( $\chi^2$ with d.f. 1)	$P$ -value
(1, 1)	-185.41	-177.13	16.56	< 0.0001
(2, 3)	-97.80	-88.43	18.75	< 0.0001
(2, 4)	-37.16	-35.42	3.47	0.0625

log(albumin) values in *node*(2, 3). Survival times were shorter for the patients with larger edema values in *node*(1, 1).

Table XII gives the likelihood ratio test statistics and  $P$ -values to test (4) at each terminal node of the tree in FIG. 6. The log-likelihoods were significantly larger for the Weibull regression models except for *node*(2, 4) at level 0.05.

## 5. CONCLUSION

This research was motivated by the goal of keeping the fitted models as simple as possible, for ease of interpretation. This goal can be met by using tree-structure and the bootstrap. Tree-structure is a powerful technique that can break down complex models into simple pieces. Several forms of splitting a node and bootstrapping were considered to study their relative effectiveness as well as to demonstrate the variety of techniques available. The examples suggest that any form of bootstrapping would produce satisfactory control of the probability of a type I error. The first and third bootstrap methods seem to yield the best power in both the M and R methods.

## ACKNOWLEDGEMENTS

Research partially supported by ARO grant DAAL03-91-G-0111 and NSF grant DMS88-03271. The author would like to express sincere thanks to Professor Wei-Yin Loh for his discussion of this paper and most valuable suggestions. The author is also grateful to the editor and the referees for their helpful comments.

## BIBLIOGRAPHY

- Ahn, H. (1992). "Survival modeling through regression trees," Unpublished Ph.D. Thesis, Department of Statistics, University of Wisconsin-Madison.
- Ahn, H. and Loh, W-Y. (1994). "Tree-structured proportional hazards regression modeling," *Biometrics*, To appear.
- Buckley, J. and James, I. (1979). "Linear regression with censored data," *Biometrika*, **66**, 429-436.
- Cox, D. R. (1972). "Regression models and life-tables," *Journal of the Royal Statistical Society B*, **34**, 187-202.
- Dickson, E. R., Grambsch, P. M., Fleming, T. R., Fisher, L. D. and Langworthy, A. (1989). "Prognosis in Primary Biliary Cirrhosis: Model for Decision Making," *The New England Journal of Medicine*, **10**, 1-7.
- Fleming, T. R., and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*, New York: Wiley.
- Levene, H. (1960). Robust tests for equality of variances, In *Contributions to probability and Statistics*, (Olkin, I., et al eds.), 278-292. Stanford University Press.
- Lin, D. Y., Wei, L. J. and Ying, Z (1992). "Checking the Cox model with cumulative sums of martingale-based residuals," *Technical Report No. 111*. Department of Biostatistics, University of Washington.
- Loh, W-Y. (1991). "Survival modeling through recursive stratification," *Computational Statistics and Data Analysis*, **12**, 295-313.
- Miller, R. G. (1976). "Least squares regression with censored data," *Biometrika*, **63**, 449-464.
- Miller, R. G. and Halpern, J. (1982). "Regression with censored data," *Biometrika*, **69**, 521-531.
- Nelson, W. B. (1972). "Graphical analysis of accelerated life test data with the inverse power law model," *IEEE Trans. Reliability*, **R21**, 2-11.
- Peto, R., and Lee, P. (1973). "Weibull distributions for continuous carcinogenesis experiments," *Biometrics*, **29**, 457-470.
- Pike, M. C. (1966). "A method of analysis of a certain class of experiments in carcinogenesis," *Biometrics*, **22**, 142-161.
- Wei, L. J., Ying, Z. and Lin, D. Y. (1990). "Linear regression analysis of censored survival data based on rank tests," *Biometrika*, **77**, 845-851.

FIG. 1: (a) Scatterplot of  $\log_{10}$  survival time (days) versus age at transplant (years) with a smooth curve for 152 Stanford heart transplant patients who survived at least 10 days. (b) Scatterplot of survival time versus age at transplant with a smooth curve for 157 Stanford heart transplant patients.

FIG. 2: Weibull regression tree with the first bootstrap and M methods for the heart transplant data. The numbers within circles or squares are sample sizes. The  $P$ -value beside each node refers to the maximum of the  $t$ -test for the means and Levene's test. One case with zero survival time was deleted.

FIG. 3: Kaplan-Meier survival curves for the terminal nodes of the tree in FIG. 2.

FIG. 4: Weibull regression tree with the second and R methods for the heart transplant data. The numbers within circles or squares are sample sizes. The  $P$ -value beside each node refers to the maximum of the  $t$ -test for the means and Levene's test. One case with zero survival time was deleted.

FIG. 5: Kaplan-Meier survival curves for the terminal nodes of the tree in FIG. 4.

FIG. 6: Weibull regression tree with the third bootstrap and R methods for the liver transplant data. The numbers within circles or squares are sample sizes. The  $P$ -value beside each node refers to the maximum of the  $t$ -test for the means and Levene's test.

FIG. 7: Kaplan-Meier survival curves for the terminal nodes of the tree in FIG. 6.