Wood Property Relationships and Survival Models in Reliability

by

Yan Cheng

B.Econ., Hunan University, 2000
B.Sc., The University of British Columbia, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Statistics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2010

Abstract

It has been a topic of great interest in wood engineering to understand the relationships between the different strength properties of lumber and the relationships between the strength properties and covariates such as visual grading characteristics. In our mechanical wood strength tests, each piece fails (breaks) after surviving a continuously increasing load to a level. The response of the test is the wood strength property – **load-to-failure**[13], which is in a very different context from the standard **time-to-failure**[16] data in Biostatistics. This topic is also called *reliability analysis*[13] in engineering.

In order to describe the relationships among strength properties, we develop joint and conditional survival functions by both a parametric method and a nonparametric approach. However, each piece of lumber can only be tested to destruction with one method, which makes modeling these joint strengths distributions challenging. In the past, this kind of problem has been solved by subjectively matching pieces of lumber, but the quality of this approach is then an issue.

We apply the methodologies in survival analysis to the wood strength data collected in the FPInnovations (FPI) laboratory. The objective of the analysis is to build a predictive model that relates the strength properties to the recorded characteristics (i.e. a survival model in reliability). Our conclusion is that a type of wood defect (knot), a lumber grade status (off-grade: Yes/No) and a lumber's module of elasticity (moe) have statistically significant effects on wood strength. These significant covariates can be used to match pieces of lumber. This paper also supports use of the accelerated failure time (AFT) model[12] as an alternative to the Cox proportional hazard (Cox PH) model[16] in the analysis of survival data. Moreover, we conclude that the Weibull AFT model provides a much better fit than the Cox PH model in our data set with a satisfying predictive accuracy.

Table of Contents

List of Tables

List of Figures

Acknowledgements

This thesis grew out of a research project provided by my supervisor Dr. Jim Zidek. I am deeply indebted to Dr. Zidek for his excellent guidance and immensely help without which the development of this thesis would not have been possible. I am sincerely grateful to my co-supervisor, Dr. Lang Wu, for his invaluable advice and constant patience. They have inspired me a lot in survival analysis with their expertise, support and encouragement, and broadened my knowledge beyond textbooks. They are undoubtedly two of the best professors that I have ever met throughtout my stay in UBC.

Also, I thank Dr. John Petkau for his invaluable advice on my consulting projects, which will benefit me a lot in the future. I thank all the faulty, staff and graduate students in the Department of Statistics for making such a wonderful study environment.

Last but not least, I am most thankful for the love, confidence and support from my parents. My huge thanks goes to my beloved husband and daughter, Yu Liu and Irene Qian Liu, for their love, tremendous support and encouragement that are always beside me.

Chapter 1

Introduction

1.1 Background

Survival analysis[16] is a collection of statistical techniques used to describe and quantify time to event data. The methodological developments with the most profound impact are the Kaplan-Meier method for estimating the survival function, the log-rank test[16] for comparing the equality of two or more survival distributions, and the Cox proportional hazards (PH) model[16] for examining the covariate effects on the hazard function. The accelerated failure time (AFT) model[16] was also proposed but less widely used. In this report, we present the basic concepts, parametric methods (univariate and bivariate Weibull distribution), nonparametric methods (the Kaplan-Meier method and the log-rank test), a semi-parametric model (the Cox PH model) and a parametric model (the AFT model) for analyzing survival data.

1.2 Weibull Distribution

Results of mechanical tests on lumber, wood composites, and wood structures are often summarized by a distribution function fit to data. The Weibull distribution (named after Waloddi Weibull, a Swedish physicist who used it in 1939 to describe the breaking strength of material) is playing an increasingly important role in this type of research and has become a part of several American Society of Testing and Materials standards. Due to one of the parameters - the shape parameter - which allows it to be like a variety of other distributions, such as the normal, lognormal, and exponential distributions, it is very popular with researchers. Its flexibility to model experimental results makes the Weibull distribution a powerful tool in wood utilization research.

The three-parameter Weibull distribution[10] is commonly used to characterize lumber strength. The density function of the Weibull is

$$f(x;\kappa,\lambda,\theta) = \frac{\kappa}{\lambda}(\frac{x-\theta}{\lambda})^{\kappa-1}\exp[-(\frac{x-\theta}{\lambda})^{\kappa}], \tag{1.1}$$

where $x \geq \theta$, $\kappa > 0$ is the shape, $\lambda > 0$ is the scale, and θ is the location.

The distribution function of the Weibull is given by

$$F(x;\kappa,\lambda,\theta) = 1 - \exp[-(\frac{x-\theta}{\lambda})^{\kappa}]. \tag{1.2}$$

Methods must be available to fit the distribution to a data set and provide statistically sound estimates of the parameters of the distribution. However, the effect that different ways of estimating a parameter has on estimating lower tail percentiles has not been widely researched. Fortunately, this limitation of using the Weibull distribution to estimate lumber properties does not affect our case since our data set is complete.

1.3 Kaplan-Meier Method

The Kaplan-Meier[16] estimator of survival is a nonparametric method of inference concerning the survivor function $S = P_r(Y > y)$. Let $y_{(i)}$ denote the ith distinct ordered observation and be the right endpoint of the interval I_i, $i = 1, 2, ..., n$. Also, let $n_i = \#$ unbroken just before the level $y_{(i)}$, while $d_i = \#$ broken at the level $y_{(i)}$. The **K–M estimator of the survivor function** is then

$$\widehat{S}(y) = \prod_{i=1}^{k}(\frac{n_i - d_i}{n_i}),$$

where $y_{(k)} \leq y < y_{(k+1)}$.

Compared to the parametric method, probability statements obtained from most nonparametric statistics are exact probabilities, regardless of the shape of the population distribution from which the random sample was drawn. However, the nonparametric method has several shortcomings such as low power and lack of software. Fortunately, there is a R function called by **survfit** which can calculate the K–M survival estimators.

1.4 Cox Proportional Hazards Model

Let Y represent survival load and the survival function be $S(y) = P_r(Y > y)$. One representation of the distribution of survival load is the hazard function, which represents the instantaneous risk of breaking at the load level y, conditional on survival to that time

$$h(y) = \lim_{\triangle y \to 0} \frac{P_r[(y \leq Y < y + \Delta y)|Y \geq y]}{\Delta y}.$$

Models for survival data usually employs the hazard function or the log hazard. Survival analysis typically examines the relationship of the survival distribution to covariates. Most commonly, this examination entails the specification of a linear-like model for the log hazard. For example, a parametric model based on the exponential distribution may be written as

$$\log h_i(y) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik},$$

or equivalently,

$$h_i(y) = \exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}),$$

that is, as a linear model for the log-hazard or as a multiplicative model for the hazard. Here, i is a subscript for observation, and the x's are the covariates. The constant α in this model represents a kind of log-baseline hazard, since $\log h_i(y) = \alpha$ (or $h_i(y) = e^{\alpha}$) when all of the x's are zero. The baseline hazard function $\alpha(y) = \log h_0(y)$ is unspecified, so the Cox PH model is

$$\log h_i(y) = \alpha(y) + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik},$$

or again equivalently,

$$h_i(y) = h_0(y) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}).$$

This model is semi-parametric because while the baseline hazard can take any form, the covariates enter the model linearly. Consider, now, two observations i and j that differ in their x-values, with the corresponding linear predictors

$$\theta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}$$

and

$$\theta_j = \beta_1 x_{j1} + \beta_2 x_{j2} + ... + \beta_k x_{jk}$$

The hazard ratio for these two observations,

$$\frac{h_i(y)}{h_j(y)} = \frac{h_0(y)e^{\theta_i}}{h_0(y)e^{\theta_j}} = \frac{e^{\theta_i}}{e^{\theta_j}}$$

is independent of the load y. This defines the "proportional hazards property". The general rule is that if the hazard functions cross over load, the PH assumption is violated.

We are not making assumptions about the form of $h_0(y)$ (the nonparametric part of model)– the shape of underlying hazard. Parameter estimates are interpreted the same way as in parametric models, except that no shape parameter is estimated.

Even though the baseline hazard is is not specified, we can still get a good estimate for regression coefficients β, hazard ratio, and adjusted hazard curves. The beauty of the Cox approach is that this vagueness creates no problems for such critical estimations.

1.5 Accelerated Failure Time Model

The accelerated failure time model is an alternative to the Cox PH model for the survival time data. Under AFT models we measure the direct effect of the predictor variables on the survival time instead of the hazard as in the Cox PH model. This characteristic provides an easier interpretation of the results since the parameters measure the effect of the corresponding covariate on the mean survival time. As with the Cox PH model, the AFT model describes the relationship between survival probabilities and covariates.

Given a set of covariates $(X_1, X_2, ..., X_p)$, the model is $S(y) = S_0(\frac{y}{\eta(x)})$, where $S_0(y)$ is the baseline survival function and $\eta(x) = \exp(\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_p x_p)$, an ‘acceleration factor’ that is a ratio of survival times corresponding to any fixed value of $S(y)$.

Under an accelerated failure time model, the covariate effects are assumed to be constant and multiplicative on the time scale, that is, the covariate impacts on survival by a constant factor (acceleration factor).

Based on the relationship between the survival function and hazard function, the hazard function for an individual with covariates $X_1, X_2, ..., X_p$ is given by:

$$h(y) = \frac{1}{\eta(x)} h_0(\frac{y}{\eta(x)}).$$

The corresponding log-linear form of the AFT model with respect to load Y is given by:

$$\log Y_i = \mu + \alpha_1 X_{1i} + \alpha_2 X_{2i} + ... + \alpha_p X_{pi} + \sigma \varepsilon_i,$$

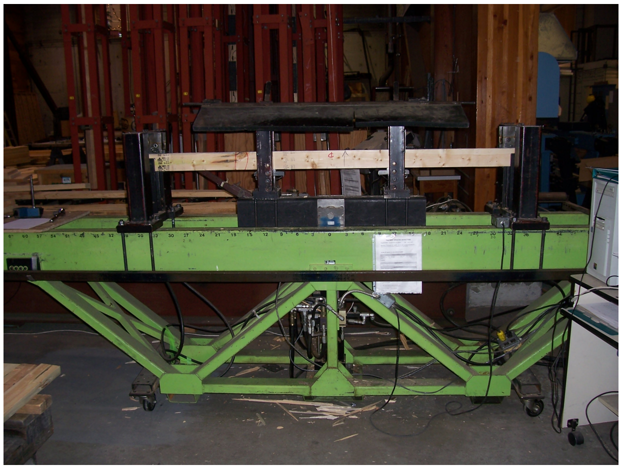
where μ is the intercept, σ is the scale parameter and ε_i is a random variable assumed with a specified distribution. For each distribution of ε_i, there is a corresponding distribution for Y. The AFT models are named for the distribution of Y rather than the distribution of ε_i or $\log Y$.

Chapter 2

Data Description

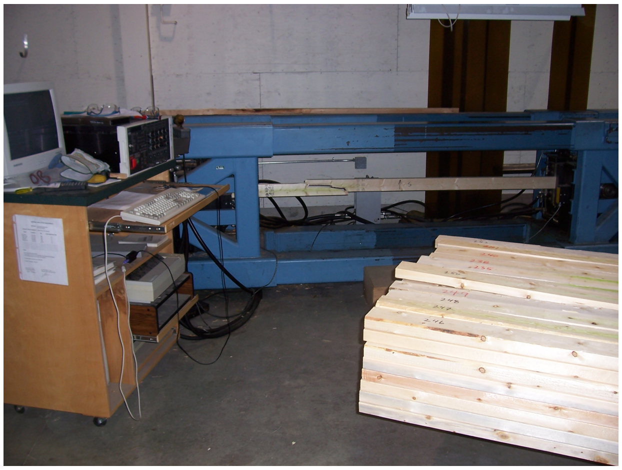
The data come from tests conducted at a FPI/Forintek laboratory. We have two samples of lumber, each of size 98. We applied the bending (R) strength test to generate one sample and the tension (T) strength test to generate the other. In these two tests, as loads (bending or tension stress) increased, each piece will remain intact ("survive") for a while until it breaks. The values of MOR and MOT are recorded (unit: psi 10^3) at the point where the stress is applied (usually at a random location near the center). The break occurs somewhere else along the board. Figure 2.1 and 2.2 show how a piece of lumber is broken in these two tests.

Figure 2.1: The bending test.



This is a transformed **time-to-failure**(**load-to-failure**) problem, and it is very typical in survival analysis. Stiffness or elasticity (E) is measured in both of the above two tests to give the values of MOE (unit: psi 10^6). As each piece of lumber can only be broken once, we only have MOE and MOR in the bending data, while in the tension data we only have MOE and

Figure 2.2: The tension test.



MOT. Interest lies in the relationships amongst MOR, MOT and MOE.

As each piece of lumber is tested, the characteristic deemed most likely to cause the lumber's failure during the test - maximum strength reducing characteristic (MSRC) - is recorded in coded form. Examples of such characteristics are "knot", "grain", "shake" and "split". The MSRC is the grader's best guess *before* testing the board as to why it will fail. The failure code (FC) is the characteristic visually judged by the grader to have caused the piece to fail *after* testing. They could be the same if the failure occurs because of the MSRC. The association between them will be explored later.

There are 10 different causes of failure recorded in the data set, including "knot combination", "grain", "shake" and "split", while around 80% of defects in MSRC and FC are due to "knot" (including both a single knot and a combination of knots). We have the data set available in the form of an excel spreadsheet. The coding system of measurements[3, 14, 15] (e.g. MSRC) is quite complicated as shown in Table 2.1.

Table 2.1: Description of failures for dimension lumber.

Code	Cause of Failure	Code	Cause of Failure
10	knot combination (pith present)	nn	% of cross-section displaced by knot (total)
20	knot combination (no pith)	nn	% of cross-section displaced by knot (total)
23	knot cluster (pith present)	nn	% of cross-section displaced
24	slope of grain (wide face)	nn	actual slope
25	grain deviation	nn	% of cross-section where deflection is greater than 1:4
26	cross grain (narrow face)	nn	actual slope
27	shake and checks	01	not through and less than 2' long
		02	not through and more than 2' long
		03	through and less than 2' long
		04	through and more than 2' long
		05	shake breaks less than 2/3 the edge
		06	shake breaks more than 2/3 the edge
28	split	nn	average length of both sides
35	bark pocket		
45	machine damage	01	saw cut through edge
		02	all other saw cuts
		03	mechanical damage at edge
		04	all other mechanical damage

For the single knot coding system, knots are allowed to be coded numerically with respect to size, orientation and location in the member of cross-section. All possible knot configurations have been incorporated into 10 ”knot classes”. For knot classes 1 through 9, the first digit designates the knot location on either the tension (0) or the compression (1) edge in bending tests. The second digit identifies the knot class (1-9). The next 4 to 8 digits are used for the required knot measurements. When the first two digits are 10, it indicates a knot class 10 and up to three sub-knots (starting from the largest) that can be individually coded with a 10 followed by the 10-digit knot code.

As an example, in Table 2.2, for the 1st piece of lumber, a knot class 1 is considered to be MSRC. For the 2nd piece, a knot class 8 and a knot class 4 are considered to be MSRC 1 and MSRC 2. For the 3rd piece, a knot class 10 is considered to be the MSRC and up to three sub-knots (starting from the largest) are individually recorded as MSRC 1, MSRC 2 and MSRC 3. Here, the MSRC 1 is regarded as the most severe one.

Table 2.2: An example of coded single knot for three pieces of lumber.

Lumber	MSRC1	MSRC2	MSRC3
1	0107001300		
2	1810151104	1413002200	
3	100810062710	100314152705	101314092920

In addition to the defects in MSRC and FC coded in the excel spreadsheet, we also have the corresponding location of MSRC coded. Location is a four-digit code describing the location of the defect or failure within the piece. The first digit indicates whether the defect or failure is located on the tension edge (0), compression edge (1), or both edges (2). The next three digits give the average location of the defect or failure along the length of the piece. As an example, look at Table 2.3.

The random number location (RNL) is the number of inches from the centre of the test span to the worst MSRC (e.g. MSRC 1) - a random integer from 0 to 36. For most of our tests, the MSRC must be randomly located in the test span, and the test span is always less than the length of the lumber.

In summary, we have two samples, MOR and MOT, each of size 98. For each specimen of these two samples, we have the following recorded variables as shown in Table 2.4.

Table 2.5 presents the original layout of bending data in the spreadsheet.

Table 2.3: An example of coded location of MSRC for four pieces of lumber.

Lumber	MSRC1	MSRC2	MSRC3	Loc1	Loc2	Loc3
1	0808131202			0025		
2	1810151104	1413002200		1057	0058	
3	101909142903	101315092722	100309122713	1042	1042	0042
4	2407	2705		2050	0052	

Table 2.4: Description of variables in both samples.

Variables	Descriptions
MOR/ MOT	Module of bending or tension (Load to break)
MOE	Module of elasticity
MSRC	MSRC(1-3 measures with 2 to 12 digits)
MLoc	locations of 3 MSRC (1-3 measures with 4 digits)
FC	Failure characteristic (1-3 measures with 2 to 12 digits)
Floc	locations of 3 FC (1-3 measures with 4 digits)
RNL	Random number location (2 digits from 0 to 36)
Off-grade	Indicator of off-grade piece (1= yes, 0= no)
Species	1 = Spruce, 2 = Pine, 3 = Fir
Moisture	Degree of moisture

Table 2.5: Original bending data.

♯	MSRC1	MSRC2	MSRC3	MLoc1	MLoc2	MLoc3	speci	mois	offg	moe	mor
1	0108131202			0025			2	14.8	0	1.65	6.04
2	1810151104	1413002200		1057	0058		2	13.7	0	1.44	6.59
3	101909142903	101315092722	100309122713	1042	1042	0042	2	15.5	0	1.43	7.46
4	2407	2705		2050	0052		2	14.4	0	1.58	8.95
5	101320172602	100904093015		1043	1028		2	13.6	0	1.36	3.09
6	1014			2111			2	15.7	0	1.46	8.74
7	101912103404			1068			2	15.4	0	1.83	9.94
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.

At this stage, to convert MSRC into meaningful covariates, I only look at MSRC1 as it is regarded as the most severe defect. Also, the first two digits of data strings in MSRC1 capture most of information of defect categories. Based on the "Forintek Knot and Failure Code" descriptions, it is reasonable to classify MSRC1 into 2 variables - knot and size of knot (ksize).

To specify the categorical variable–"knot", we take the first two digits of the MSRC1 data string as they capture most of relevant information on defects:

1. If the first two digits belong to $(0, 9] \bigcup [11, 20)$, knot $= 1$(a single knot);
2. If the first two digits are equal to $10 \bigcup 20 \bigcup 23$, knot $= 2$(a knot combination);
3. Otherwise, knot $= 0$(defects other than knot).

To quantify the numerical variable – "ksize":

1. The value of ksize for a single knot, class 20 or class 23 knot combination is given by the 3rd and 4th digits of MSRC1 data string.
2. The value of ksize for a class 10 knot combination is mainly given by the 5th and 6th digits, or 3rd and 4th digits in some few cases.
3. The value of ksize for other defect is 0.

Therefore, for bending data, we have variables defined as in Table 2.6, and the layout of bending data with transformed covariates is in Table 2.7.

Table 2.6: Variables definition for the transformed bending data.

Variables	Descriptions
knot	1=a single knot, 2=a knot combination, 0=other
ksize	the size of knot or 0 for non-knot defects
rnl	random number for location of MSRC
Off-grade	Indicator of off-grade piece (1= yes, 0= no)
loc	location of defect
face	edge of defect:0=tension, 1=compression, 2=both
Species	1 = Spruce, 2 = Pine, 3 = Fir
Moisture	Degree of moisture
moe	module of elasticity
mor	module of rupture

Table 2.7: Transformed bending data.

Specimen	knot	ksize	rnl	offg	loc	face	species	moisture	moe	mor
1	1	12	7	0	44	0	2	14.8	1.65	6.0424
3	0	0	22	1	22	2	2	13.7	1.44	6.5902
5	1	101	9	0	29	1	2	15.5	1.43	7.4588
9	1	9	5	0	13	0	2	14.4	1.58	8.9549
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.

Chapter 3

Exploratory Analysis and Preliminary Conclusions

3.1 Introduction

Exploratory data analysis (EDA) is detective work. It comprises techniques to visualize patterns in data.

3.2 Graphical Presentation of Strength Properties Data

3.2.1 Histogram of Strength Data

For bending and tension tests, let's first explore the shape of distributions of the strength properties data: MOR, MOT and MOE in both tests. Based on their histograms and density curves in Figure 3.1, we see that all of the distributions are asymmetrical and in fact right-skewed, which is very typical for survival data. Moreover, the two density curves of MOE from the two tests seem to be identical, and the side-by-side boxplots of MOE in these two tests are almost overlapped. This indicates that there may be no significant difference between the two MOE's in the two tests.

3.2.2 Exploring the Relationship Between the Strength Data and Covariates

We next explore the relationships between the strength data and all other variables. With bending test data, we classify these variables into continuous and categorical. Then, we use scatterplots and side-by-side boxplots to visualize the relationships between MOR and these two types of variables respectively.

Figure 3.2 displays MOR against continuous variables. The non-parametric curve using lowess shows the pattern of association between the MOR and

Figure 3.1: Distributions of the strength properties data.

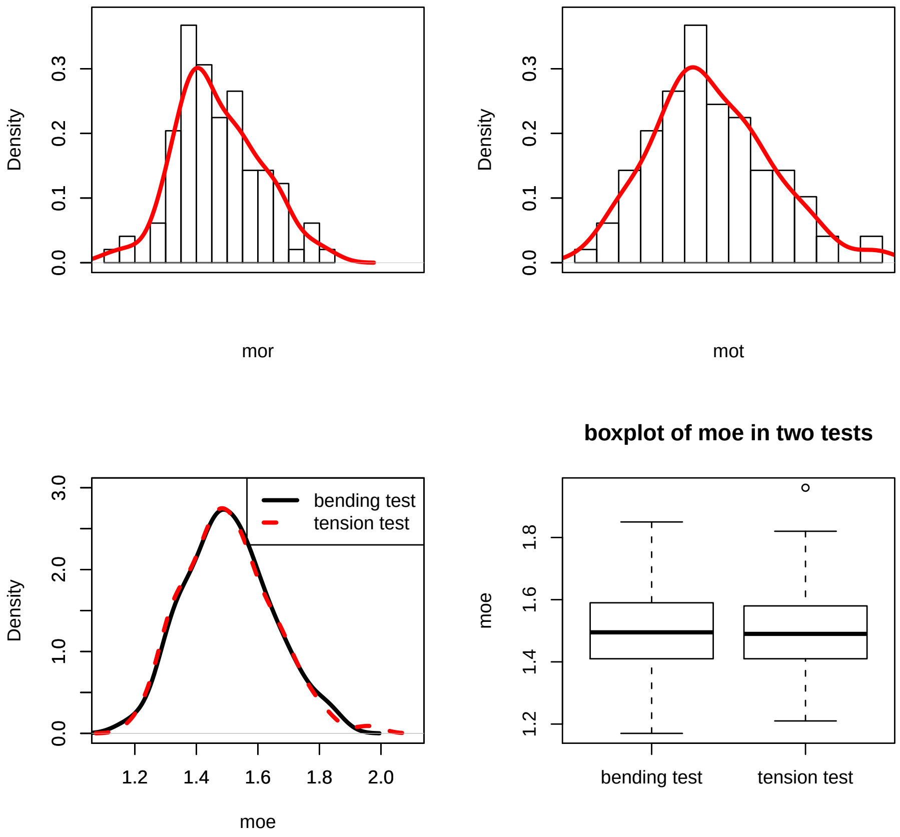
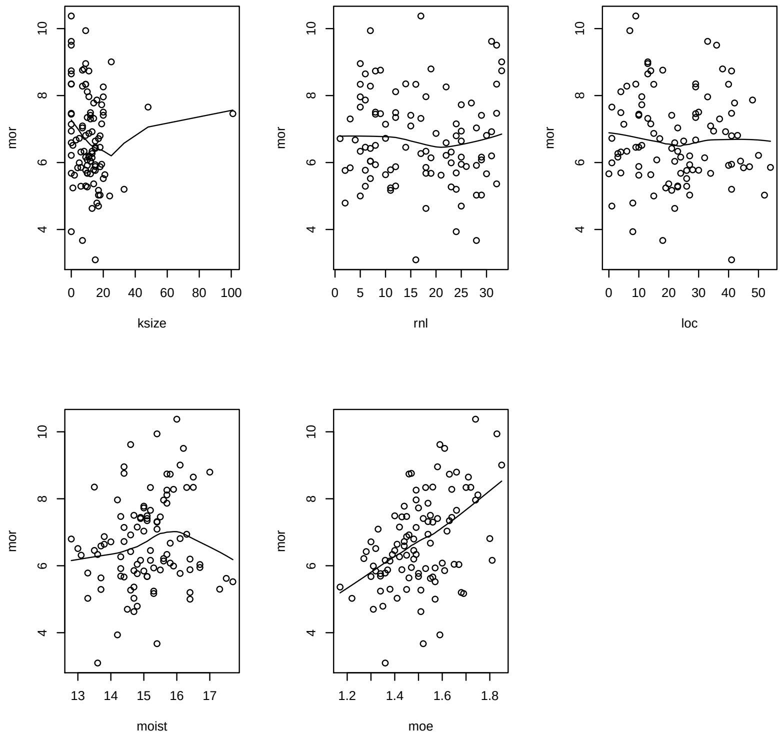


Figure 3.2: MOR against continuous variables, with a lowess smooth curve.

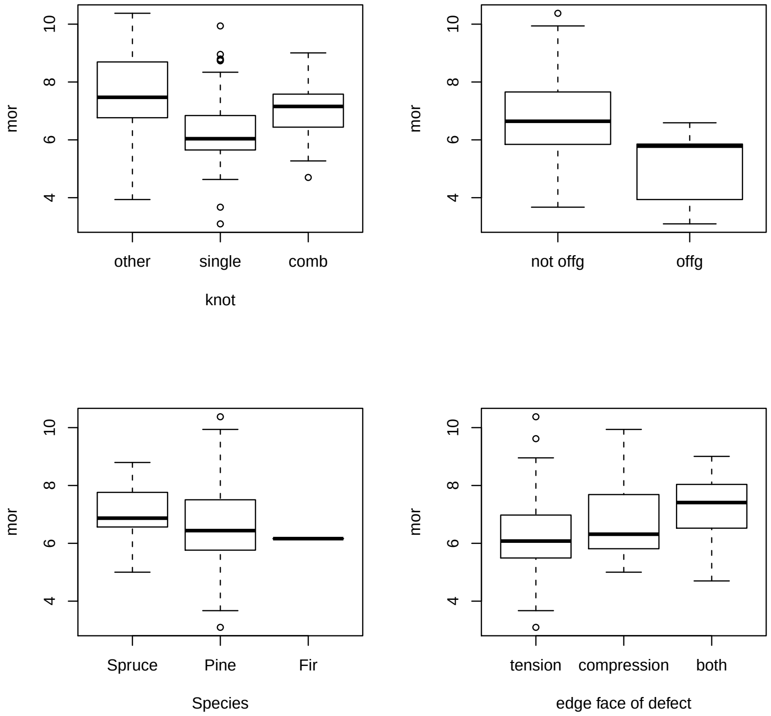


other variables in pairs. We see that there is a positive association between MOR and MOE, but no specific patterns for MOR and other variables.

Figure 3.3 shows the side-by-side boxplots of MOR against the categorical variables – “knot”,“offg”,“species” and “face”. It shows that a piece of lumber with a “single knot” as MSRC1, “off-grade”, “pine” species, or the defect is on the tension edge will produce a relatively lower MOR.

The tension test data display the same patterns as the bending test data in terms of associations between the strength property MOT and other variables. From the plots above, we can see that distributions of strength properties are very typical for survival data. Thus, to model their distributions, we may consider both a parametric approach (e.g. Weibull distribution) and a non-parametric method (e.g. Kaplan-Meier estimator).

Figure 3.3: MOR against categorical variables.



3.3 Univariate Approaches to Modeling the Distributions

3.3.1 Introduction

Interests lies in the relationships between the MOR, MOT and MOE. Moreover, in lumber strength testing, people are usually interested in the weakest boards (e.g. the population strength 5th percentile– $\zeta_{0.05}$ or lower). Next, I will use both parametric and nonparametric approaches to estimate the $\zeta_{0.05}$ for each type of strength as well as their ratio. Using the population 5th percentiles for MOR and MOT as an example, the ratio is $\rho = \zeta^R_{0.05}/\zeta^T_{0.05}$.

3.3.2 Univariate Weibull Distribution

Assuming Weibull population distributions and independent samples, the three parameters in (1.1) can be estimated using maximum likelihood via numerical optimization in R.

Let $(\kappa_i, \lambda_i, \theta_i)$, $i = 1, 2$, be the true parameters for two independent 3-parameter Weibull distribution populations, and $(\widehat{\kappa}_i, \widehat{\lambda}_i, \widehat{\theta}_i)$, $i = 1, 2$, be the corresponding maximum likelihood estimates from two samples, where $\kappa_i > 0$ is the shape, $\lambda_i > 0$ is the scale, and θ_i is the location.

Table 3.1 displays the maximum likelihood estimates (MLEs) of parameters in the three parametric Weibull distribution for the **MOR** data. where $\lambda(psi \times 10^3)$, κ(unitless) and $\theta(psi \times 10^3)$:

Table 3.1: MLEs of univariate Weibull parameters for the bending data.

Quantity	Value	Standard Error
$\widehat{\lambda}_1$	4.726	0.590
$\widehat{\kappa}_1$	3.325	0.511
$\widehat{\theta}_1$	2.460	0.537

Similarly, Table 3.2 displays the maximum likelihood estimates (MLEs) of parameters in the three parametric Weibull distribution for the **MOT** data. where $\lambda(psi \times 10^3)$, κ(unitless) and $\theta(psi \times 10^3)$:

Since the distribution function of the Weibull is given by (1.1),

$$F(x; \kappa, \lambda, \theta) = P(X \leq x) = 1 - \exp[-(\frac{x - \theta}{\lambda})^\kappa],$$

Table 3.2: MLEs of univariate Weibull parameters for the tension data.

Quantity	Value	Standard Error
$\widehat{\lambda}_2$	3.610	0.362
$\widehat{\kappa}_2$	2.556	0.335
$\widehat{\theta}_2$	0.901	0.297

the population 5th percentile $\zeta_{0.05}$ is then given by

$$P(X \leq \zeta_{0.05}) = 0.05 = 1 - \exp[-(\frac{\zeta_{0.05} - \theta}{\lambda})^{\kappa}].$$

Solving this equation we get

$$\zeta_{0.05} = \lambda[-\ln(0.95)]^{\frac{1}{\kappa}} + \theta.$$

Thus, we can easily get the ratio given by

$$\rho = \frac{\zeta^R_{0.05}}{\zeta^T_{0.05}} = \frac{\lambda_1[-\ln(0.95)]^{\frac{1}{\kappa_1}} + \theta_1}{\lambda_2[-\ln(0.95)]^{\frac{1}{\kappa_2}} + \theta_2}.$$

By the invariance property of MLEs, we can obtain the corresponding MLEs of $\widehat{\zeta^R_{0.05}}$, $\widehat{\zeta^T_{0.05}}$ and $\widehat{\rho}$ can be calculated by substituting $(\widehat{\kappa}_i, \widehat{\lambda}_i, \widehat{\theta}_i)_{i=1,2}$ in Table 3.3.

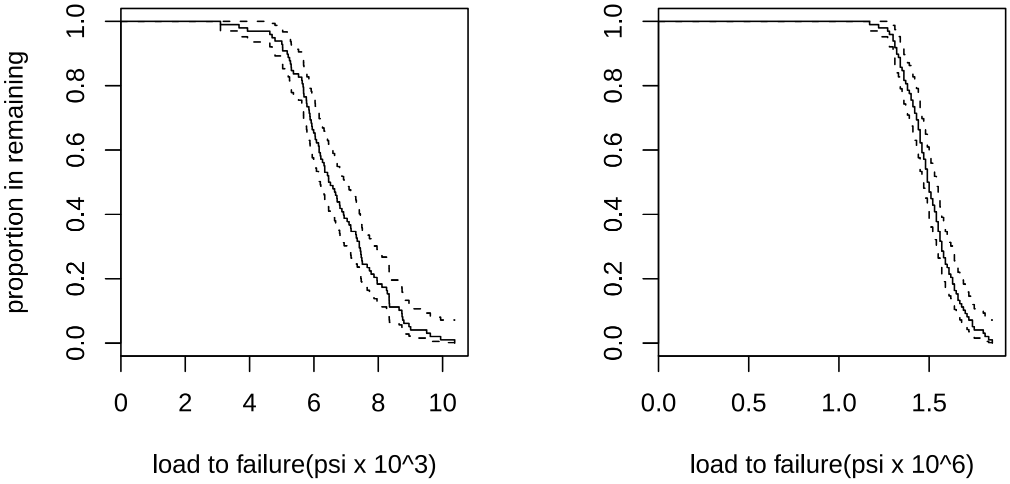
Table 3.3: MLEs of $\widehat{\zeta^R_{0.05}}$, $\widehat{\zeta^T_{0.05}}$ and $\widehat{\rho}$.

Quantity	Value	Standard Error	95% Confidence Interval
$\widehat{\zeta^R_{0.05}}$	$4.394(psi \times 10^3)$	$0.180(psi \times 10^3)$	(4.041 , 4.747) $(psi \times 10^3)$
$\widehat{\zeta^T_{0.05}}$	$2.030(psi \times 10^3)$	$0.137(psi \times 10^3)$	(1.761 , 2.299) $(psi \times 10^3)$
$\widehat{\rho}$	2.164	0.171	(1.829 , 2.499)

3.3.3 Univariate Kaplan–Meier Estimator

To explore the distribution of MOR, MOE and MOT, we could also use the non-parametric Kaplan-Meier estimators of their survival functions $S(y) =$

Figure 3.4: KM curves for MOR and MOE in the bending test.



$P_r(Y > y)$. Using the R function **survfit**, we plot the Kaplan-Meier curves of MOR (left) and MOE (right) with 95% error bands in Figure 3.4.

As well as the KM estimators of $P_r(MOR > mor)$ in Table 3.4, where n_i= # at risk before mor_i, d_i= # that break at mor_i.

Table 3.4: KM estimation of survival function for MOR.

mor_i	n_i	d_i	$P_r(MOR > mor_i)$	std.err	lower 95% CI	upper 95% CI
3.09	98	1	0.9898	0.0102	0.97010	1.0000
3.67	97	1	0.9796	0.0143	0.95199	1.0000
3.94	96	1	0.9694	0.0174	0.93587	1.0000
4.63	95	1	0.9592	0.0200	0.92080	0.9992
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

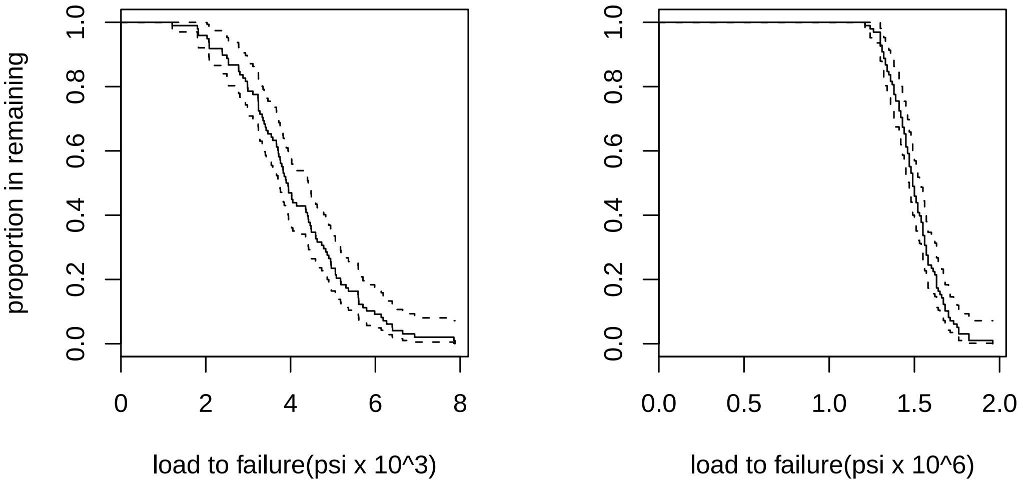
Similarly, the KM curves for MOT (left) and MOE (right) with 95% error bands in the tension test are shown in Figure 3.5.

as well as the KM estimators of $P_r(MOT > mot)$ in Table 3.5, where n_i= # at risk before mot_i, d_i= # that break at mot_i.

Table 3.5: KM estimation of survival function for MOT.

mot_i	n_i	d_i	$P_r(MOT > mot_i)$	std.err	lower 95% CI	upper 95% CI
1.21	98	1	0.9898	0.0102	0.97010	1.0000
1.80	97	1	0.9796	0.0143	0.95199	1.0000
1.82	96	1	0.9694	0.0174	0.93587	1.0000
1.83	95	1	0.9592	0.0200	0.92080	0.9992
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Figure 3.5: KM curves for MOT and MOE in the tension test.



3.3.4 The 5th Percentile Estimators by the KM Approach

It is very handy to use the KM method to get the 5th percentile estimates ($\widehat{\zeta^R_{0.05}}$, $\widehat{\zeta^T_{0.05}}$ and $\widehat{\zeta^E_{0.05}}$) using a formula given by Mara and Jong (2004)[16]. Table 3.6 and 3.7 show the KM estimators of percentiles for bending and tension, respectively.

Table 3.6: KM estimators of percentiles in bending test.

Quantity	Value	Standard Error	95% Confidence Interval
$\widehat{\zeta^R_{0.05}}$	$4.70(psi \times 10^3)$	$0.503(psi \times 10^3)$	(3.714 , 5.686) $(psi \times 10^3)$
$\widehat{\zeta^E_{0.05}}$	$1.30(psi \times 10^6)$	$0.0.0395(psi \times 10^6)$	(1.223 , 1.378) $(psi \times 10^6)$

Table 3.7: KM estimators of percentiles in tension test.

Quantity	Value	Standard Error	95% Confidence Interval
$\widehat{\zeta^T_{0.05}}$	$2.03(psi \times 10^3)$	$0.285(psi \times 10^3)$	(1.471 , 2.589) $(psi \times 10^3)$
$\widehat{\zeta^E_{0.05}}$	$1.30(psi \times 10^6)$	$0.0.028(psi \times 10^6)$	(1.245 , 1.355) $(psi \times 10^6)$

It has been shown that the KM estimators are pretty close to the estimators for the Weibull distribution approach in Table 3.3, but the standard errors of KM estimators are relatively larger than the ones by Weibull approach. One reason is the nonparametric method is usually less precise than the parametric one. Also, the two estimated values of $\zeta^E_{0.05}$ for bending and tension are almost the same, which indicates that the lumbers in two different tests might be homogeneous in terms of elasticity.

3.4 Bivariate Approaches to Modeling the Distributions of (R,E) and (T,E)

3.4.1 Bivariate Weibull Distribution

The density function of the bivariate Weibull[12] is:

$$\begin{aligned} f(x, y; \kappa_1, \lambda_1, \theta_1, \kappa_2, \lambda_2, \theta_2, \delta) &= \frac{\kappa_1}{\lambda_1}(\frac{x-\theta_1}{\lambda_1})^{\frac{\kappa_1}{\delta}-1}\frac{\kappa_2}{\lambda_2}(\frac{y-\theta_2}{\lambda_2})^{\frac{\kappa_2}{\delta}-1} \\ &\times\{(\frac{x-\theta_1}{\lambda_1})^{\frac{\kappa_1}{\delta}} + (\frac{y-\theta_2}{\lambda_2})^{\frac{\kappa_2}{\delta}}\}^{\delta-2}\{[(\frac{x-\theta_1}{\lambda_1})^{\frac{\kappa_1}{\delta}} + (\frac{y-\theta_2}{\lambda_2})^{\frac{\kappa_2}{\delta}}]^{\delta} + \frac{1}{\delta} - 1\} \\ &\times \exp\{-[(\frac{x-\theta_1}{\lambda_1})^{\frac{\kappa_1}{\delta}} + (\frac{y-\theta_2}{\lambda_2})^{\frac{\kappa_2}{\delta}}]^{\delta}\} \qquad (3.1) \end{aligned}$$

For estimating of the bivariate Weibull parameters, a feasible method has been developed by Richard, James and David (1999)[2]. We first estimated the shape (κ), scale (λ) and location (θ) parameters from the two marginal distributions, using standard theory for the univariate Weibull. Given these parameter estimates $(\kappa_1, \lambda_1, \theta_1, \kappa_2, \lambda_2, \theta_2)$, we can find the dependence parameter estimate δ using maximum likelihood be numerical optimization in R. We can get the log of the likelihood $\log L$ for a random and uncensored sample, and the MLEs of parameters can be obtained by minimizing $-2\log L$.

A three-parameter Weibull distribution has the survival function,

$$\begin{aligned} \overline{F}(x, y) &= P[X > x, Y > y] \\ &= \exp\{-[(\frac{x-\theta_1}{\lambda_1})^{\frac{\kappa_1}{\delta}} + (\frac{y-\theta_2}{\lambda_2})^{\frac{\kappa_2}{\delta}}]^{\delta}\}, 0 < \delta \leq 1 \qquad (3.2) \end{aligned}$$

Therefore, once the parameters $\kappa_1, \lambda_1, \theta_1, \kappa_2, \lambda_2, \theta_2$ and δ are estimated, we can easily estimate the survival probability for the bivariate data (x, y).

3.4.2 Bivariate KM Estimator

A bivariate version of the KM estimator does exist. To describe it we let $(X_i, Y_i)(i = 1, ..., n)$ be n independent and identically distributed pairs of loads to failure with survival function $F(x, y) = Pr(X \geq x, Y \geq y)$. Since X_i and Y_i are the observed loads, it is natural to estimate $Pr(X \geq x, Y \geq y)$ by the empirical survival function:

$$\widehat{S}(x, y) = n^{-1}\sum_{i=1}^{n} I(X_i \geq x, Y_i \geq y) \qquad (3.3)$$

And the asymptotic variance of this estimator is given by:

$$\widehat{Var}(\widehat{S}(x,y)) = \widehat{S}(x,y) - [\widehat{S}(x,y)]^2.$$

Lin and Ying (1993)[9] provide evidence in favor of this approach. Also, our data are uncensored, which makes our problem much easier than the censored case.

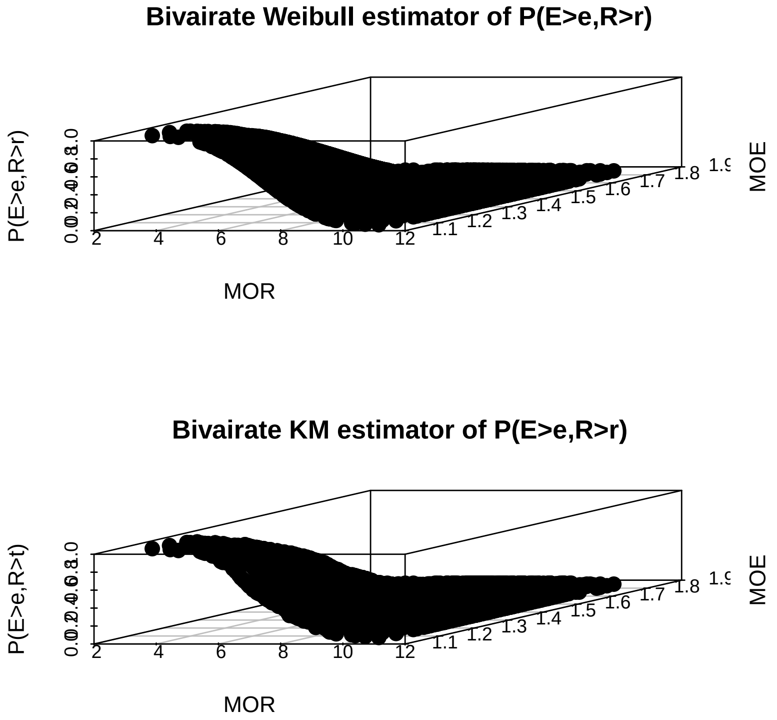
Then, as an example, for the MOR data, we can compare the estimates of the survival function $S(e,r)$ computed with (3.2) and by (3.3) in the following Table 3.8:

Table 3.8: $\widehat{S}(e,r)$ by Bivariate Weibull and by Bivariate KM.

$(e,r)(psi \times 10^6, psi \times 10^3)$	$\widehat{S}(e,r)$ by (4)	$\widehat{S}(e,r)$ by (5)
(1.65, 6.042)	0.1327	0.1330
(1.65, 6.590)	0.1122	0.1146
(1.65, 7.459)	0.1020	0.0762
(1.36, 7.867)	0.2143	0.1960
(1.36, 4.791)	0.8061	0.7952
(1.36, 5.664)	0.7041	0.6840
(1.36, 5.363)	0.7347	0.7318
(1.36, 7.318)	0.3367	0.3112
(1.17, 7.459)	0.2857	0.2982
(1.17, 8.955)	0.0612	0.0558
(1.17, 3.095)	1.0000	0.9974
(1.17, 8.740)	0.0918	0.0757
(1.17, 9.939)	0.0204	0.0100
.	.	.
.	.	.
.	.	.

It seems that these two estimates are pretty close to each other, which confirms that both parametric and nonparametric survival analysis approaches to lumber strength appears to work well. Besides, we could graph the 3-dimensional scatterplot for each method as shown in Figure , and it is obvious that these two estimates are almost the same.

Figure 3.6: Comparison of Bivariate Weibull and KM Estimates.



Bivairate Weibull estimator of P(E>e,R>r)

Bivairate KM estimator of P(E>e,R>r)

3.5 Tests for the Difference of Distributions

3.5.1 Graphical Approach by the KM Estimator and Log-rank Test

A central objective of the study described in this thesis is the relationship between strength and its covariates. For a categorical covariate, we may graph the the KM curves for strength data for different covariate categories, so that we can see if different categories make a difference in the distribution of strength.

With bending data, Figure 3.7 displays the KM curves of “mor” against 4 categorical covariates – “knot”, “offg”, “species” and “face”, respectively. It seems that the KM curves are parallel for “offg” and “knot” (overall - there are slight cross-overs when MOR is either small or large). But they are decidedly nonparallel for “species” and “face”. That is, the differences between KM curves for “knot” and “offg” are relatively larger than the other two covariates.

The KM curves give us an insight into the difference of survival functions in two or more groups, but whether this observed difference is statistically significant requires a formal statistical test. One commonly used non-parametric tests for comparing two or more survival distributions is the log-rank test. The log-rank test compares the observed number of failures with the expected number of failures for each group. The null hypothesis asserts no difference between survival curves in two or more groups.

That test yields p-values of 0.00623 (knot), 0.00215 (offg), 0.749 (species) and 0.312 (face). Therefore, the differences we observed above of MOR survival curves made by “knot” and “offg” are statistically significant, which indicates that “knot” and “off” may be the important predictors for MOR.

3.5.2 Test for the Difference Between Two MOE in the Two Tests

Another topic of interest is that difference between the two MOE population distributions for bending and tension. The two KM curves are sketched in Figure 3.8 and we observe that they are almost identical. Also, by the **log-rank** test, their difference is not statistically significant with a very large p-value 0.995, a finding consistent with the previous conclusion suggested by Figure 3.1 – the two density curves of MOE in the two cases are almost identical.

Figure 3.7: KM curves of MOR against categorical covariates. Notice that unlike the curves for “species” and “face”, those for “offg” and “knot” are quite parallel.

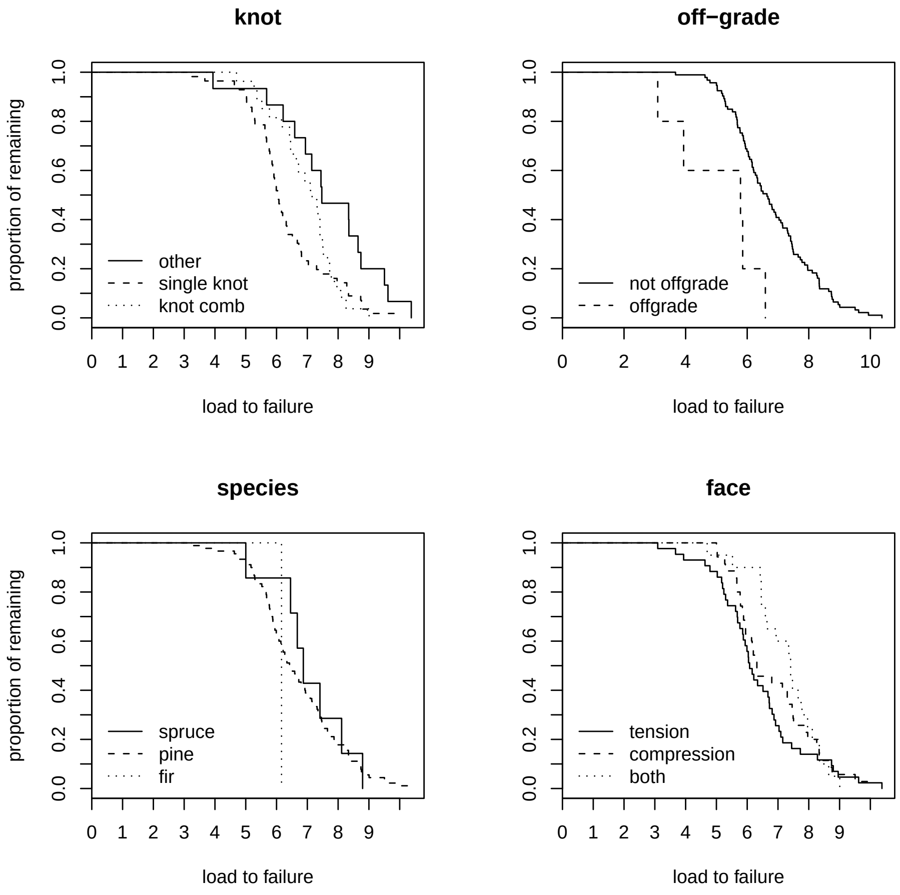
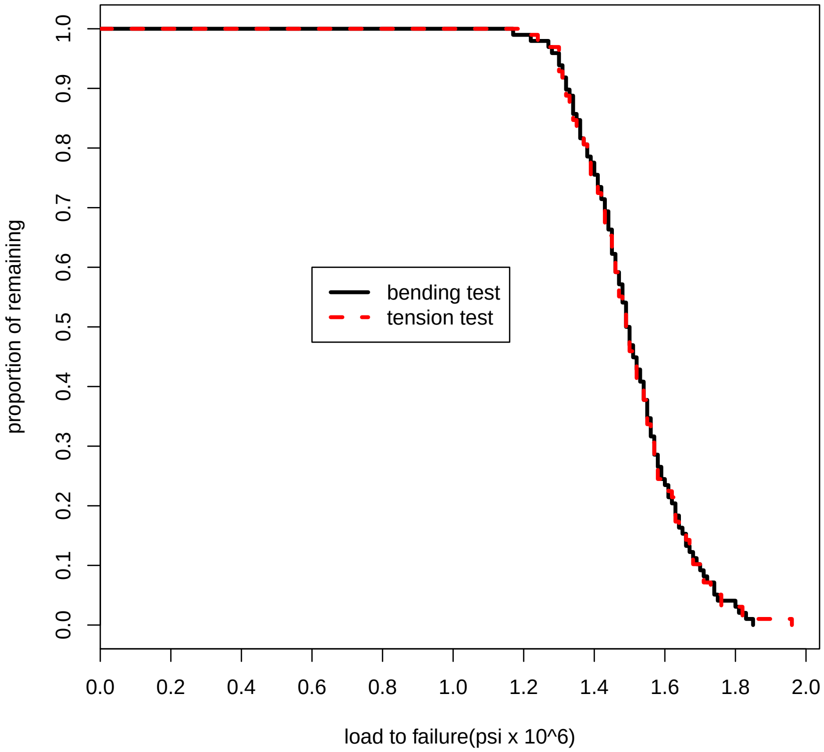


Figure 3.8: KM curves of MOE in the two cases. Notice that the curves for two tests are almost identical.



3.6 Exploring the Association between MSRC and FC

Recall that MSRC means the grader's best guess *before* testing the board as to why it will fail, while FC is the characteristic visually judged by the grader to have caused the piece to fail *after* testing. They could be the same if the failure occurs because of MSRC.

3.6.1 Two-way Contingency Table

If two variables are measured at categorical levels (eg. nominal or ordinal), we assess their relationship by crosstabulating the data in a two-way contingency table[1]. A two-way contingency table is a two-dimensional (rows × columns) table formed by 'cross-classifying' subjects or events on two categorical variables. One variable's categories define the rows while the other variable's categories define the columns. The intersection (crosstabulation) of each row and column forms a cell, which displays the count (frequency) of cases classified as being in the applicable category of both variables. Table 3.9 is a simple example of a hypothetical contingency table that crosstabulates student gender against answer on one question of an exam; a total of 100 students are described.

So, we can set up the 2-way contingency table between MSRC and FC, as shown in Table 3.10, using the first two digits in the characteristic descriptions since they capture the most of the visual information on lumber defects. Note the total of observations is 195 (not 196), since we have one missing datum in the data set.

Table 3.9: Example of A Hypothetical Two-way Contingency Table. Here we see "gender" being broken down by a subject's answer to an examination question (1= "Yes"; 0= "No").

	Answer		
Gender	Yes	No	Total
Male	38	12	50
Female	10	40	50
Total	48	52	100

Table 3.10: Two-way Contingency Table of MSRC and FC.

	FC			
MSRC	01-09	10-19	20-60	Total
01-09	42	14	20	76
10-19	20	41	15	76
20-60	7	11	25	43
Total	69	66	60	195

3.6.2 Test of Independence (Chi-square and Related Tests)

For ease of understanding, let's take the data in Table 3.9 for example. If the characteristics *Gender* and *Answer* were not associated (the null hypothesis of independence), we can easily calculate the expected counts in each cell, i.e., the number of cases we would expect based on their total distribution in the sample. Given that the sample contains exactly 50% male and 50% female, were there no association between *Gender* and *Answer*, we would expect exactly half of those answering 'Yes' (48) to be male, i.e., $48 \div 2 = 24$. The actual formula for computing the expected count (E) in any cell of a contingency table is: $E = (row\,total \times column\,total) \div (grand\ total)$. Thus, for the "Male/Yes" cell, $E = (50 \times 48) \div 100 = 24$.

The larger the difference between the observed (O) and expected (E) cell counts, the less likely that the null hypothesis of independence holds true, i.e., the stronger the evidence that the two variables are related. In our example, the large difference between the observed ($O = 38$) and expected ($E = 24$) cell counts for the Male/Yes cell suggests that being male is associated with greater likelihood of answering 'Yes'.

To determine whether or not the row and column categories for the table as a whole are independent of each other, we compute **Pearson's chi-square statistic** (X^2):

$$X^2 = \sum[\frac{(O-E)^2}{E}]\,,$$

where $O = observed\,frequency$ and $E = expected\,frequency$. As indicated in the formula, one first computes the difference between the observed and expected frequencies in a cell, squares this difference, and then divides the squared difference by that cell's expected frequency. These values are then summed (the $\sum$ symbol) over all the cells, yielding the value of X^2. In our example, $X^2 = 31.41$.

The value of X^2 is then compared to a critical value that is based on the number of rows and columns ($df = degrees\, of\, freedom = (number\, of\, rows - 1) \times (number\, of\, columns - 1)$) and obtained from a chi-square distribution table. If the value of X^2 is less than this critical value, then we cannot reject the null hypothesis and we conclude that the data do not provide evidence of an association. If the value of X^2 exceeds the critical value, then we reject the null hypothesis and conclude that the variable categories are indeed associated.

In our example, $df = 1$ and the chi-square critical value for a significance level of $\alpha = 0.05$ is 3.84. Since our calculated X^2 is 31.41 which clearly exceeds this critical value, we may conclude that gender is associated with answer in the exam.

If the minimum expected count for any cell in a contingency table is less than 5, then the chi-square approximation to the distribution of the X^2 statistic may not be accurate. In this case, an alternative is **Fisher's Exact Test**. If one or more of the expected counts in the cells of a contingency table are less than 5 or when the row or column totals are very uneven, Fisher's exact test is more desirable.

In our real 2-way contingency table,where

H_0: there is no association between MSRD and FC
H_1: there is association between MSRC and FC

our calculated X^2 is 43.9383, and the corresponding p-value is approximately 0, which indicates that we should reject the null hypothesis and in favor of the hypothesis that independence doesn't hold here, there is association between MSRD and FC. The Fisher's exact test also produces a p-value close to 0, which confirms the conclusion of the Chi-square test.

3.6.3 Describing the Strength of Association

If there is an association, it may be desirable to then describe the **strength** of the association. We use correlation-like measures such as the $Phi\ coefficient$ and $Cramer's\ V$ to describe the strength of relationship between nominal variables, since MSRC and FC are measured at nominal level. These coefficients range from 0 to 1 since you cannot have a 'negative' relationship between nominal variables.

The $Phi\ coefficient$ (ϕ) is a measure of nominal association applicable only to 2×2 tables. It is calculated as:

$$\phi = \sqrt{\frac{X^2}{N}}$$

where $X^2 = the\ value\,of\ Pearson's\ \ chi{-}square$, and $N = the\ sample\ size$. In our example, the $Phi\ \ coefficient = \sqrt{\frac{31.41}{100}} = 0.56$, suggesting a moderately strong association.

For contingency tables that are larger than 2×2, $Cramer's\ V$[1] is the choice of nominal association measure. The formula for $Cramer's\ V$ is given by:

$$V = \sqrt{\frac{X^2}{N(k-1)}}$$

where N is the sample size and k is the lesser of the number of rows or columns. Since in 2×2 tables $k = 2$, $Cramer's\ V$ equals the $Phi\ \ coefficient$ for 2×2 tables.

Therefore, since our calculated X^2 is 43.9383, the strength of association between MSRC and FC is $\sqrt{\frac{43.9383}{195(3-1)}} = 0.34$, suggesting a relatively weak association. However, making a low V level is inevitable with such a small data set. If we also include information on MSRC2 and MSRC3 to construct the two-way contingency table, a larger V should be produced in no doubt.

Chapter 4

Semi-parametric Survival Model

4.1 Introduction

Let's first fit a semi-parametric survival regression model - $CoxPH$ model[6, 7, 17]. Since in the $CoxPH$ model, the baseline hazard function $h_0(t)$ is nonparametric and no distributional assumption is needed for the survival data, it is easier to start with it.

As an example, for bending data, we may fit a $CoxPH$ model for MOR with covariates: knot, ksize, random number location(rnl), off-grade indicator(offg), location of defect(loc), face of defect(0=on the tension edge, 1=on the compression edge, 2=on the both edges) , species, moisture and MOE.

4.2 AIC Procedure For Variable Selection

Comparisons between a number of possible models, which need not necessarily be nested nor have the same error distribution, can be made on the basis of the statistic

$$AIC = -2 \times \log(maximumlikelihood) + k \times p,$$

where p is the number of parameters in each model under consideration and k is a predetermined constant. This statistic is called **Akaike's (1974) information criterion (AIC)**; the smaller the value of this statistic, the better the model. This statistic trades off goodness of fit (measured by the maximized log likelihood) against model complexity (measured by p). Here we shall take k as 2.

So, we can rewrite the AIC in the context of the Cox PH model:

$$AIC = -2 \times \log(maximumlikelihood) + 2 \times b,$$

where b is the number of β coefficients in each model under consideration. The maximum likelihood is replaced by the maximum partial likelihood. The smaller the AIC value the better is the model.

4.3 Application to Variable Selection

First, we fit the initial Cox PH model for the bending data using all possible covariates:

$coxph.fit1 < -coxph(Surv(mor) \sim factor(knot) + ksize + rnl +$
$factor(offg) + loc + factor(face) + factor(species) + moist + moe)$

Table 4.1 presents $summary(coxph.fit1)$ as below:

Table 4.1: Summary of the initial Cox PH model.

	coef	exp(coef)	se(coef)	z	p	
factor(knot)1	1.81	6.13	0.42	4.27	0.00	$***$
factor(knot)2	0.92	2.50	0.39	2.32	0.02	$*$
ksize	0.00	0.99	0.00	-0.20	0.84	
rnl	0.02	1.01	0.01	1.36	0.17	
offg	1.70	5.49	0.53	3.16	0.00	$**$
loc	0.00	1.00	0.00	0.14	0.88	
factor(face)1	-0.48	0.62	0.26	-1.81	0.07	
factor(face)2	0.46	1.57	0.35	1.28	0.20	
factor(species)2	0.47	1.59	0.43	1.07	0.28	
factor(species)3	1.25	3.47	1.18	1.04	0.29	
moist	0.22	1.24	0.13	1.56	0.12	
moe	-5.82	0.00	1.06	-5.46	0.00	$***$

Thus, we can see the covariates “ knot”, “off-grade” and “moe” are significant at level of 0.05.

4.3.1 Method I: *step* () to select the best model according to AIC statistic

Table 4.2 shows p-values corresponding to variables selected by $step(coxph.fit1)$.

From Table 4.3, we may see that the stepwise method chooses 3 covariates: **knot**, **off-grade** and **moe**.

4.3.2 Method II: Single term deletions

Table 4.4 displays the result of single term deletions method $drop\,1\,(coxph.fit1, test = “Chi”)$:

Table 4.2: Stepwise model path for the main effects model on the bending data.

Step	Df	AIC
		667.16
- moist	1	667.89
- factor(face)	2	668.45
- factor(offg)	1	672.74
- factor(knot)	2	682.11
- moe	1	698.23

Table 4.3: p-values of covariates in the model selected by *step* ().

	coef	exp(coef)	se(coef)	z	p	
factor(knot)1	1.440	4.22207	0.363	3.97	7.3e-05	$***$
factor(knot)2	0.730	2.07554	0.367	1.99	4.7e-02	$*$
factor(offg)1	1.692	5.42779	0.528	3.20	1.4e-03	$**$
factor(face)1	-0.407	0.66565	0.241	-1.69	9.1e-02	
factor(face)2	0.374	1.45321	0.350	1.07	2.9e-01	
moist	0.214	1.23845	0.130	1.65	9.9e-02	
moe	-5.931	0.00266	1.069	-5.55	2.9e-08	$***$

So, we see that deletion of **knot**, **off-grade** and **moe** will lead to a significant increase in AIC values, which indicates that these 3 variables are likely to be the most important covariates.

4.3.3 Comparing Nested Models

So far, we obtain the same reduced model by **Method I** and **Method II**. Next, we will compare this reduced model to the initial full model. Nested models can be compared using the likelihood ratio test (LRT).

Symbolically we may describe a model as follows:

full model : $coxph.fit1 < -coxph(Surv(mor) \sim factor(knot) + ksize + rnl + factor(offg) + loc + factor(face) + factor(species) + moist + moe)$

Table 4.4: Drop 1 model path for the main effects model on the bending data.

	Df	AIC	LRT	Pr(Chi)	
		673.69			
factor(knot)	2	690.73	21.042	2.697e-05	$***$
ksize	1	671.73	0.044	0.834587	
rnl	1	673.56	1.866	0.171930	
factor(offg)	1	679.24	7.554	0.005987	$**$
loc	1	671.71	0.021	0.883448	
factor(face)	2	676.11	6.417	0.050423	
factor(species)	2	671.32	1.635	0.441493	
moist	1	674.17	2.483	0.115061	
moe	1	704.00	32.315	1.311e-08	$***$

reduced model by method I and II : $cox1 < -coxph(Surv(mor) \sim factor(knot) + factor(offg) + moe)$

$anova(cox1, coxph.fit1)$ gives:

	loglik	Chisq	Df	p
1	-330.38			
2	-324.84	11.063	8	0.20

Conclusion: the LRT test shows no evidence against the reduced model ($p - value$= 0.20), which indicates the difference between these two models is not significant, and we prefer the smaller reduced model $cox1$.

4.3.4 Checking for Interaction

$step(cox1, \sim .^2)$

Conclusion: Adding the interaction term makes the AIC values increase and we may conclude that there is no need to add interactions, so our final model is **cox1**.

$cox1 < -coxph(Surv(mor) \sim factor(knot) + offg)$

Table 4.5 presents the results of $summary(cox1)$:

Based on the above summary output of $cox1$, we may make the following **comments**:

Step	Df	AIC
		668.75
+ factor(offg):moe	1	669.66
+ factor(knot):moe	2	671.33
+ factor(knot):factor(offg)	2	672.65
- factor(offg)	1	673.06
- factor(knot)	2	677.90
- moe	1	696.64

Table 4.5: p-values of covariates in the final model.

	coef	exp(coef)	se(coef)	z	p	
factor(knot)1	1.049262	2.855543	0.316801	3.312	0.000926	$***$
factor(knot)2	0.691564	1.996836	0.361133	1.915	0.055495	
factor(offg)1	1.452420	4.273444	0.491035	2.958	0.003098	$**$
moe	-4.811266	0.008138	0.936015	-5.140	2.75e-07	$***$

1. The estimated coefficient for the single knot as MSRC is 1.049 with very small p-value. Hence, fixing other covariates, the hazard ratio between the lumber with a single knot as MSRC and the one with knot combination as MSRC is exp(1.049)/exp(0.692) = 2.85554/1.99684 = 1.43, which means that the prior ones are 1.43 times more likely than the later ones to fail(having shorter survival). Similarly, the hazard ratio between the lumber with a single knot as MSRC and the ones with other defects than knot is 2.856, which means that the prior ones are 2.856 times more likely than the later ones to fail(having shorter survival). This is consistent with the side-by-side boxplots of "mor" against "knot" in the exploratory data analysis (EDA), which shows lumbers with a single knot as MSRC posses the lowest "mor" than ones with other two categories of "knot" as MSRC.

2. The estimated coefficient for the offgrade pieces of lumber is 1.452, and exp(1.452) = 4.273, which means the offgrade pieces of lumbers are 4.273 times weaker than standarded ones. This is also consistent with the conclusion in EDA.

3. Fixing other covariates, lumbers with higher moe have a decreased

hazard than the ones with lower moe. This is quite reasonable in common sense as the higher elasticity a piece of lumber the less likely that the failure will occur.

4.4 Model Diagnostics for the Cox PH Model

As in the case of a linear or generalized linear model, it is desirable to determine whether a fitted Cox regression model adequately describes the data. The model checking procedures below are based on residuals. In linear regression methods, residuals are defined as the difference between the observed and predicted values of the dependent variable. However, when the partial likelihood function is used in the Cox PH model, the usual concept of residual is not applicable.

We will discuss three major residuals that have been proposed for use in connection with the Cox PH model: the **Scaled Schoenfeld residuals**[5], the **Deviance residuals**[18] and the **Cox-Snell residuals**[8]. Then we will talk about influence assessment and strategies for analysis of nonproportional data.

4.4.1 Checking for the Proportional Hazards Assumption

The main assumption of the Cox PH models is proportional hazards[16]. Proportional hazard means that the hazard function of one individual is proportional to the hazard function of the other individual, i.e., the hazard ratio is constant over time. There are several methods for verifying that a model satisfies the assumption of proportionality.

The kth *Schoenfeld residual* (Schoenfeld, 1982) defined for the kth subject on the jth explanatory variable x^j is given by

$$r_{sjk} = \delta_k x_k^j - a_k^j,$$

where δ_k is the kth subject's censoring indicator, x_k^j is the value of the jth explanatory variable on the kth individual in the study,

$$a_k^j = \frac{\sum_{m \in R(y_k)} \exp(\underline{x}_m' \hat{\beta}) x_m^j}{\sum_{m \in R(y_k)} \exp(\underline{x}_m' \hat{\beta})},$$

and $R(y_k)$ is the risk set at time y_k. The MLE $\hat{\beta}$ is obtained from maximizing Cox's partial likelihood function. The Shoenfeld residuals for each predictor x^j must sum to zero. We define the *scaled Schoenfeld residuals* by the product of the inverse the estimated variance-covariance matrix of the kth Schoenfeld residual and the kth Schoenfeld residual, so that the kth Schoenfeld residual has an easily computable variance-covariance matrix.

Tests and graphical diagnostics for proportional hazards may be based on the $scaled\ Schoenfeld$
$residuals$. Conveniently, the **cox.zph** function calculates tests of the proportional hazards assumption for each covariate, by correlating the corresponding set of scaled Schoenfeld residuals with a suitable transformation of load (the default is based on the $Kaplan - Meier\ estimate$ of the survival function, i.e., $\widehat{S}(r)$ for the bending data). If the PH assumption holds for a particular covariate then the scaled Schoenfeld residual for that covariate will not be related to survival time. Using the cox.zph function, rho is the Pearson product-moment correlation between the scaled Schoenfeld residuals and survival time. The null hypothesis is that the correlation between the scaled Schoenfeld residuals and the ranked survival time is zero. Rejection of the null hypothesis concludes that the PH assumption is violated.

As mentioned, **cox.zph** computes a test for each covariate, along with a global test for the model as a whole:
$cox.zph(cox1)$ gives:

```
factor(knot)1   -0.0875   0.694   0.4048
factor(knot)2    0.1215   1.491   0.2220
offg             0.0439   0.190   0.6633
moe             -0.0524   0.319   0.5722
GLOBAL             NA     9.179   0.0568
```

Therefore, there is no statistically significant evidence of non-proportional hazards for any of the covariates, and the global test is also not quite statistically significant. These tests are sensitive to linear trends in the hazard. Moreover, we may plot the scaled Schoenfeld residuals against load-to-failure for each covariate in Figure 4.1:

Interpretation of these graphs is greatly facilitated by smoothing, for which purpose **cox.zph** uses a smoothing spline, shown on each graph by a solid line; the broken lines represent $\pm 2 - standard - error$ envelopes around the fit. **Systematic departures from a horizontal line are indicative of non-proportional hazards**. The assumption of proportional hazards appears to be supported for the covariate **offg** (which is, recall, a dummy variable, accounting for the two bands in the graph) and **moe**. However, there appears to be a trend in the plot for **knot**, with the **knot** effect

Figure 4.1: Scaled Schoenfeld residuals against load-to-failure.

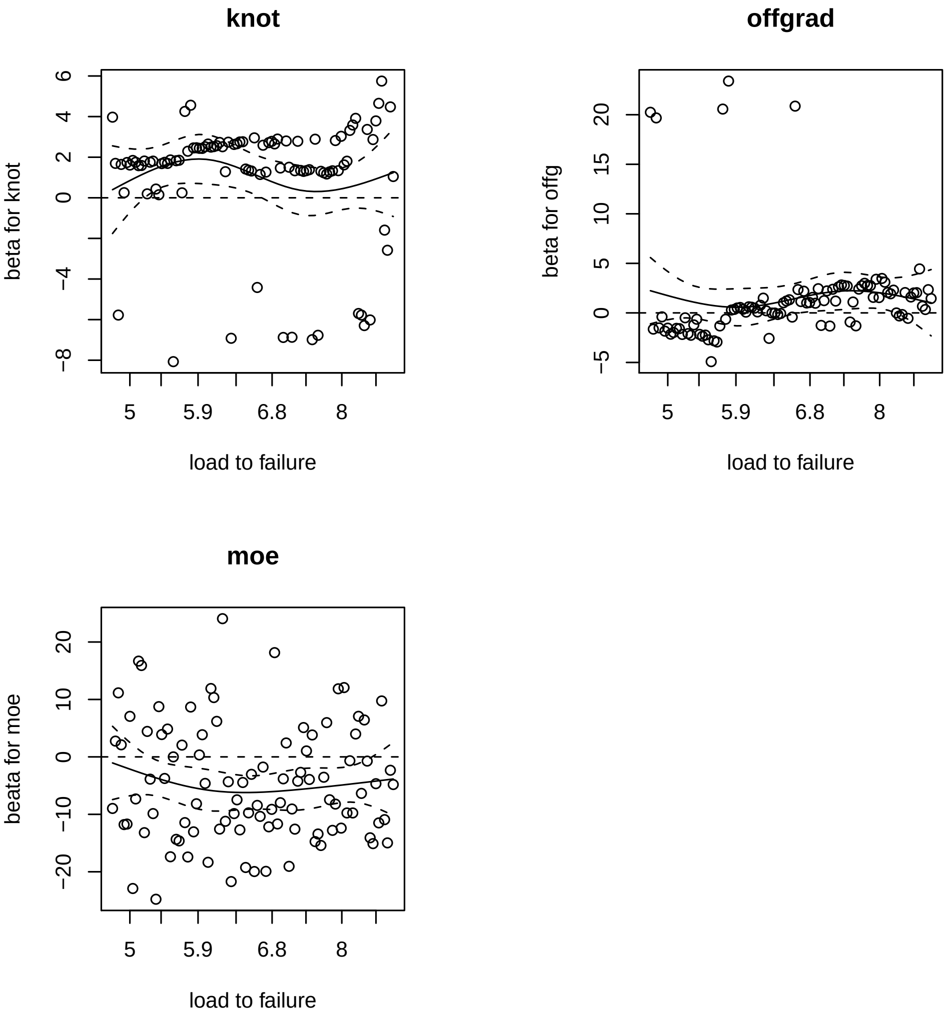
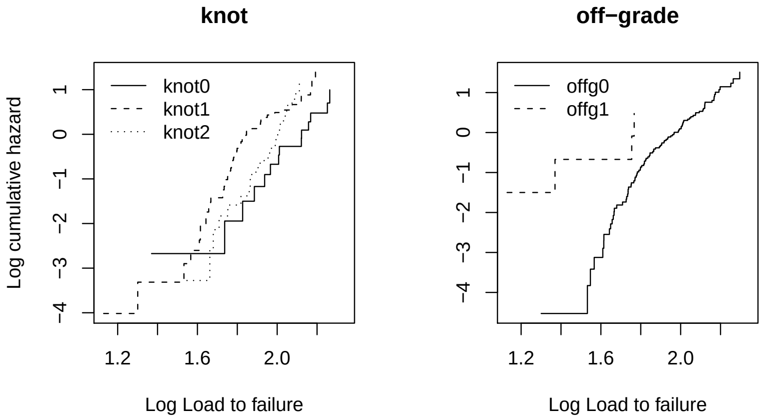


Figure 4.2: Graphical check of the PH assumption.



increasing with load. That is, the variability band for **knot** (a categorical variable with 3 levels, accounting for the 3 bands in the graph) displays a positive slope over load, suggesting non-proportionality of hazard and conflicting with the finding of the cox.zph test.

An alternative (and less sensitive) means of testing the proportional hazards assumption is to plot $\log[-\log S(r)]$ vs $\log(r)$ in Figure 4.2.

We conclude that the $\log[-\log S(r)]$ vs load plots are parallel for **offg** while nonparallel for **knot**, implying that the proportional hazards assumption has been violated for **knot**, which is supported by the Schoenfeld residual plots. Therefore, it gives us some concern about whether the Cox PH model is appropriate.

4.4.2 Assessing Goodness-of-Fit

The *ith* Cox-Snell residual is defined as

$$r_{Ci} = \hat{H}_0(t_i) \times \exp(\underline{x}_i'\hat{\beta}) = \hat{H}_i(t_i) = -\log \hat{S}_i(t_i),$$

where $\hat{H}_0(t_i)$ and $\hat{\beta}$ are the MLE's of the baseline cumulative hazard function and coefficient vector, respectively.

$r_{Ci} = -\log \hat{S}_i(t_i)$ will have a unit exponential distribution with $f_R(r) = \exp(-r)$. Let $S_R(r)$ denote the survival function for the Cox-Snell residual r_{Ci}. Then,

$$S_R(r) = \int_r^\infty \exp(-x)dx = \exp(-r),$$

and

$$H_R(r) = -\log S_R(r) = -\log(\exp(-r)) = r.$$

Therefore, we plot the cumulative hazard function $H_R(r_{Ci})$ versus Cox-Snell residual r_{Ci} to check the fit of the model. This gives a straight line with unit slope and zero intercept if the fitted model is correct. Note the Cox-Snell residuals will not be symmetrically distributed about zero and cannot be negative.

Then, we assess the goodness of fit for this Cox PH model by residual plots. A plot of the Cox-Snell residuals against the cumulative hazard of Cox-Snell residuals is presented in Figure 4.3. There is some obvious evidence of a systematic deviation from the straight line with an intercept zero and a slope one, which gives us some concern about the adequacy of the fitted model.

4.4.3 Checking for Outliers

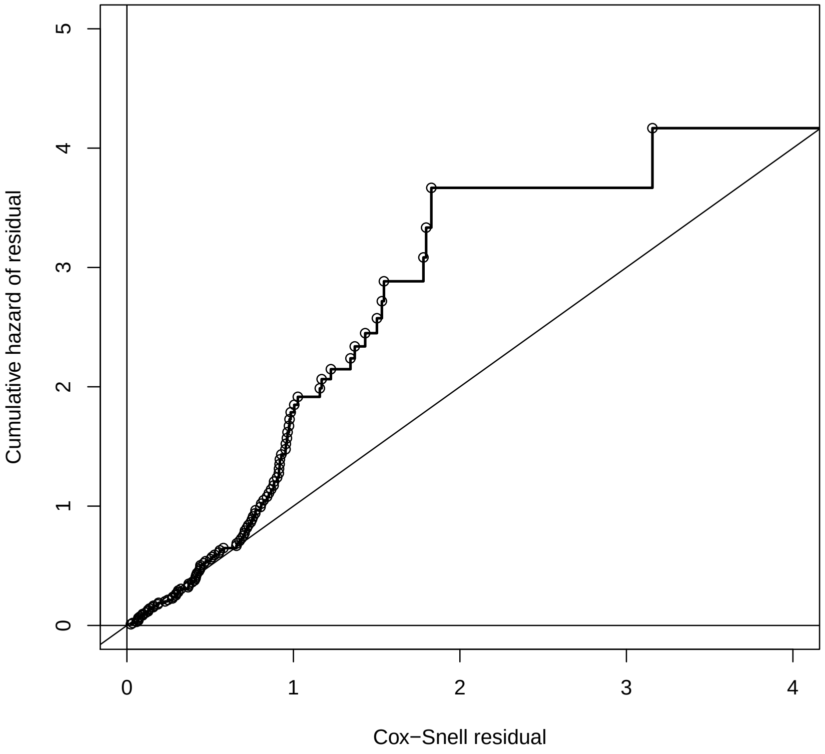
The *ith* deviance residual is defined by

$$r_{Di} = sign(r_{m_i})\sqrt{-2\{r_{m_i} + \delta_i \log(\delta_i - r_{m_i})\}},$$

where the function sign() is the sign function which takes the value 1 if r_{m_i} is positive and -1 if r_{m_i} is negative; $r_{m_i} = \delta_i - r_{Ci}$ is the martingale residual; and $\delta_i = 1$ for uncensored observation, $\delta_i = 0$ for censored observation.

In a fitted Cox PH model, the hazard of failure for the *ith* individual at any time depends on the value of $\exp(\beta' x_i)$ that is called the **risk score**. A plot of deviance residuals versus the risk score is a helpful diagnostic to assess a given individual on the model. Potential outliers will have deviance

Figure 4.3: Cumulative hazard plot of the Cox-Snell residual for Cox PH model.



residuals whose absolute values are very large. This plot will give information about characteristics of observations that are not well fitted by the model.

A plot of deviance residuals against the covariates can also be obtained. Any unusual patterns may suggest features of the data that have not been adequately fitted for the model. Very large or very small values suggest that the observation may be an outlier in need of special attention.

The plots of deviance residuals against the risk score, index and covariates are given in Figure 4.4. They show only one possible outlier, but none of them seems to be systematically distributed about zero. Therefore, overall, we have some concern about the adequacy of the fitted Cox PH model.

4.4.4 Influential Observations

Figure 4.5 shows the change in each regression coefficient when each observation is removed from the data (influence statistics). The changes plotted are scaled in units of standard errors and changes of less than 0.1 are of little concern.

These plots give us an idea of the influence individual observations have on the estimated regression coefficients for each covariate. Most of the changes in the regression coefficients are less than 0.1 s.e.'s of the coefficients and all others are less than 0.2 s.e.'s. Therefore, data sets where the influence plot is tightly clustered around zero indicate an absence of influential observations.

4.4.5 Dealing with the Violation of the Proportional Hazards Assumption

From the analyses conducted so far, we conclude that the proportional hazards assumption has been violated for the variable "knot". One method of dealing with this is to stratify the model by "knot". This means that we produce a separate baseline hazard function for each level of "knot". However, by stratifying, we cannot obtain a hazard ratio for "knot" since the 'knot effect' is absorbed into the baseline hazard.

The two models are given as below:

$cox1 < -coxph(Surv(mor) \sim factor(knot) + offg + moe, method =$
$\text{``}breslow\text{''})$
$cox2 < -coxph(Surv(mor) \sim$
$strata(factor(knot)) + offg + moe, method = \text{``}breslow\text{''})$

Figure 4.4: Deviance residuals against the risk score,index and covariates.

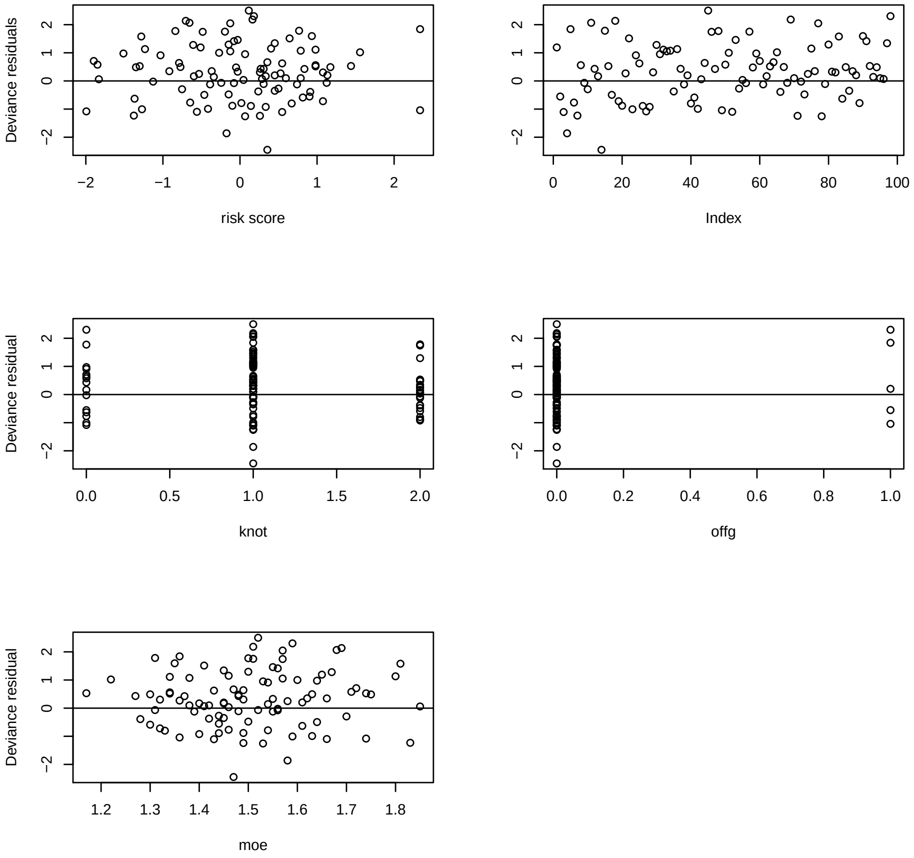
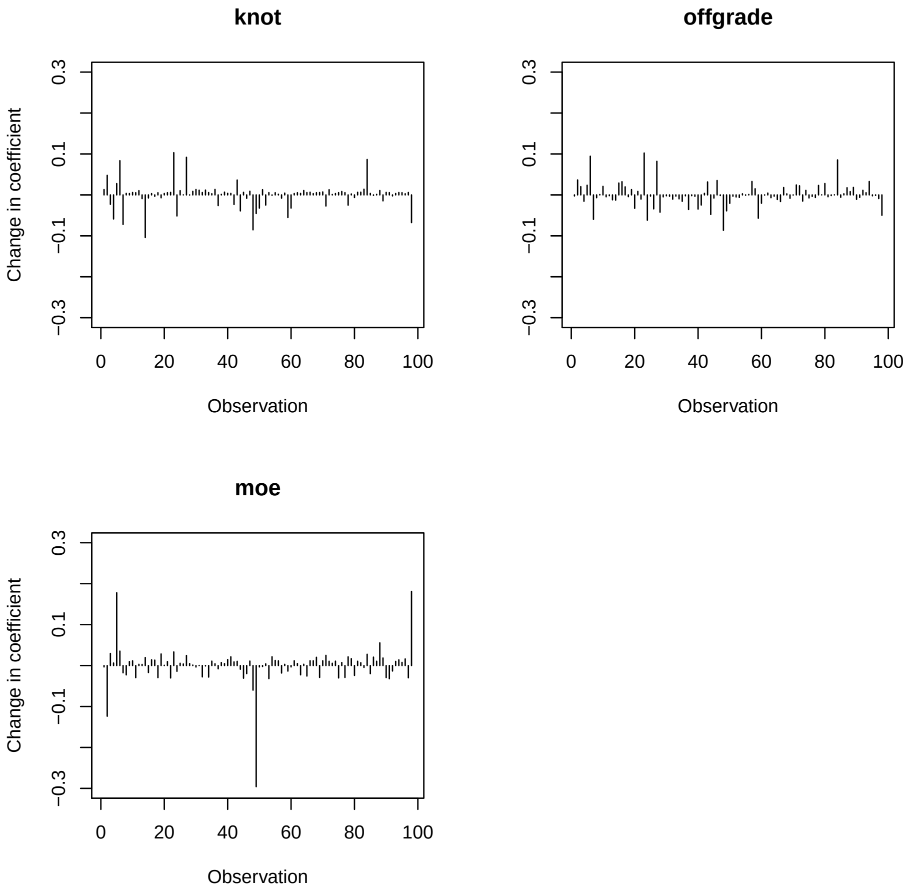


Figure 4.5: Influence statistics.



We may compare these two models using the AIC criterion. Since the stratified model *cox*2 provides a smaller AIC value than the previous model *cox*1, we may conclude that the stratified model gives a better fit for this data. However, if the covariate "knot" is of primary interest, this method is not recommended. Therefore, we may try other appropriative alternatives, such as the accelerated failure time model that will be discussed in the sequel.

Chapter 5

Parametric Survival Models

The accelerated failure time (AFT) model[4, 7, 11] is another alternative of the Cox PH model when the PH assumption is violated. The AFT model can be used to express the magnitude of effect in a more accessible way in terms of the difference between groups in survival strength. Under AFT models we measure the direct effect of the explanatory variables on the survival strength instead of hazard, as we do in the PH model. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the corresponding covariate on the mean survival strength.

5.1 Exploring the Distribution of Load to Failure

The most commonly used AFT models include the exponential AFT model, Weibull AFT model, log-logistic AFT model, and log-normal AFT model. The AFT models are named for the distribution of survival data.

Since each parametric distribution is defined by a different hazard function, we can check the consistency of survival data with a specific distribution by investigating the corresponding underlying linearity. Four different plots can be obtained and the corresponding distributions indicated if these plots form a straight line pattern. The plots and their associated distributions are given in Table 5.1, where $Z(p)$ means the pth quantile from the standard normal distribution.

Table 5.1: Plots and associated distributions.

Plot	Distribution indicated by a straight line pattern
-log[$S(t)$] vs. t	Exponential, through the origin
log[$-log(S(t))$] vs. log(t)	Weibull
log[$(1-S(t))/S(t)$] vs. log(t)	Log-logistic
Z[1-S(t)] vs. log(t)	Log-normal

We present these four different plots based on the bending data in Figure 5.1. By comparing the straightness of these lines, we may see that the distribution of bending data is more likely to be one of Weibull, log-normal, or log-logistic. Also, it seems that the bending data should not be from the exponential distribution because the line is far away from the straight line through the origin. Note it is the left hand tail that accounts in applications, and the left hand tail observations seem more likely to be from the Weibull distribution as they present a slightly better straight line through the origin.

5.2 Variable Selection

We fit the bending data using exponential, Weibull, log-logistic, and log-normal AFT models. In both univariate and multivariate AFT models, "knot" , "offg" and "moe" are statistically significantly associated with load to failure MOR. No interactions are statistically significant in multivariate AFT models. There is no big difference for the estimated Weibull, log-logistic and log-normal models, but the estimated exponential model is quite different. This indicates the distribution of MOR may be far away from the exponential distribution. The results from the different AFT models applied to the bending data are presented in Table 5.2, where η is the estimated acceleration factor.

Figure 5.1: Exploring distribution of load to failure.

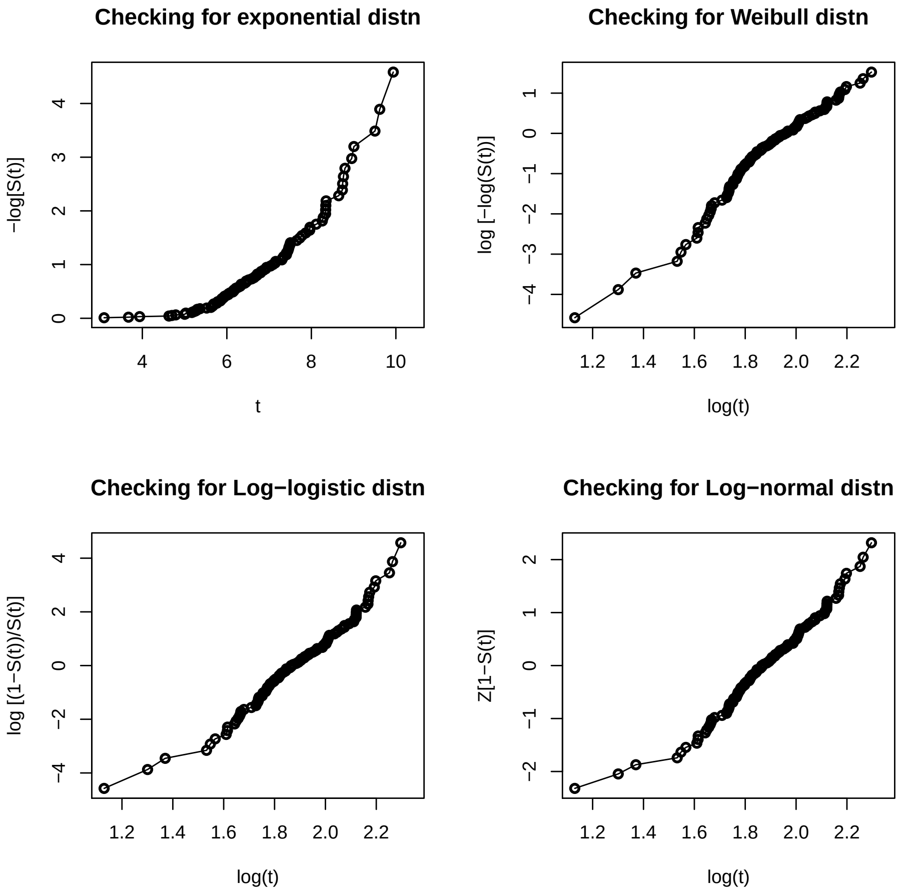


Table 5.2: Results from AFT models for the bending data.

Coef	Exponential				Weibull				Log-logistic				Log-normal			
	α	sd	η	p	α	sd	η	p	α	sd	η	p	α	sd	η	p
μ	1.41			0.45	1.44			0.00	1.41			0.00	1.42			0.00
knot1	-0.21	0.36	0.81	0.55	-0.23	0.04	0.79	0.00	-0.23	0.06	0.79	0.00	-0.21	0.05	0.81	0.00
knot2	-0.10	0.38	0.90	0.78	-0.12	0.04	0.88	0.02	-0.10	0.06	0.90	0.10	-0.10	0.06	0.90	0.11
ksize	-0.00	0.01	1.00	0.98	0.00	0.00	1.00	0.97	0.00	0.00	1.00	0.97	-0.00	0.00	1.00	0.88
rnl	-0.00	0.01	1.00	0.85	0.00	0.00	1.00	0.15	-0.00	0.00	1.00	0.32	-0.00	0.00	1.00	0.24
offg	-0.30	0.50	0.74	0.54	-0.23	0.06	0.79	0.00	-0.25	0.10	0.77	0.02	-0.33	0.08	0.71	0.00
loc	0.00	0.01	1.00	0.95	0.00	0.00	1.00	0.99	0.00	0.00	1.00	0.77	0.00	0.00	1.00	0.61
face1	0.07	0.24	1.07	0.76	0.07	0.03	1.07	0.03	0.07	0.03	1.07	0.07	0.07	0.03	1.07	0.06
face2	-0.00	0.34	1.00	0.99	-0.06	0.04	0.94	0.19	-0.02	0.05	0.98	0.77	0.01	0.05	1.01	0.83
spec2	-0.04	0.41	0.96	0.91	-0.06	0.05	0.94	0.29	-0.05	0.06	0.95	0.43	-0.04	0.06	0.96	0.55
spec3	-0.11	1.13	0.89	0.92	-0.18	0.15	0.83	0.23	-0.13	0.15	0.87	0.38	-0.09	0.18	0.91	0.62
mois	-0.02	0.12	0.98	0.86	-0.03	0.01	0.97	0.16	-0.02	0.01	0.98	0.30	-0.02	0.01	0.98	0.29
moe	0.66	0.83	1.93	0.42	0.77	0.11	2.15	0.00	0.64	0.12	1.89	0.00	0.64	0.13	1.89	0.00

For the parametric models we discuss here, the AIC is given by

$$AIC = -2 \times \log(maximumlikelihood) + 2 \times (a + b),$$

where a is the number of parameters in the specific model and b the number of one-dimensional covariates. For example, $a = 1$ for the exponential model, $a = 2$ for the Weibull, log-logistic, and log-normal models. In Table 5.3, we compared all these AFT models using statistical criteria–AIC. Note the smaller AIC is the better. The Weibull AFT model appears to be an appropriate AFT model according to AIC compared to other AFT models. However, the exponential model provides the worst fit, which is consistent with the conclusion we drawn from Figure 5.1.

Table 5.3: AIC in the AFT models.

Model	Log-likelihood	a	b	AIC
Exponential	-283.6	1	12	593.1685
Weibull	-139	2	12	306.0937
Log-logistic	-144.4	2	12	316.8026
Log-normal	-144.5	2	12	317.0219

5.3 Q-Q Plot to Check the AFT Assumption

An initial method for assessing the potential for an AFT model is to produce a quantile-quantile plot. For any value p in the interval (0,100), the *pth* percentile is

$$t(p) = S^{-1}(\frac{100 - p}{100}).$$

Let $t_0(p)$ and $t_1(p)$ be the *pth* percentiles estimated from the survival functions of the two groups of survival data. The percentiles for the two groups may be expressed as

$$t_0(p) = S_0^{-1}(\frac{100 - p}{100}),$$

$$t_1(p) = S_1^{-1}(\frac{100 - p}{100}),$$

where $S_0(t)$ and $S_1(t)$ are the survival functions for the two groups. So we can get

$$S_1[t_1(p)] = S_0[t_0(p)].$$

Under the AFT model, the assumption is $S_1(t) = S_0[t/\eta]$, and so

$$S_1[t_1(p)] = S_0[t_1(p)/\eta].$$

Therefore, we get

$$t_0(p) = \eta^{-1} t_1(p).$$

The percentiles of the survival distributions for the two groups can be estimated by the KM estimates of the respective survival functions. If the accelerated failure time model is appropriate, a plot of percentiles of the KM estimated survival function from one group against another should be given an approximate straight line through the origin. The slop of this line will be an estimate of the acceleration factor η^{-1}.

For the 3-level categorical covariate “knot”, we have 3 possible pairwise combinations. The Q-Q plot in Figure 5.2 approximates well a straight line from the origin indicating that the AFT model may be appropriate.

5.4 Model Diagnostics for the AFT Model

5.4.1 Overall Goodness-of-Fit

We check the goodness of fit of the model using residual plots. The cumulative hazard plot of the Cox-Snell residuals in the Weibull model is presented in Figure 5.3. The plotted points mostly lie on a line that has a unit slope and zero intercept. So there is no reason to doubt the suitability of this fitted Weibull model. Comparing Figure 4.3 with Figure 5.3, we may see that the Weibull AFT model provides a much better fit than the Cox PH model. We conclude that the Weibull produces the best fitting AFT model based on AIC criteria and residuals plot.

5.4.2 Checking for Outliers

Similarly, the plots of deviance residuals against the risk score, index and covariates are given in Figure 5.4. They display only one possible outlier, but none of them seem to be systematically distributed about zero. Therefore, overall, we have little concern about the adequacy of the fitted log-normal AFT model.

Figure 5.2: Q-Q plot for load to failure.

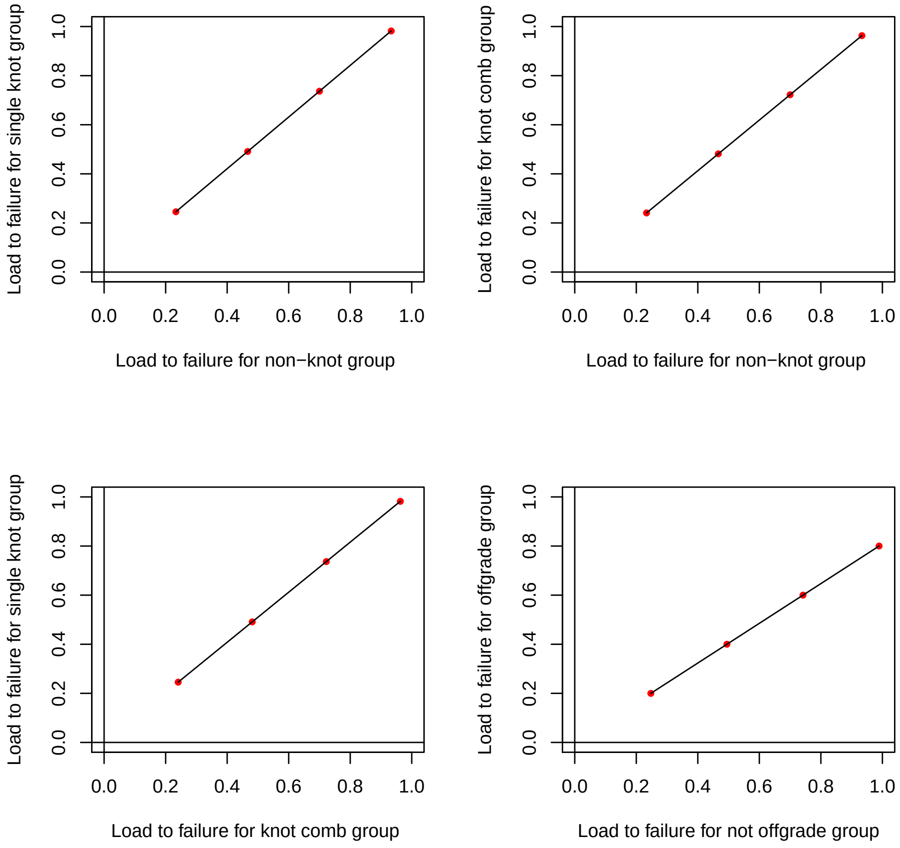


Figure 5.3: Cumulative hazard plot of the Cox-Snell residual for the Weibull AFT model.

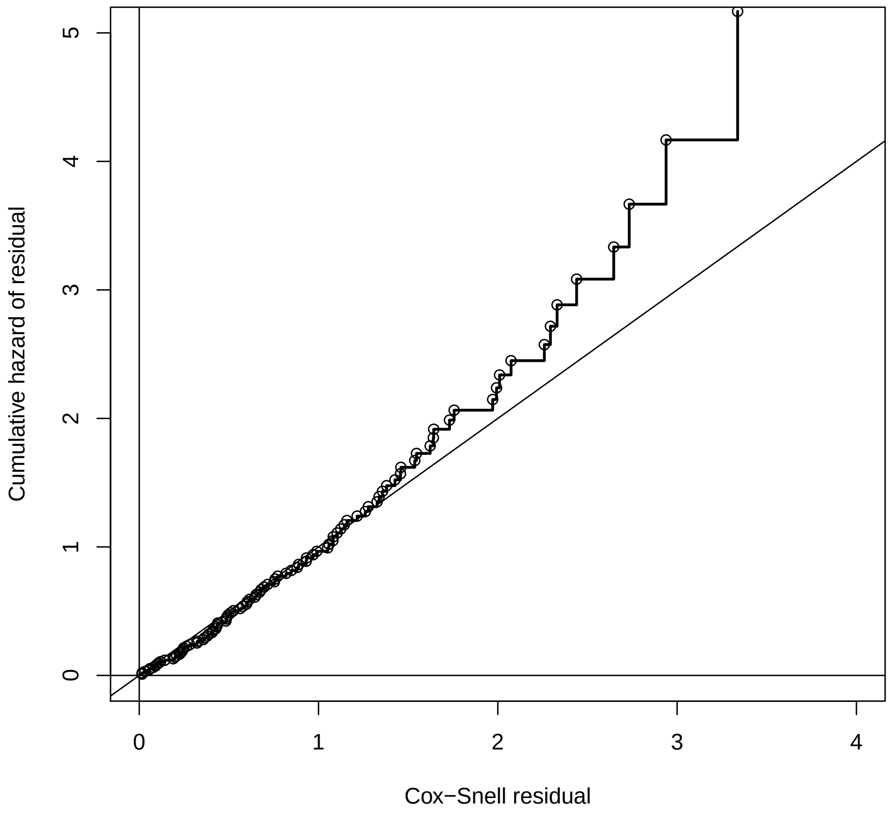
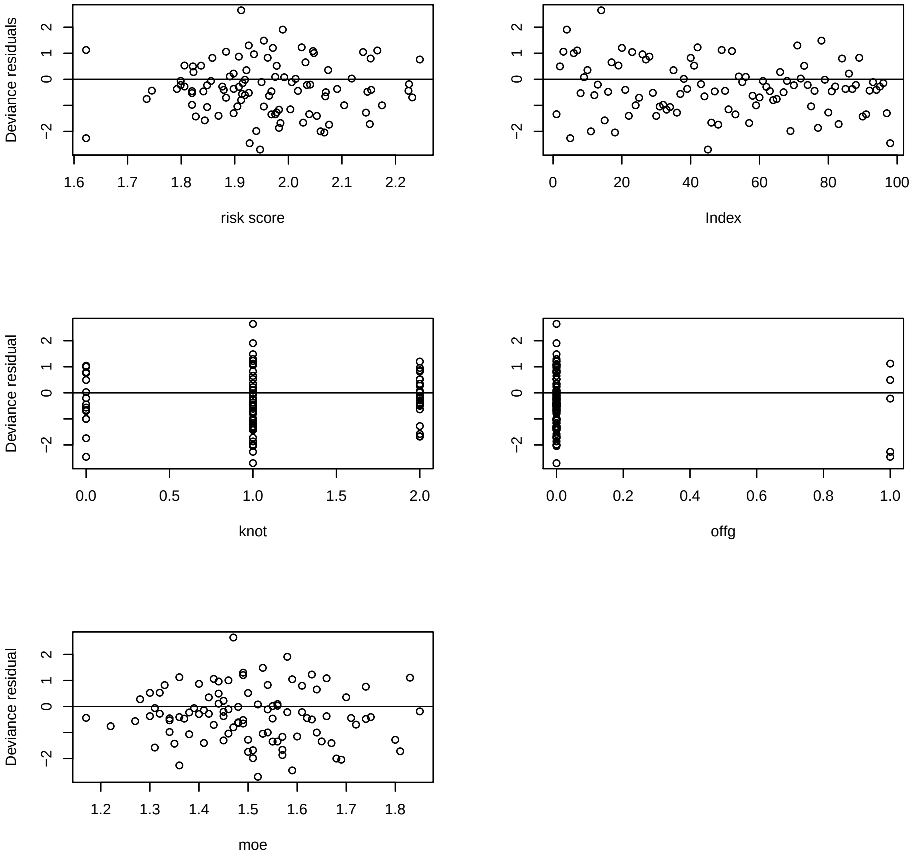


Figure 5.4: Deviance residuals against the risk score, index and covariates.



5.4.3 Influential Assessment

Figure 5.5 shows the change in each regression coefficient when each observation is removed from the data (influence statistics). The changes plotted are scaled in units of standard errors. Changes of less than 0.04 are of little concern.

These plots give us an idea of the influence individual observations have on the estimated regression coefficients for each covariate. Most of the changes in the regression coefficients are less than 0.02 s.e.'s of the coefficients and all others are less than 0.03 s.e.'s. Therefore, data sets where the influence plot is tightly clustered around zero indicate an absence of influential observations.

5.5 Interpretation of Results

Finally, we may fit the Weibull AFT model with only statistically significant covariates – "knot", "offg" and "moe":

$wei < -survreg(Surv(mor) \sim factor(knot){+}offg{+}moe, dist = \text{“}weibull\text{”})$

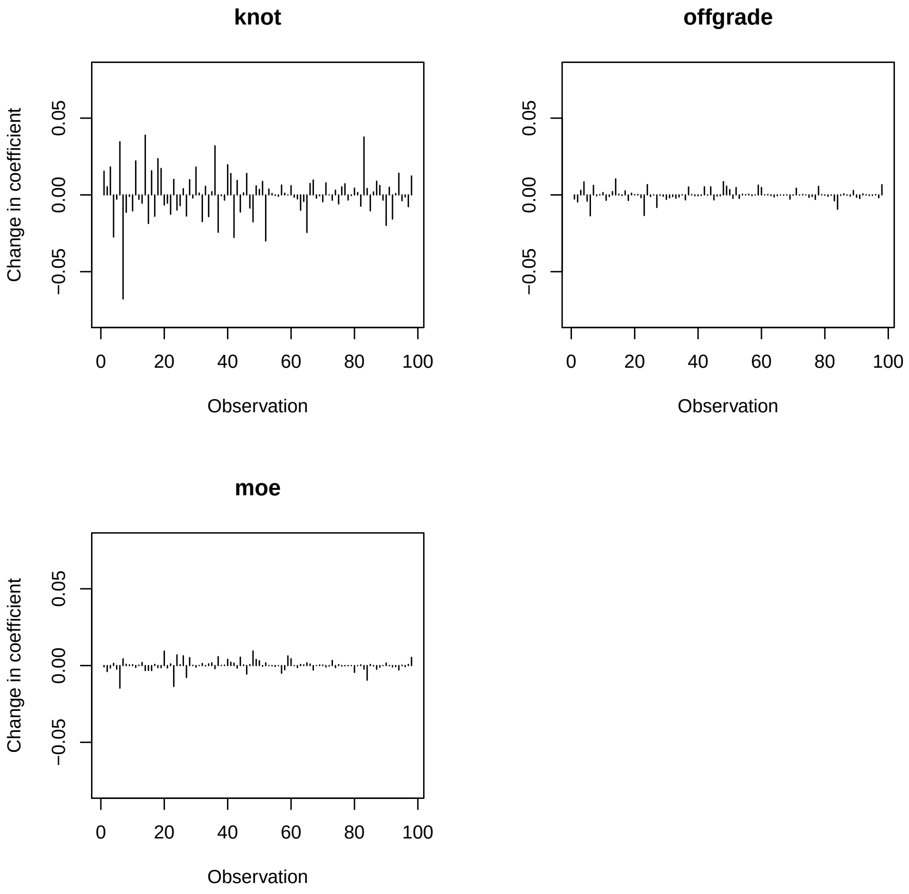
The summary for this model is given in Table 5.4:

Table 5.4: Summary for the final Weibull AFT model.

	coef	se(coef)	η(coef)	p	
(Intercept)	1.0174	0.1796		1.47e-08	
factor(knot)1	-0.1428	0.0412	0.8669274	5.31e-04	* * *
factor(knot)2	-0.0978	0.0471	0.9068303	3.79e-02	*
offg	-0.2118	0.0661	0.8091265	1.35e-03	**
moe	0.7057	0.1119	2.0252639	2.86e-10	* * *
Log(scale)	-1.9688	0.0786		- 2.00e-138	
Loglik(model)=	-145				

Conclusion: the acceleration factor (η) for "offg" is 0.81 (less than 1), which indicates that the smaller survival load is more likely for off-grade lumber. The η's for "knot" is also less than 1 imply that this variable yields a lower load to failure, and the "single knot" group is more likely to break than the "knot combination" group since it has a even smaller acceleration factor. The acceleration factor (η) for "moe" is 2.03 (more than 1), which indicates that the larger survival load is more likely for the piece of lumber with higher MOE. These conclusions are consistent with the ones drawn from application of the Cox PH model.

Figure 5.5: Influence statistics.



5.6 Simulation Study

5.6.1 Introduction

In practice the model relating the strength of a piece of lumber to its covariates cannot be known and we explore through simulations studies the inferential effect of mis-specifying that model. However, to constrain the scope of our study to a practical limit, we will assume that the structural link between the response and the covariates is correct based on our belief that diagnostic assessments of data would suggest a reasonable choice for that link. Thus we restrict our studies to the effect of mis-specifying the random error component of an AFT model for the strength. More precisely, we looked at the estimates for the coefficients in that link when the standard Normal distribution, the Cauchy t_1 distribution, the Student t_2 distribution and the standard Gumbel distribution are assumed for the error distribution when the true distribution is none of these. The details follow below.

A simulation study was conducted to compare the estimates for the AFT models with Weibull, exponential, log-normal and log-logistic distribution assumptions. Also, one of our interests is to investigate predictive accuracy. One commonly used measure of predictive accuracy is the *expected squared error* of the estimate. This quantity is defined as the expected squared difference between predicted and observed values, that is, the average squared difference between predicted and observed values if the experiment were repeated finitely often and new estimates were made at each replication.

5.6.2 Description of Method

Our final log-linear form of the AFT model with respect to load Y is given by:

$$\log Y_i = \mu + \alpha_1\, knot_i + \alpha_2\, offg_i + \alpha_3\, moe_i + \sigma\, \varepsilon_i,$$

where $\mu = 1.0$, $\alpha_{11} = -0.1$,$\alpha_{12} = -0.1$,$\alpha_2 = -0.2$, $\alpha_2 = 0.7$ and $\sigma = 0.1$ are fixed. The significant X variables **knot**, **offg** and **moe** values from the original sample are also fixed with respect to replication of the study. The errors ε_i was generated parametrically from a standard Normal distribution, from a Cauchy t_1 distribution, from a Student t_2 distribution and from a standard Gumbel distribution. The response values Y_i, however, are randomly generated by the AFT model, because of the error component of the model. We would then regress the response values Y_i on the fixed X matrix (knot, offg and moe) to obtain the regression coefficients estimates at each replication.

We may also obtain the average squared difference between predicted and observed values $\frac{1}{98}\sum(Y_i - \widehat{Y}_i)^2$ at each replication.

Since there are 4 settings of the errors term distributions and 4 settings of AFT models with different distribution assumptions (Weibull, exponential, log-normal and log-logistic), there were total 4 × 4 (16) different settings of simulation conducted. Each simulation involved 1000 replications with a sample size 98.

5.6.3 Results for the Simulation

For each simulation, the estimates were computed using the Weibull AFT model, the exponential AFT model, the log-normal AFT model and the log-logistic model. Let's take the coefficient α_2 for the covariate "offg" for example. Table 5.5 shows the average values of the parameter estimates for α_2 and their standard deviations over the 1000 replications with 4 different error terms.

Table 5.5: True value $\alpha_2 = -0.2$. Expected value, standard deviation of parameter estimates.

Setting	Weibull	Exponential	Log-normal	Log-logistic
Normal	-0.211(0.08)	-0.214(0.07)	-0.212(0.07)	-0.211(0.07)
Cauchy t_1	-0.224(0.15)	-0.261(0.32)	-0.240(0.35)	-0.253(0.34)
Student t_2	-0.233(0.46)	-0.276(0.57)	-0.265(0.47)	-0.269(0.57)
Gumbel	-0.245(2.10)	-0.292(6.12)	-0.283(5.10)	-0.288(5.12)

Overall, based on this simulation study, the Weibull AFT model shows better estimations on this coefficient, which are closer to the true values. The patterns of other coefficients are the same in most cases. Moreover, the mean of predictive accuracy $\frac{1}{98}\sum(Y - \widehat{Y})^2$ over 1000 replications shows that the Weibull AFT model with a standard normal error performed better than other models since it gave the smallest mean predictive accuracy. This confirms our choice of the Weibull AFT model once again.

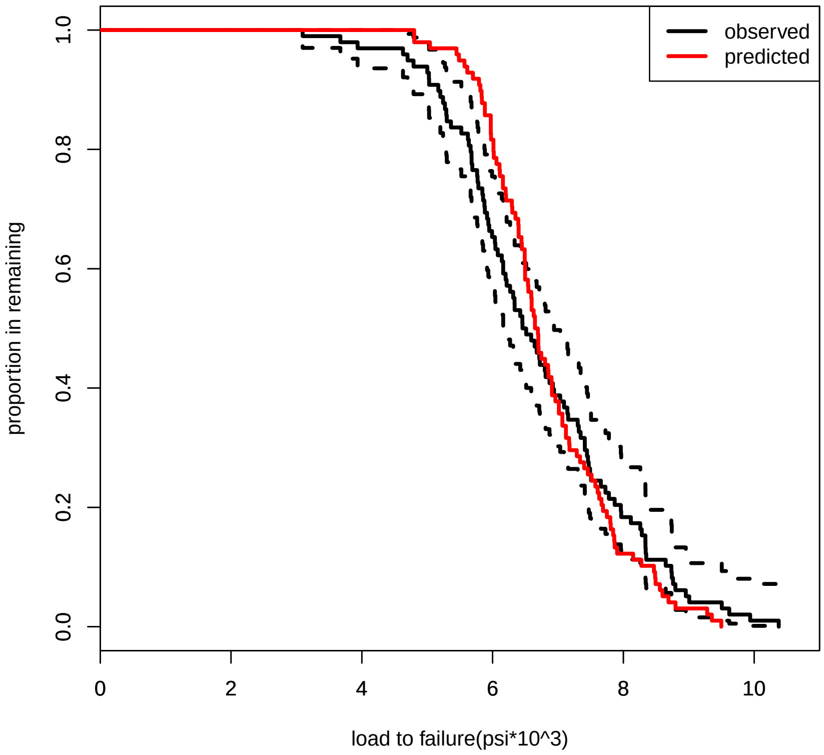
5.7 Cross-Validation

A stringent test of a model is an external validation - the application of the 'frozen' model to a new population. It is often the case that the failure of a model to validate externally could have been predicted from an honest (unbiased) ‘internal’ validation. One well-known method for obtaining nearly unbiased internal assessments of accuracy is cross-validation. To uncover problems that may make prediction models misleading or invalid, the predictive accuracy has to be unbiasedly validated using cross-validation.

Each time, we drop one record from the sample and the remaining data are used as a training (model development) sample. That model is ‘frozen’ and applied to the dropped out sample for computing predictive survival probability. For example, we drop record 98, then fit a model on records 1 to 97 and use this model to predict the 98th record, so on so forth.

The following plot Figure 5.6 gives us an idea of how well the predicted survival curve from the final Weibull AFT model tracks observed Kaplan-Meier estimates. The predicted survival is slightly larger than the observed in the lower tail and smaller than the observed in the upper tail. However, we see that predicted survival curve mainly falls within the 95% error bounds of the observed survival curve. Therefore, it does not produce large deviations from the true values.

Figure 5.6: Comparing observed and predicted survival curves.



Chapter 6

Conclusion and Discussion

This study is based on the wood strength data collected in a FPInnovations (FPI) laboratory. We employed survival analysis methods in this very different context - load to failure problem. A finding of the present study shows that a type of wood defect (knot), a lumber grade status (off-grade: Yes/No) and a lumber's modulus of elasticity (moe) have statistically significant effects on wood strength properties including bending strength and tension strength.

Forms of non-parametric and parametric bivariate-strength survival functions (Biv-KM and Biv-Weibull) have been explored to obtain the joint strength distributions. Association between MSRC and FC was also examined by the Cramer's V statistic and found to be just 0.3, indicating the strength of association is not that strong. However, this measure of strength highly depends on how the covariate values are aggregated into sub categories and in our case, these lumber categories were fairly fine, making a low V level inevitable with such a small dataset.

The Cox PH model is routinely applied to the analysis of survival data, but the proportional hazards (PH) assumption does not hold for 'knot' in this analysis. We also use four different accelerated failure time (AFT) models to fit the data. We found that the Weibull AFT model was the best fit for this dataset. The study considered here provides an example of a situation where Cox PH model is inappropriate and where the Weibull AFT model provides a better description of the data. We see that the Weibull AFT model is a more valuable and realistic alternative to the Cox PH model in some situation. Moreover, the AFT model has a more realistic interpretation in terms of an effect on expected load to failure and provides more informative results. To this content the AFT model has explanatory advantage in that covariates have a direct effect on load to failure rather on hazard functions as in the Cox PH model. Therefore, we suggest that using the Cox PH model may not be the optimum approach. The AFT model may provide an alternative method to fit some survival data. This final Weibull AFT model can be used to make the current lumber grading system (currently highly relays on graders' experience) more powerful and reliable.

Both of the Cox PH model and the Weibull AFT model yield exactly the same significant covariates - ‘knot’, ‘off-grade’ and ‘moe’, indicating these three are the most important predictors in our reliability modeling. In our study, a piece of lumber with a ‘knot’ defect is more likely to break than one with other defects; in particular, a piece of lumber with a ‘single knot’ defect is even more likely to break than one with a ‘knot combination’ defect. Not surprising, off-grade lumber is more likely to have lower survival loads than the standard ones. Also, the piece of lumber with a higher ‘moe’ is more likely to have a higher survival load.

As mentioned above, after applying these survival analysis methods to wood strength properties, we obtained the same significant covariates - ‘knot’, ‘offg’ and ‘moe’ in both bending strength data and tension strength data. These significant covariates can be used to match pieces of lumber in describing the relationships among strength properties. Obtaining matched pairs in this way helps solve the challenging problem that a single piece of lumber cannot be broken twice by two different strength tests. This is a major potential application of reliability modeling analysis conducted here, and this could even be used in other situations for example in analyzing the duration of load for lumber. We should also recognize the fact that a small sample size makes it difficult to find significant predictors and that in future work, a larger sample should be collected to find more others.

In practice the model relating the strength of a piece of lumber to its covariates cannot be known and we explore through simulations studies the inferential effect of misspecifying that model. These studies were conducted to compare the coefficients estimates from the AFT models with Weibull, exponential, log-normal and log-logistic distribution assumptions. The Weibull AFT model leads to somewhat better estimates of coefficients than the other incorrectly specified models. As well, it provides the best mean predictive accuracy. This confirms our choice of the Weibull AFT model once again.

Finally, to uncover problems that may make prediction models misleading or invalid, predictive accuracy has been unbiasedly assessed using cross-validation. We observe that predicted survival curve from the final Weibull AFT model tracks the observed Kaplan-Meier estimates very well. This study has shown the power of employing survival analysis methods in reliability in this very different context from that which originally led to its development.

Bibliography

[1] Agresti A. *An Introduction to Categorical Data Analysis.* Wiley-Interscience, 2008.

[2] Richard A.J., James W.E., and David W.G. Some Bivariate Distributions for Modeling the Strength Properties of Lumber. *Forest Products Laboratory Research Paper*, 1999.

[3] National Lumber Grades Authority. *Canadian Lumber Grading Manual.* National Lumber Grades Authority, 8th edition, 2001.

[4] Collett D. *Modelling Survival Data in Medical Research.* Chapman and Hall, London, 2003.

[5] Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika*, 69:239–241, 1982.

[6] Cox D.R. Regression Models and Life Tables(with Discussion). *Journal of the Royal Statistical Society*, 34:187–220, 1972.

[7] Cox D.R. and Oakes D. *Analysis of Survival Data.* Chapman and Hall, 1984.

[8] Cox D.R. and Snell E.J. A general definition of residuals with discussion. *Journal of the Royal Statistical Society*, 30:248–275, 1968.

[9] Lin D.Y. and Ying Z. A simple Nonparametric Estimator of the Bivariate Survival Function Under Univariate Censoring. *Biometrika*, 80:573–581, 1993.

[10] P. Hougaard. A class of multivariate failure time distributions. 73:671–678, 1986.

[11] Lawless J.F. *Statistical Models and Methods for Lifetime Data Analysis.* Wiley, New York, 1982.

[12] Peter J.S. *Analysis of Failure and Survival Data.* Chapman and Hall/CRC, 2002.

[13] Crowder M.J., Kimber A.C., Smith R.L., and Sweeting T.J. *Statistical Analysis of Reliability Data*. Chapman and Hall, 1991.

[14] Abbott R. Commentary on the Maximum Strength Reducing Defects(MSRD) and Failure Coding System. 2002.

[15] Abbott R. Forintek Knot and Failure Code. 2002.

[16] Mara T. and Jong S.K. *Suvival Analysis Using S: Analysis of Time-to-Event Data*. Chapman and Hall/CRC, 2004.

[17] Therneau T.M. A package for Survival Analysis in S. *Technical Report Mayo Foundation*, 1999.

[18] Therneau T.M., Grambsch P.M., and Fleming T.R. Martingale-based residuals for survival models. *Biometrika*, 77:147–160, 1990.