THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF STATISTICS

TECHNICAL REPORT #269

Wood Property Relationships and Survival Models in Reliability

BY

YAN CHENG, LANG WU, CONROY LUM,
JAMES V ZIDEK & XIAOLI YU

November 2012

Wood Property Relationships and Survival Models in Reliability

Yan Cheng, Lang Wu, Conroy Lum, Jim Zidek and Xiaoli Yu*

November 12, 2012

Abstract

This report studies the relationship between the strength properties of lumber and their covariate visual grading characteristics. In our mechanical wood strength tests, each piece fails under a continuously increasing load. This topic is central to the analysis of the reliability of lumber products in that it underlies the calculation of structural design values

The approaches described in the report, are adaptations of joint and conditional survival distributions based on both a parametric method (the Weibull distribution) and a nonparametric approach (Kaplan-Meier method). 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.

The development of the theory in the report is based on 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). The paper concludes that 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. It finds that that the Weibull AFT model provides a much

---

*The work reported in this manuscript was partially supported by FPInnovations and grants from the Natural Sciences and Engineering Research Council of Canada .

better fit than the Cox PH model in our data set with a satisfying predictive accuracy and leads to the development of a Bayesian version of the AFT.

1 Introduction

This paper proposes an adaption of time–to–event theory as applied for example in survival analysis, for modeling the load–to–failure in reliability analysis. It leads to a new theory for potential application in characterizing the strength of manufactured dimension lumber. In particular it may be applied should a need arise to change the grading rules.

Grading rules are based of observable features of a piece of lumber of specified length and volume from a prescribed group of species Only those features deemed to be predictors of the strength (say Y) of that piece of lumber would be relevant for this article and these we include in a vector X. An example of something that would be represented in X is the presence or absence of "shake" i.e. a separation along a grain of a piece of lumber and its length if present. Graders use these features (both those in X as well as those not) to classify lumber into subpopulations called grades. Under the grading rules each grade is uniquely associated with a particular range of the observable features and so in particular a subset of the range space of X say G. For a given grade, the length of a shake if present would have to lie in a specified range (a,b) for example. Thus the distribution of Ys would implicitly be determined by the limitations that the grading rules impose on X and the stochastic relationship between Y and X, in other words through

$$P(Y > y|X \in G) = \int_{\{x \in G\}} P(Y > y|X = x) dF_G(x) \tag{1}$$

with $F_G(x) = F(x)/P(X \in G)$.

Structural engineering design values are based on random in–grade test samples and the resulting Y–data they yield. In other words they are based on an empirical estimate of the so–called conditional survival function $P(Y > y|X \in G)$. These values say $\xi_{G\alpha}$ are chosen so that with very high conditional probability say 0.95, they are exceeded by randomly selected pieces in that grade after allowing for such things as sampling and estimation error.

The genesis of this report is the recognition that the representativeness of the original in–grade sample could conceivably decline over time. Equation 1 points to the two technical reasons why that change could occur. First F_G say it could become more skewed in the direction of regions in the range of X for which the associated Y values were lower. The second is a change would be in the stochastic relationship between Y and X that makes Y stochastically lower than it was. In other words, the survival probability $P(Y > y|X = x)$ might decline for in-grade x's for values of y in a region of importance in setting the design values

Substantive reasons for change can be broadly divided into two classes. The first are changes in the resource from which lumber is manufactured. For example one longterm monitoring program was initiated due to an increasing reliance on plantation lumber. Another possible cause of change in the resource is the changing climate. Quoting [7]

> "Gradual increases in temperature, changes in rainfall patterns, or modification in solar radiation will likely impact growth, regeneration and natural rates of mortality, while more abrupt changes in climate may lead to extensive fires, killing frosts, droughts, and outbreaks of insects and diseases [5]."

These authors conclude based on their analysis that 13 out of 15 native trees species have become more favorable to potential migration. Catastrophic changes like those described above from such things as fires could lead to the harvesting of more dead wood and hence a change in $F_G(x)$ in Equation 1. More specifically, the shake lengths for in–grade lumber might tend to be closer to b, the upper end of the allowable range (a,b). Things like migration suggested above could change the mix of species in the grade group in a way that increases the supply from species with stochastically lower Y–values for a given x. In other words, $P(Y > y|X = x)$ might change.

The second broad category of substantive reasons for change are those that point to possible changes the in–grade population even if the resource does not change. For example, the in–grade lumber population might change due to selection biases arising from changing market conditions that could pull lumber out of that population for special purpose applications.

The changes described above could lead to the need to either adjust the design values or to change the grading rules in order to maintain the reliability of the lumber. The second option would seem more appealing from a practical point of view in many cases, since publishing new design values would create a serious discontinuity in a traditional and well understood system if these were decreased in size. Both possibilities point to the potential need to reestimating their stochastic representatives in Equation 1 as part of a long term monitoring program.

This paper presents approaches for dealing with the second of the two technical reflections of change described above, namely reestimating $P(Y > y|X = x)$ on the basis of test data from a sample of in–grade lumber. The approaches taken in this paper are based on our adaptations of time– to–event theory for characterizing that conditional probability. The first is approach is based on the proportional hazards (PH) model which assumes that

$$\begin{aligned} P(Y > y|X = x) &\doteq S(y|x) \\ &= [S(t)]^{exp(\beta x')} \\ &= \exp\{-e^{\beta x'} \int_y^\infty h(u)du\}. \end{aligned}$$

Here $S(y) \doteq P(Y > y)$ denotes the baseline survival function while $h(u)$ denotes the baseline hazard function defined by

$$\begin{aligned} h(u)du &= P(u < Y < u + du|Y > u) \\ &= -\frac{d\log S(u)}{du}\,du \end{aligned}$$

so that

$$S(u) = \exp\{-\int_y^\infty h(u)du\}.$$

It follows that the harzard function corresponding to $S(y|x)$ is given by

$$h(u)e^{\beta x'}.$$

This expression is what gives rise to the terminology "proportional hazards"; each coordinate of x like x_1 gives rise too a multiplicative factor $exp(\beta_1 x_1$ that increases the hazard. The great success of this method derives from the fact that the h can be treated as a nuisance parameter and eliminated in making inferences about β by relying on the use of the so–called partial likelihood. Note that once $S(y)$ and β have been estimated by $\widehat{S}(y)$ and $\hat{\beta}$ respectively, we can use

$$\hat{S}^{\exp(\hat{\beta}x')}$$

as a predictive survival function once x is specified.

The second approach is based on the accelerated failure time (SFT) model, “time” in its name reflecting its orgins when time–to–survival rather than load–to–survival dominated research on this topic. We will for expository simplicity continue here to refer to it as the AFT model. This model, which has an appealingly simple interpretability, assumes that the load is modified also in a way that is proportional to the x hazards, more precisely Y is transformed as

$$Y = e^{\beta x'} Y_0$$

where Y_0 would represent a baseline load–to–failure level.

This report compares the two approaches albeit with one important modification. Instead of x, all of whose elements cannot be recorded for practical reasons, we are forced to use in our analysis a censored version of x as described in Section 3. With that modification, that analysis points to a superiority of the AFT approach over the PH approach. That leads in Section 7 to what we consider to be an enhanced Bayesian version of the AFT, namely a Bayesian version of AFT. However our conclusions are very tentative given the small sizes of the only samples we had to work with.

Still a third approach, which is not an adaption of survival analysis methods is given in a companion paper [20], is based on a purpose built hierarchical Bayes framework. A comparison of these two approaches will be undertaken in future work with larger samples.

For completeness, this report begins with a summary in Section 2 of the background theory needed for the work to follow. Section 3 described the data that were collected through destructive testing in the FPInnovations laboratory and used to help in the development of our theory. Section 4 looks at what might be learned from the data about load–to–failure models. Modeling proceeds from there in Sections 6 and refsect:nonparametricsurv with parametric and semiparametric approaches. The Bayesian version of AFT appears in Section 7. Our summary and conclusions appear in Section 8. The report wraps up with technical details presented in its appendices.

2 Background theory

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), a parametric model (the AFT model) and a Bayesian AFT model (in Section 7) for analyzing survival data .

2.1 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. One of the parameters - the shape parameter - allows it to resemble a variety of other distributions, such as the normal, lognormal, and exponential distributions. That flexibility also allows it to model experimental results, making it a powerful tool in lumber properties research.

The three-parameter Weibull distribution [11] is commonly used to characterize lumber strength. The density function of the Weibull is

$$f(x;\kappa,\lambda,\theta)=\frac{\kappa}{\lambda}(\frac{x-\theta}{\lambda})^{\kappa-1}\exp[-(\frac{x-\theta}{\lambda})^{\kappa}], \tag{2}$$

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

$$F(x;\kappa,\lambda,\theta)=1-\exp[-(\frac{x-\theta}{\lambda})^{\kappa}]. \tag{3}$$

Methods are needed 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.

2.2 The Kaplan-Meier estimator

The Kaplan-Meier (K–M) [16] estimator, estimates the survivor function $S = P_r(Y > y)$. To describe it we let $y_{(i)}$ denote the ith distinct ordered observation and let it be the right endpoint of the interval I_i, $i = 1, 2, ..., n$. Next let n_i denote the number of unbroken pieces for loads below level $y_{(i)}$, while d_i denotes the number broken at level $y_{(i)}$. Then the K–M estimator of the survivor function is defined by

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

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

Compared with a parametric methods, probability statements obtained from most nonparametric methods of inference are exact regardless of the shape of the population distribution from which the random sample was drawn. However these method has several shortcomings such as low power and a lack of available software. Fortunately, the R function survfit can calculate the K–M survival estimators.

2.3 The Cox proportional hazards (PH) approach

Let Y represent the load–to–survival and the survival function be $S(y) = P_r(Y > y)$. The latter may be represented by means of its hazard function, which represents the instantaneous risk of breaking at 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 commonly rely on the hazard function or the log hazard instead of the survivor function itself. Survival analysis typically examines the relationship between the conditional survival distribution and covariates on which it is conditioned. Most commonly, this examination begins with a linear–like model for the log hazard function or equivalently, a multiplicative model for the hazard. For example, a parametric model for the log hazard function of 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,

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

where the x's are the covariates. The constant α in this model for the exponential represents a constant log-baseline hazard, since $\log h_i(y) = \alpha$ (or $h_i(y) = e^{\alpha}$) when all of the x's are zero. More generally, 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 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 thus 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 inferences.

2.4 The accelerated failure time (AFT) approach

The accelerated failure time method is an alternative to the Cox PH method for estimating the survival function. The former models 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 method, the AFT method is intended to provide a description of the relationship between survival probabilities and covariates.

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

Based on the relationship between the survival function and hazard function, the hazard function for an specimenwith 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–to–failure Y is given by:

$$\log Y_i = \mu + \alpha_1 X_{1i} + \alpha_2 X_{2i} + ... + \alpha_p X_{pi} + \sigma\varepsilon_i, \tag{4}$$

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, so the approach provides a good deal of modeling flexibility.

3 The data

Our data come from testing experiments conducted by our research group, notably Samuel Wong, at a FPI/Forintek laboratory located in Vancouver, British Columbia, with assistance from the Lab's technical support staff. These experiments have yielded the two samples of lumber, each of size 98, used in the research described in this report. The first sample consists of bending strength or rupture (MOR) test data and the second, ultimate tensile strength or tension (MOT) data. In these two tests, as loads (bending or tension stress) are increased, each piece remains intact ("survives") until it reaches its critical load at which point it fails. The values of MOR and MOT are computed from the recorded failure loads (unit: psi 10^3) at the point where the stress is applied (usually at a random location near the center in the case of MOR and at the ends of the piece in the case of MOT). The break occurs at a random location along the board. Figures 1 and 2 show pieces of lumber in the bending and tension test devices.

Figure 1: The bending test. Notice that the piece of lumber is held firmly in place by the vertical suspenders at its ends. A load is applied in the center simultaneously at two sites by vertical arms that are hydraulically lifted by the platform to which they are attached. The whole process is automated and the load–to–failure recorded by the electronic equipment on the right.

Stiffness or elasticity (MOE psi 10^6) is also measured in a non-destructive way as each of the above two tests are carried out. 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 MOT. Interest lies in the relationships amongst MOR, MOT and MOE.

Before the tests are carried out, the specimens are conditioned to achieve a specified moisture content and subsequently they are examined by a professional grader who records the observable features referred to in Section 1. In particular the characteristic deemed most likely to cause the specimen to fail during the test, referred to as its maximum strength reducing characteristic (MSRC), is recorded in a coded form. Examples of such characteristics are ‘knot’, “grain”, “shake” and “split”. The MSRC is the feature that the Grader expects to be the cause of failure. In contrast, the failure code (FC) is the characteristic the Grader visually judges to have caused the piece to fail after testing. These two features can be identical. Although the specimen may present numerous other features, for practical reasons only these two are recorded along with the one that determines the grade of the specimen. The latter may not be a strength reducing feature. For example “wane”, which are remnants of bark left on the corners of the specimen, may lower the commercial value of the piece without affecting its engineering characteristics.

We see an aspect of reliability analysis in this context that would usually not be seen in survival analysis, namely censorship of the covariates rather than the response. At the same time, the censoring mechanism, although somewhat subjective, is clearly related to the response. In contrast, in medical applications of survival analysis for example, the inclusion of all covariates would normally be required, even though some might be missing due to chance. Moreover, covariate would often be missing, but not always, at random for reasons not related to the response of interest.

Figure 2: The tension test. Here the piece of lumber is gripped at at its ends. A load is applied laterally at those end points by by the hydraulically controlled suspenders The whole process is automated and the load–to–failure recorded by the electronic equipment on the left.

Ten different causes of failure are found in our data, notably “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). The coding system of measurements[1, 2, 4] (e.g. MSRC) is quite elaborate as can be seen in Table 1.

Table 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, 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 severest one.

Table 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$_{\copyright}$ data 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 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, one for each of MOR and MOT, each being of size 98. For each specimen in these two samples, we have the feature

Table 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	

measurements shown in Table 4.

Table 4: Description of measured features in the two samples on which model development in this report are based.

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 5 presents a section of the layout of dataset for the bending and tension tests.

Table 5: This table shows a transcription of an illustrative section of the dataset including the preliminary assessments of visual features made by a professional Grader.

♯	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
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.

The small size of the dataset led the authors to convert MSRC into meaningful covariates using only MSRC1 in as much as it represented the severest 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).

Quantifying the numerical variable – “ksize” was done in accordance with the following rules:

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 a few cases.
3. The value of ksize for other defects is 0.

Therefore, for bending data, we have the variables defined as in Table 6, and the layout of bending data with transformed covariates is in Table 7.

4 Exploratory data analysis

Exploratory data analysis (EDA) is preliminary detective work done to see problems and patterns in the data that informs subsequent model development.

Table 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

4.1 Graphical visualizations of strength properties

Histograms. For bending and tension tests, we 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, 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 suggests no significant difference between the two MOE's in the two tests.

Strength and covariate relationships. 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 4 displays MOR against continuous variables. The non-parametric curve using lowess shows the pattern of association between the MOR and 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 5 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

Figure 3: Distributions of the strength properties data.

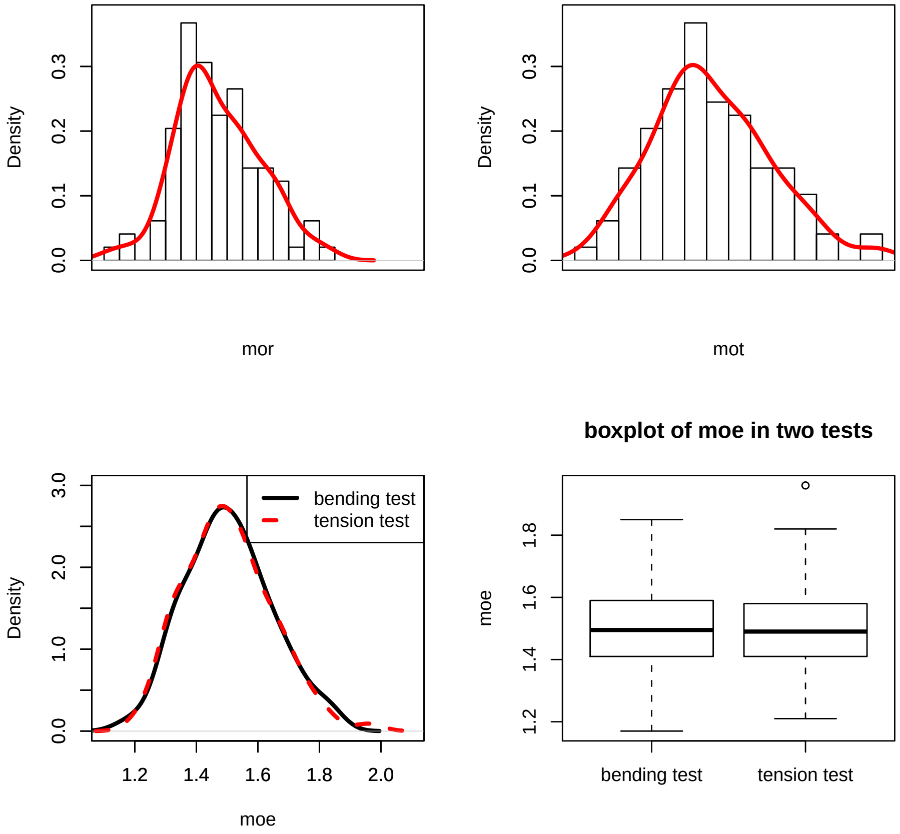


Table 7: Transformed bending data.

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
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.

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).

4.2 Univariate approaches to modeling the distributions

In reliability analysis for lumber, much interest lies in the relationships between the MOR, MOT and MOE. Moreover in lumber strength testing, interest focuses specifically on the pieces with the lowest strengths, for example the subpopulation with strengths below the 5^{th} percentile, say $\zeta_{0.05}$. We will use both parametric and nonparametric approaches to estimate $\zeta_{0.05}$ 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 $\rho = \zeta^R_{0.05}/\zeta^T_{0.05}$. The latter is of particular interest because its estimate can be used to estimate by multiplication, the 5^{th} bending percentile from an estimate of that for tension, error bands included as we will see below.

Univariate Weibull distribution. Assuming Weibull population distributions and independent samples, the three parameters in (1.1) can be

Figure 4: MOR against continuous variables, with a lowess smooth curve.

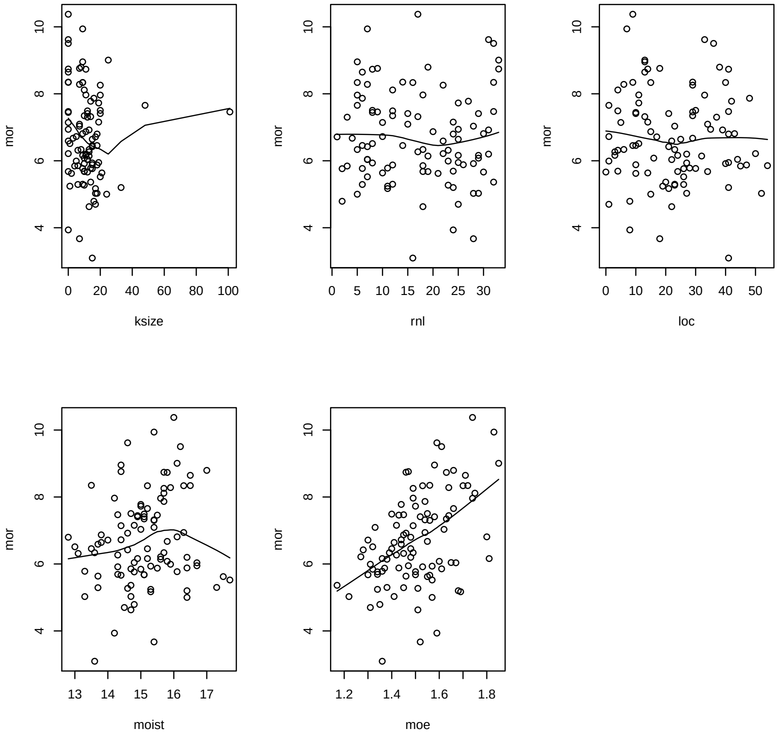
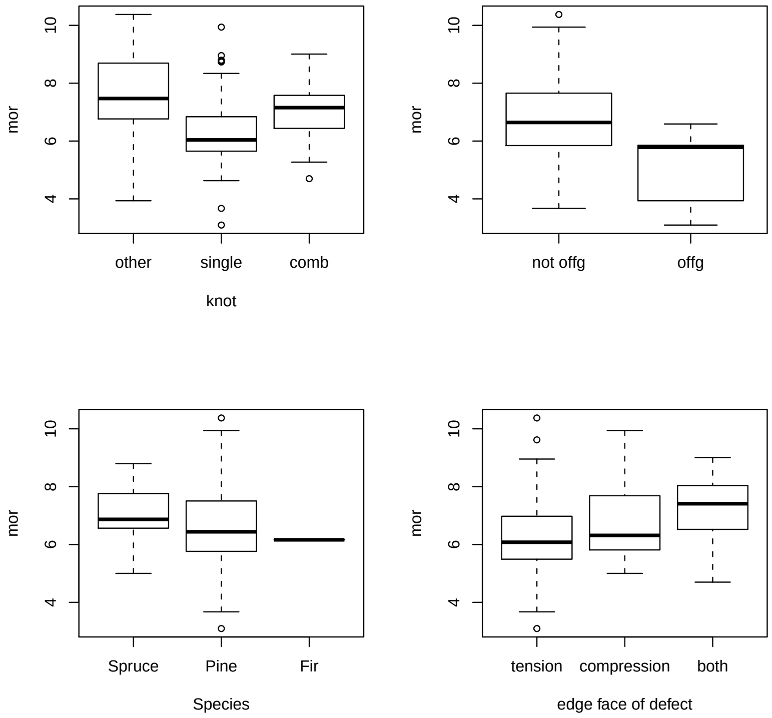


Figure 5: MOR against categorical variables.



estimated using maximum likelihood obtained through numerical optimization. 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 8 displays the maximum likelihood estimates (MLEs) of parameters in the three parametric Weibull distribution for the MOR data. .

Table 8: MLEs of univariate Weibull parameters for the bending (MOR) data.

Quantity	Value	Standard Error
$\widehat{\lambda}_1$ ($psi \times 10^3$)	4.726	0.590
$\widehat{\kappa}_1$	3.325	0.511
$\widehat{\theta}_1$ ($psi \times 10^3$)	2.460	0.537

Similarly, Table 9 displays the maximum likelihood estimates (MLEs) of parameters in the three parametric Weibull distribution for the MOT data.

Table 9: MLEs of univariate Weibull parameters for the tension (MOT) data.

Quantity	Value	Standard Error
$\widehat{\lambda}_2$ ($psi \times 10^3$)	3.610	0.362
$\widehat{\kappa}_2$	2.556	0.335
$\widehat{\theta}_2$ ($psi \times 10^3$)	0.901	0.297

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

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

the population 5$^{\text{th}}$ percentile $\zeta_{0.05}$ is given by solving

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

to get

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

Thus the estimated ratio is given by

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

The invariance property of MLEs implies that 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 10.

Table 10: 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)

Univariate Kaplan–Meier estimators. To explore the distribution of MOR, MOE and MOT, we can also use the non-parametric Kaplan-Meier estimators of their survival functions $S(y) = P_r(Y > y)$. Using the R function survfit, we plot the Kaplan-Meier curves of MOR (left) and MOE (right) with 95% error bands in Figure 6.

As well as the KM estimators of $P_r(MOR > mor)$ in Table 11, where n_i denotes the number below mor_i while d_i denotes the number that are recorded as failing at load level mor_i. Similarly, the KM curves for MOT (left) and MOE (right) with 95% error bands in the tension test are shown in Figure 7. as well as the KM estimators of $P_r(MOT > mot)$ in Table 12, where n_i denotes the number at risk below mot_i while d_i denotes the number recorded as failing right at mot_i.

Estimators of the population 5th percentile implied by KM approach. The KM method easily generates estimates of the population's 5th percentile ($\widehat{\zeta^R_{0.05}}$, $\widehat{\zeta^T_{0.05}}$ and $\widehat{\zeta^E_{0.05}}$) using a standard method [16]. Table 13 and 14 show the KM estimators of percentiles for bending and tension, respectively. They show the KM estimators to be pretty close to the estimators for the Weibull distribution given in Table 10. However not surprisingly the standard errors of the KM estimators are relatively larger than the ones

Figure 6: Kaplan – Meier nonparametric estimates of survival curves for MOR (left) and MOE (right) in the bending test.

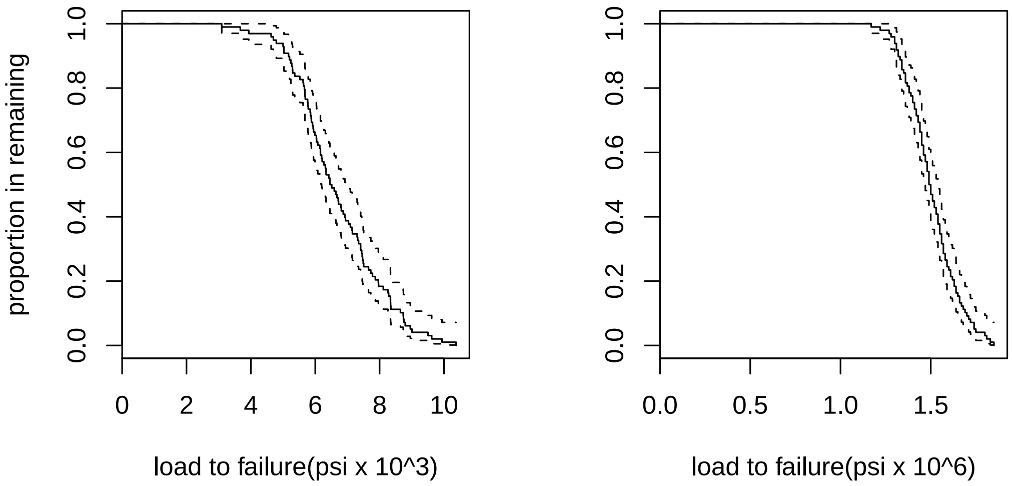


Figure 7: Kaplan–Meier curves for MOT and MOE in the tension test.

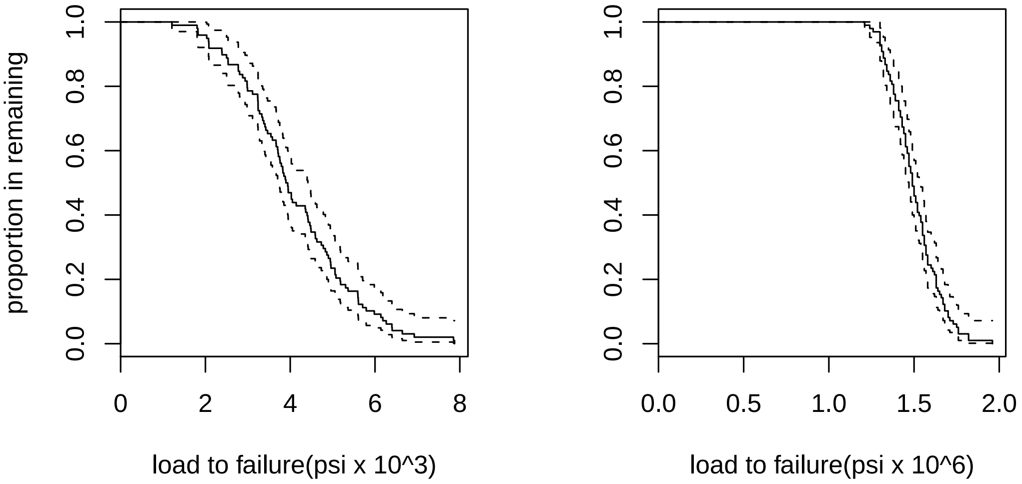


Table 11: Kaplan–Meier estimation of survival function for MOR.

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
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Table 12: Kaplan–Meier estimates 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
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Table 13: Kaplan–Meier 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 14: Kaplan–Meier 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) $(psi \times 10^6)$

produced by the Weibull approach. This is because in adopting the KM approach we have not expressed the same high level of certainty about the shape of the survivor function as when we asserted our belief that it had a Weibull shape with just three unknown parameters. Also, the two estimated values of $\zeta^E_{0.05}$ for bending and tension are almost the same, which suggests the lumber in the two samples are homogeneous in terms of elasticity.

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

A Bivariate Weibull distribution. The density function of a bivariate Weibull is given by

$$\begin{aligned} f(x, y; \kappa_1, \lambda_1, \theta_1, \kappa_2, \lambda_2, \theta_2, \delta) &= \frac{\kappa_1}{\lambda_1}(\frac{x-\theta_1}{\lambda_1})^{\frac{\kappa_1}{\delta}-1}\frac{\kappa_2}{\lambda_2}(\frac{y-\theta_2}{\lambda_2})^{\frac{\kappa_2}{\delta}-1} \\ &\times\{(\frac{x-\theta_1}{\lambda_1})^{\frac{\kappa_1}{\delta}} + (\frac{y-\theta_2}{\lambda_2})^{\frac{\kappa_2}{\delta}}\}^{\delta-2}\{[(\frac{x-\theta_1}{\lambda_1})^{\frac{\kappa_1}{\delta}} + (\frac{y-\theta_2}{\lambda_2})^{\frac{\kappa_2}{\delta}}]^{\delta} + \frac{1}{\delta} - 1\} \\ &\times \exp\{-[(\frac{x-\theta_1}{\lambda_1})^{\frac{\kappa_1}{\delta}} + (\frac{y-\theta_2}{\lambda_2})^{\frac{\kappa_2}{\delta}}]^{\delta}\}. \end{aligned} \tag{5}$$

A feasible method for fitting that distribution has already been developed [12]. We first estimate the shape (κ), scale (λ) and location (θ) parameters from the two marginal distributions as above. Given these estimates $(\kappa_1, \lambda_1, \theta_1, \kappa_2, \lambda_2, \theta_2)$, we can find the dependence parameter estimate δ using maximum likelihood by numerical optimization. We get the log–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,

$$\begin{aligned}\overline{F}(x,y) &= P[X > x, Y > y] \\ &= \exp\{-[(\frac{x-\theta_1}{\lambda_1})^{\frac{\kappa_1}{\delta}} + (\frac{y-\theta_2}{\lambda_2})^{\frac{\kappa_2}{\delta}}]^{\delta}\}, 0 < \delta \leq 1 \qquad (6)\end{aligned}$$

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).

A 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 the $\{X_i\}$ and $\{Y_i\}$ are the observed loads, it is natural to estimate $Pr(X \geq x, Y \geq y)$ by the empirical survival function:

$$\widehat{S}(x,y) = n^{-1}\sum_{i=1}^{n} I(X_i \geq x, Y_i \geq y) \qquad (7)$$

The asymptotic variance of this estimator is given by:

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

Evidence in favor of this estimator has been published [15]. Note that since our data are uncensored, the problem we face is more easily solved than if they were censored.

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 15: These results suggest the two approaches yield estimates that are pretty close to one another. In other words both the parametric and more unrestricted nonparametric survival analysis methods seem to work well. The 3-dimensional scatterplot for each method shown in Figure makes it apparent that the two estimators are nearly identical.

4.4 Testing for differences between the estimated distributions

A graphical approach using the KM estimator and log-rank test. A central objective of the study described in this report is the characterization of the relationship between strength and its covariates. For a categorical

Figure 8: A visual comparison of the bivariate Weibull and KM estimators.

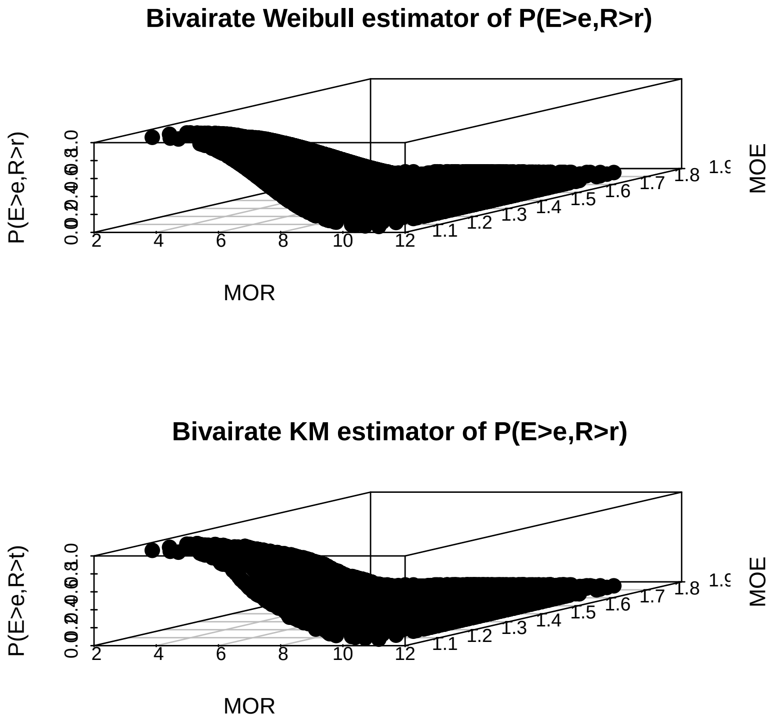


Table 15: A comparison of the estimators of the survivor function provided by the bivariate Weibull and by bivariate Kaplan–Meier approaches.

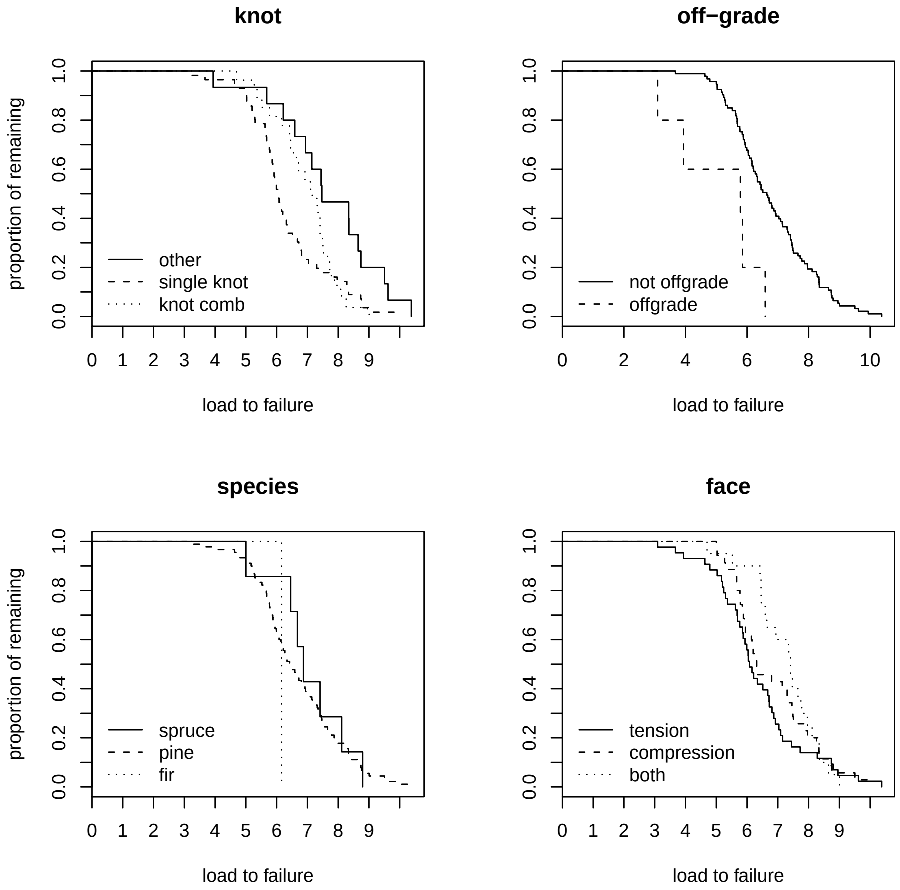
$(e, r)(psi \times 10^6,\, psi \times 10^3)$	$\widehat{S}(e, r)$ by (4)	$\widehat{S}(e, r)$ by (5)
(1.65, 6.042)	0.1327	0.1330
(1.65, 6.590)	0.1122	0.1146
(1.65, 7.459)	0.1020	0.0762
(1.36, 7.867)	0.2143	0.1960
(1.36, 4.791)	0.8061	0.7952
(1.36, 5.664)	0.7041	0.6840
(1.36, 5.363)	0.7347	0.7318
(1.36, 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
.	.	.
.	.	.
.	.	.

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 9 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.

Figure 9: KM curves of MOR against categorical covariates. Notice that unlike the curves for ‘species’ and ‘face’, those for ‘offg’ and ‘knot’ are quite parallel.



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 'offg' may be the important predictors for MOR.

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 10 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 – the two density curves of MOE in the two cases are almost identical.

4.5 Exploring the association between MSRC and FC

Recall that MSRC is the grader's assessment before testing of a piece of lumber of the most likely cause of failure, while FC is the characteristic visually judged by the grader to have caused the piece to fail after testing. They would be the same if the failure occurs because of MSRC.

The two-way contingency table. If two variables are measured at categorical levels (eg. nominal or ordinal), we assess their relationship by cross–tabulating the data in a two-way contingency table [3]. 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 16 is a simple example of a hypothetical contingency table that cross–tabulates 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 17, 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.

Test of independence (chi-square and related tests). For ease of understanding, let's take the data in Table 16 for example. If the characteristics *Gender* and *Answer* were not associated (the null hypothesis of

Figure 10: KM curves of MOE in the two cases. Notice that the curves for two tests are almost identical.

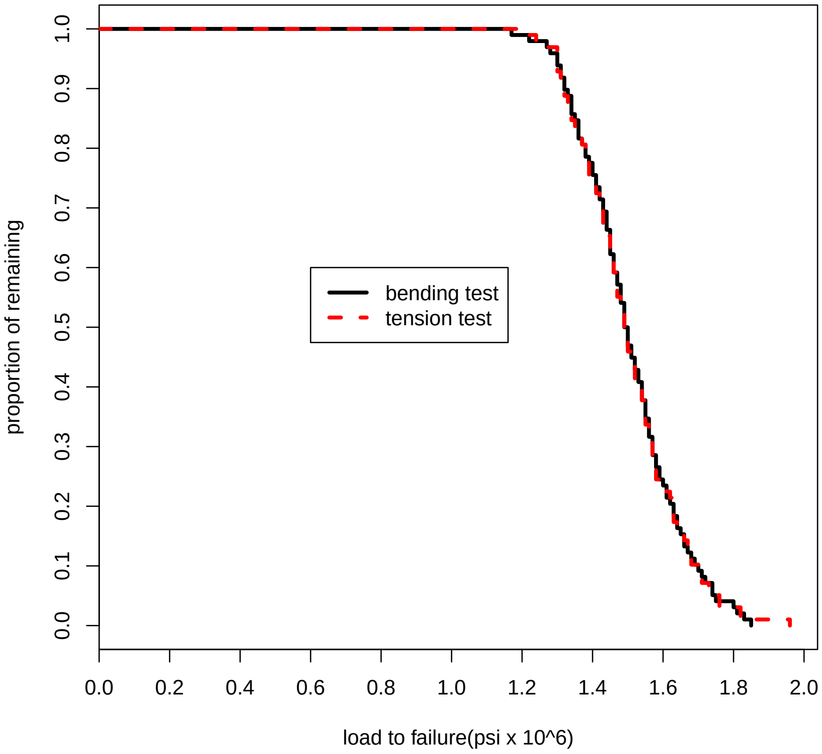


Table 16: Example of a hypothetical two-way contingency table. Here we see 'gender' being broken down by a subject's answer to an examination question (1= 'Yes'; 0= 'No').

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

Table 17: Two-way contingency table for 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

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.

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, $Cramer's\ V$[3] 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 $2{\times}2$ 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.

5 Semi–parametric survival models

Let's first fit a semi-parametric survival regression model, namely the Cox PH model [8, 9, 18]. Since in the Cox PH 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 Cox PH model for MOR with covariates: 'knot', 'ksize', random number location ('rnl'), off-grade indicator('offg'), location of defect('loc'), face of defect (0 means on the tension edge, 1, on the compression edge and 2, on both edges) , 'species', 'moisture' and 'moe'.

5.1 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 the model.

5.2 Variable selection

First using R-code, 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 18 presents a summary of the results as generated by the R-code, $summary(coxph.fit1)$. Thus, we can see the covariates ' knot', 'off-grade' and 'moe' are significant at level 0.05. We next apply two alternate methods to select the best possible model.

Method I: ***step*** **() to select the best model according to AIC statistic.** Table 19 shows p-values corresponding to variables selected by $step(coxph.fit1)$.

From Table 20, we see that the stepwise method chooses 3 covariates: 'knot', 'off-grade' and 'moe'.

Table 18: Summary of the initial application of the Cox PH model to the bending data.

	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	$***$

Method II: Single term deletions. Table 21 displays the result of applying a single term deletions method using the code $drop\,1\,(coxph.fit1, test = \text{“}Chi\text{”})$. 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.

Comparing nested models. Methods I and II have generated the same reduced model. We now compare that model to the initial full model by means of the likelihood ratio test (LRT). Symbolically we may describe a model using R-code as follows:

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

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

We can compare these models using R by means of the command $anova(cox1, coxph.fit1)$, which gives the results in Table 22. We conclude that 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.

Table 19: 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 20: p-values of covariates in the model selected by *step* (). The significant factors are the ones where at least one * has been affixed.

	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	$***$

Checking for Interaction. Here we use $step(cox1, \sim .^2)$ and the results appear in Table 23. Adding an interaction term increases the AIC so we conclude there is no need to add interactions. Our final model is labelled cox1.

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

Table 24 presents the result of applying the R function $summary(cox1)$. Based on that we can make the following comments:

1. The estimated coefficient for having a single knot as the MSRC is 1.049 and it has a very significant, small p-value. Hence, fixing other covariates, the hazard ratio between the lumber with a single knot as MSRC and the one having a knot combination as the MSRC is

Table 21: 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	$***$

Table 22: This table compares the fits of the full model with those generated by the reduced model produced by Methods I and II.

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

exp(1.049)/exp(0.692) = 2.85554/1.99684 = 1.43, which means that the former are 1.43 times more likely than the latter to fail at any given load given survival to that load–to–failure. Similarly, the hazard ratio between the lumber with a single knot as MSRC and the ones with defects other than 'knot' is 2.856 with a similar interpretation to the previous case. This is consistent with the side–by–side boxplots of 'mor' against 'knot' in the exploratory data analysis (EDA), which shows that pieces of lumber with a single knot as its MSRC possess a lower 'mor' than ones with the other two categories of 'knot' as MSRC.

2. The estimated coefficient for the off–grade pieces of lumber is 1.452, and exp(1.452) = 4.273, which means the off–grade pieces of lumber have load–to–failure levels 4.273 times lower than standard ones. This is also consistent with the conclusions suggested by our EDA.

3. Fixing the other covariates, pieces of lumber with a higher 'moe' level have a decreased hazard of failure than ones with lower 'moe' levels.

Table 23: In this table we show the results of including interaction terms in the model.

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 24: Here we see the p-values for the various covariates in the final model for the bending data.

	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	$***$

This seems quite reasonable.

5.3 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 [17], the deviance residuals [19] and the Cox-Snell residuals [10]. Then we will talk about influence assessment and strategies for analysis of nonproportional data.

Checking for the proportional hazards assumption. The main assumption of the Cox PH models is that of proportional hazards [16]. Proportional hazard means that the hazard function of one specimen is proportional to that of any other, i.e., the hazard ratio is constant over increasing load. Several methods are available for assessing that assumption. The k^{th} Schoenfeld residual [17] defined for the k^{th} specimen on the j^{th} explanatory variable x^j is given by

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

where δ_k is the k specimen's censoring indicator, x_k^j is the value of the j^{th} explanatory variable on the k^{th} specimen,

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

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 of the estimated variance-covariance matrix of the k^{th} Schoenfeld residual and the k^{th} Schoenfeld residual, so that the k^{th} 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 implies the PH assumption is violated.

As mentioned above, the R-function cox.zph computes a test for each covariate, along with a global test for the model as a whole. The R-code $cox.zph(cox1)$ gives the results in Table 25.

Table 25: This table reports the results obtained by application of the R-code $cox.zph(cox1)$. They provide a test of the model as a whole as well as for each individual covariate. The columns of the matrix are from left to right, the correlation coefficient between transformed survival time and the scaled Schoenfeld residuals, a chi-square statistic, and the two-sided p-value. The NA is just a placeholder, there since is no appropriate correlation for the global model.

factor			p-value
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

These results show now no statistically significant evidence of non-proportional hazards for any of the covariates. The global test is also not quite statistically significant at the 5% level. 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 as in Figure 11. Interpreting 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 ± two-standard-error envelopes around the fit. Systematic departures from a horizontal line indicate 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 we see what appears to be a trend in the plot for 'knot', with the 'knot' effect 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 test based on the R function cox.zph.

An alternative and less sensitive way of assessing the proportional hazards assumption is to plot $\log[-\log S(r)]$ vs $\log(r)$ as in Figure 12. We see parallel plots for 'offg' but nonparallel ones for 'knot' in confirmation of what we learned above from the Schoenfeld residual plots. These results cast doubt

Figure 11: Scaled Schoenfeld residuals against load-to-failure. The solid line represents a spline smooth of those residuals with two sigma limits around it. Note the apparent trend for ‘knot’ unlike the other two, pointing to a possible problem with the proportional hazards assumption for this case.

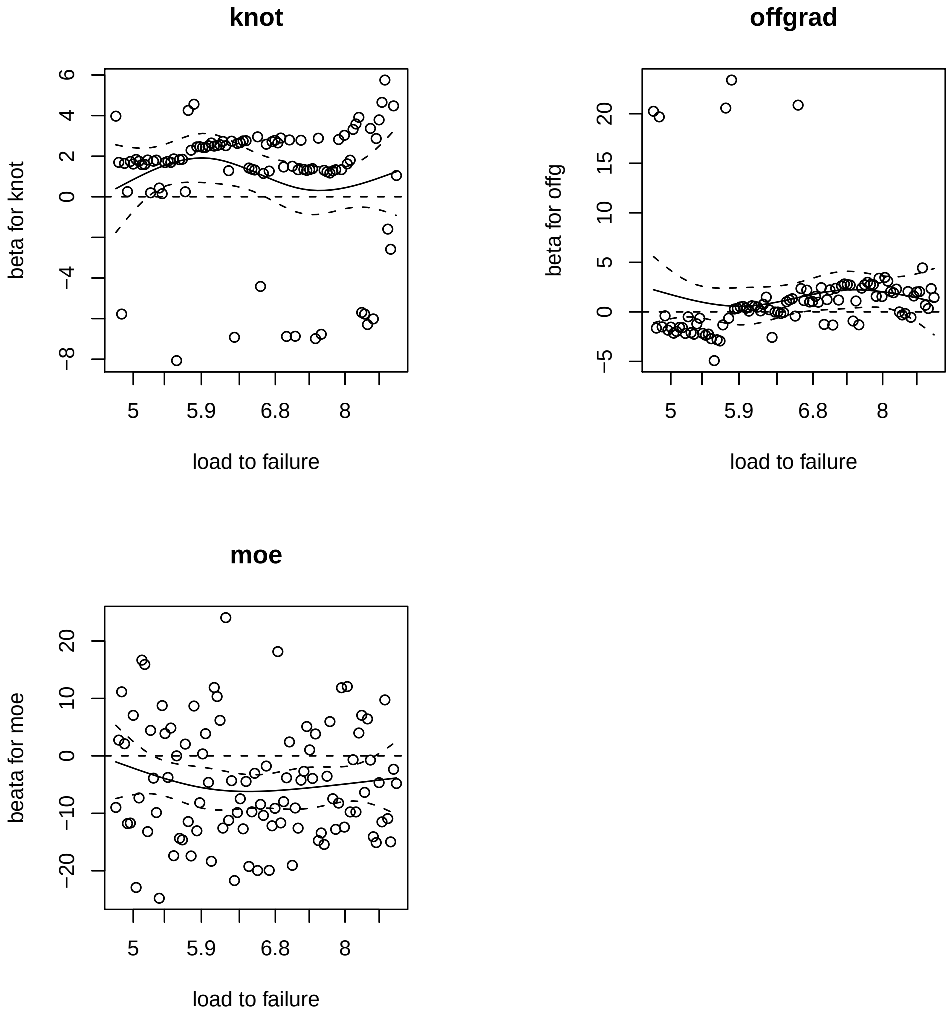
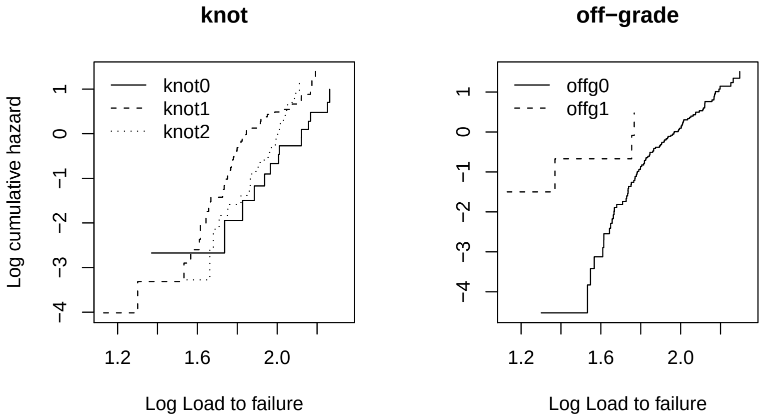


Figure 12: Graphical check of the proportional hazard assumption. We see a slight lack of parallelism in the plots for the cases associated with ‘knot’. The situation with off–grade is a little less clear but at any rate does not give us grounds for rejecting the assumption of proportional hazards for that factor.



on the appropriateness of the Cox PH model, a major finding in our analysis.

Assessing goodness-of-fit. The i^{th} 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. The quantity $r_{Ci} = -\log \hat{S}_i(t_i)$ 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 can plot the cumulative hazard function $H_R(r_{Ci})$ against Cox-Snell residual r_{Ci} to check the fit of the model. We should see a straight line with unit slope and zero intercept if the fitted model is correct. [Note however, that the Cox-Snell residuals will not be symmetrically distributed about zero and cannot be negative.] We see that plot in Figure 13. 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.

Checking for outliers. The i$^{\text{th}}$ deviance residual is defined by

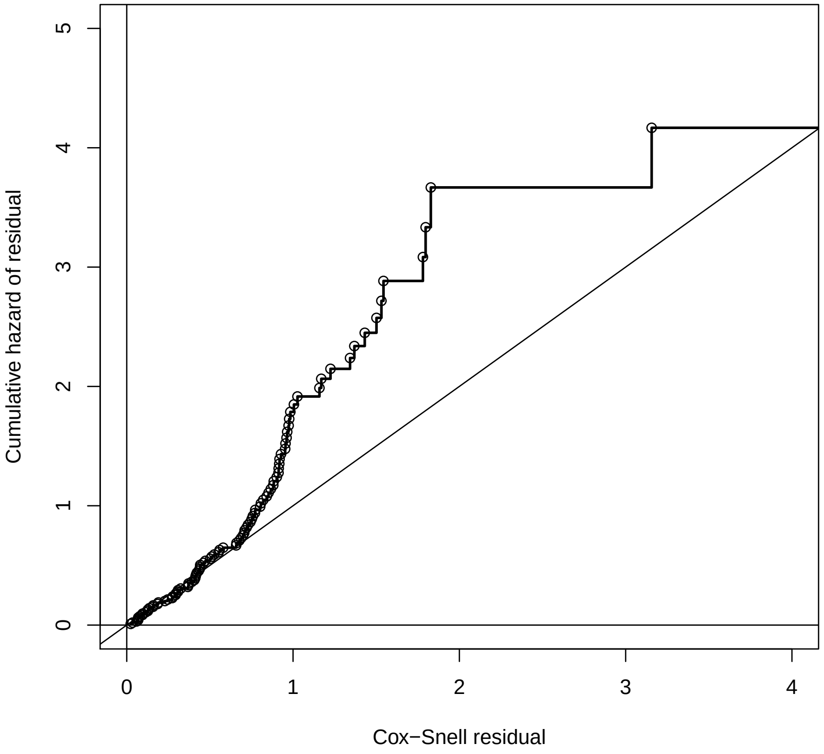
$$r_{Di} = sign(r_{m_i})\sqrt{-2\{r_{m_i} + \delta_i \log(\delta_i - r_{m_i})\}},$$

where: the sign function sign(), which is 1 or -1 according as $r_{m_i} > 0$ or $r_{m_i} < 0$; $r_{m_i} = \delta_i - r_{Ci}$ is the martingale residual; δ_i is 1 or 0 according as the observation is uncensored or censored.

In a fitted Cox PH model, the hazard of failure for the i$^{\text{th}}$ specimen at any load 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 specimen relationship to the model. Potential outliers will have deviance residuals whose absolute values are 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

Figure 13: Cumulative hazard plot of the Cox-Snell residual for Cox PH model. Validity of the Cox PH model would mean that the cumulative hazard plot would lie along the line which it does not, thus casting doubt on that model.



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 14. They show only one possible outlier, but none of them seems to be well–scattered about zero. Therefore overall, they lead to further doubt about the adequacy of the Cox PH model.

Influential observations. Figure 15 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 show the influence of individual observations on the estimated regression coefficients for each covariate. Most of the changes in the regression coefficients are less than 0.1 se.'s of the coefficients and all others are less than 0.2 se.'s. Data sets where the influence plot is tightly clustered around zero indicate an absence of influential observations and that is the case with these data leading us to conclude that none of the observations are outliers.

Violation of the proportional hazards assumption. The analyses described above have led to doubt about the validity of the proportional hazards assumption for the predictor 'knot'. One method for dealing with this problem is to stratify the model by 'knot', meaning that we produce a separate baseline hazard function for each level of 'knot'. However by stratifying the results in this way, we cannot get a hazard ratio for 'knot', its effect being absorbed into the baseline hazard.

The two models are described symbolically below:

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

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

Since the stratified model $cox2$ provides a smaller AIC value than the previous model $cox1$, we 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.

Figure 14: Deviance residuals against the risk score, index and covariates. They show only one possible outlier but their lack of systematic scatter about the zero line casts doubt on the suitability of the Cox PH model.

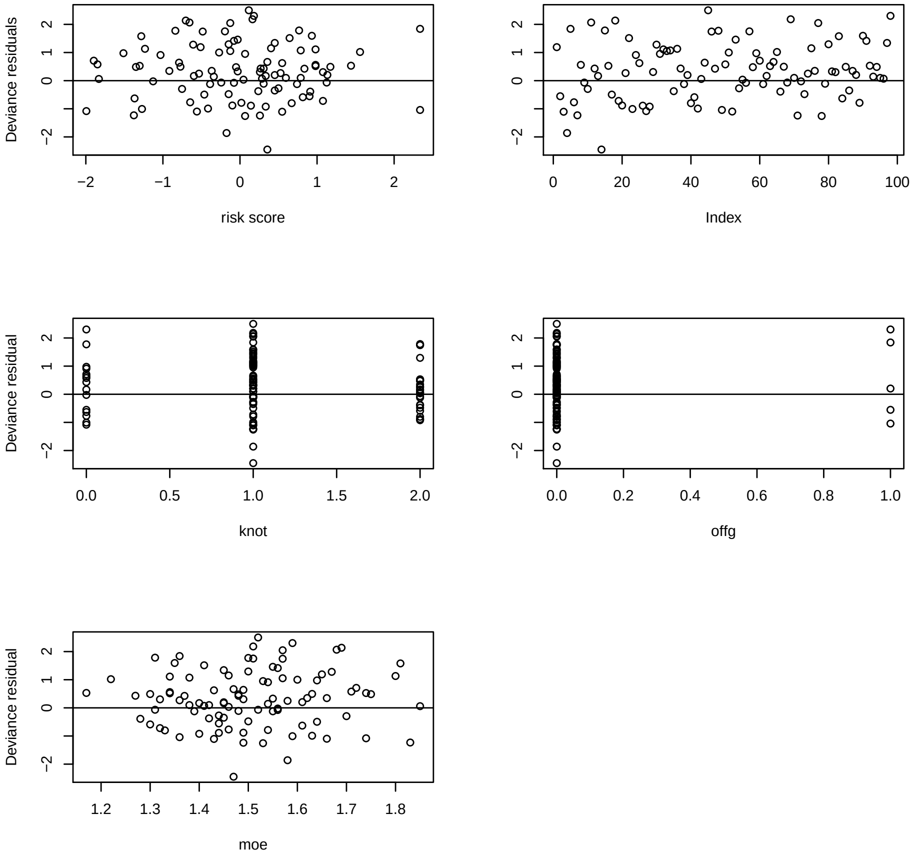
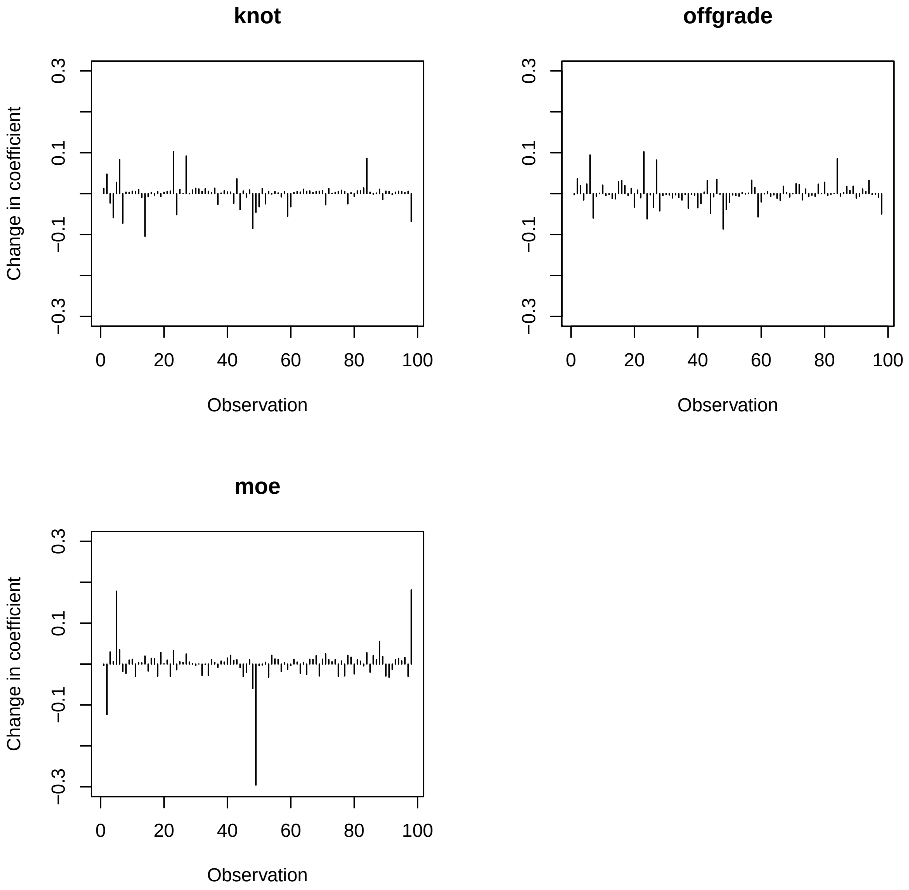


Figure 15: Influence statistics. Since none of these normalized changes in the regression coefficients exceeds 0.1 in absolute value, we see no outliers in these data.



6 Parametric survival models

The accelerated failure time (AFT) model [6, 9, 14] is an alternative to the Cox PH model when the PH assumption is violated. The former is more tractable than the latter for representing the difference in strengths between two groups. The AFT models the direct effect of the explanatory variables on the load–to–failure instead of hazard, as in the PH model. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the corresponding covariate directly on the mean load–to–failure response.

6.1 The distribution of load–to–failure

AFT models are named for the distribution of strength data in our application. The most commonly used ones include the exponential, Weibull, log-logistic, and log-normal AFT models. 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 26, where $Z(p)$ means the p^{th}-quantile from the standard normal distribution.

Table 26: 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

For bending data, we present these four different plots in Figure 16. By comparing the straightness of these lines, we may see that the distribution of bending data is most likely to be one of Weibull, log-normal, or log-logistic. Note that the left hand tail matters most in applications pointing to the Weibull distribution in this case as it offers a better fit in that region. The exponential distribution, a special case of the Weibull would certainly not be acceptable.

6.2 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 27, where η^{-1} is the estimated load acceleration factor, the multiplicative constant which scales that load up or down.

Figure 16: Exploring potential distributions for the load–to–failure. Note that a straight line fit is required for acceptability of a strength distribution, particularly in the left hand tail. This rules out the exponential AFT model.

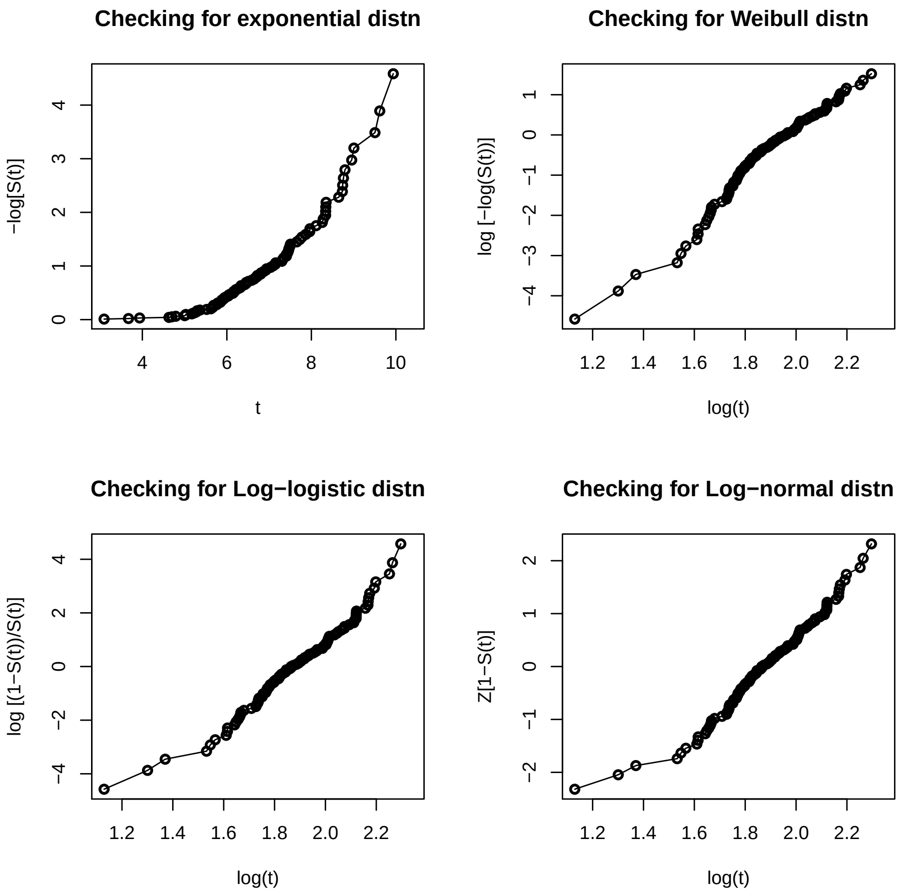


Table 27: Results from fitted AFT models for the bending data. Here we use α stands generically for the coefficient of the explanatory variable. η is the acceleration parameter, and p stands for p–value.

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 28, we compare all these AFT models using the AIC, the smaller AIC the better. The Weibull AFT model appears to be an appropriate AFT model according to the AIC compared to the other AFT models. The exponential model provides the worst fit, which is consistent with the conclusion we draw from Figure 16.

Table 28: The AIC for the the various AFT models under consideration.

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

6.3 Checking the AFT assumption

A preliminary method for assessing an AFT model's performance is through a quantile-quantile (Q-Q) plot. For any value p in the interval (0,100), the p^{th} percentile is

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

Let $t_0(p)$ and $t_1(p)$ be the p^{th} 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 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 give an approximate straight line through the origin. The slop of this line will be an estimate of the acceleration factor η.

For the 3-level categorical covariate 'knot', we have 3 possible pairwise combinations. The Q-Q plot in Figure 17 approximates well a straight line from the origin indicating that the AFT model may be appropriate.

6.4 Model Diagnostics for the AFT Model

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 18. The plotted points mostly lie on a line that has a unit slope and zero intercept, although we see some divergence at the right hand end. In any event, comparing Figures 13 and 18, we 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.

Checking for outliers. Similarly, the plots of deviance residuals against the risk score, index and covariates are given in Figure 19. 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.

Influential observations. Figure 20 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

Figure 17: Q-Q plot for load–to–failure.

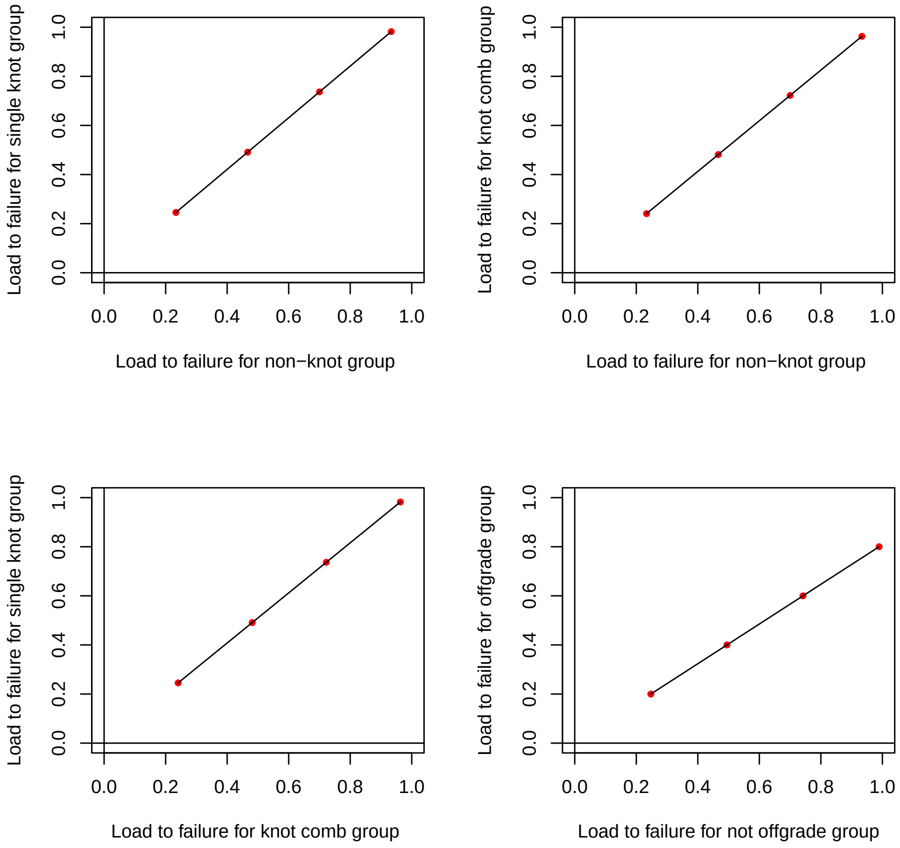


Figure 18: Cumulative hazard plot of the Cox-Snell residual for the Weibull AFT model.

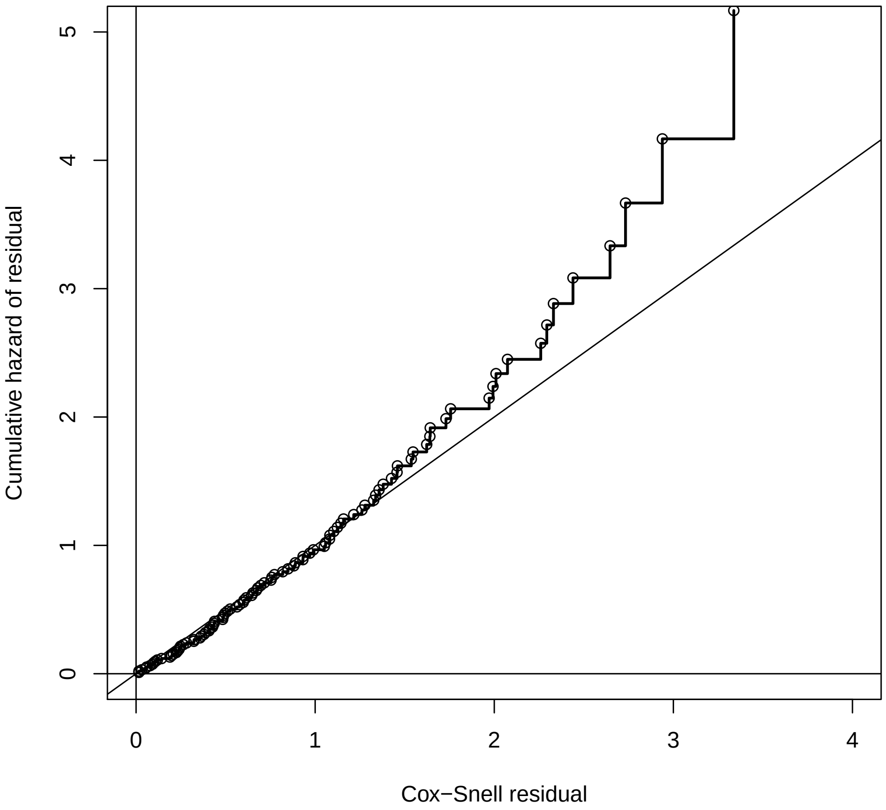
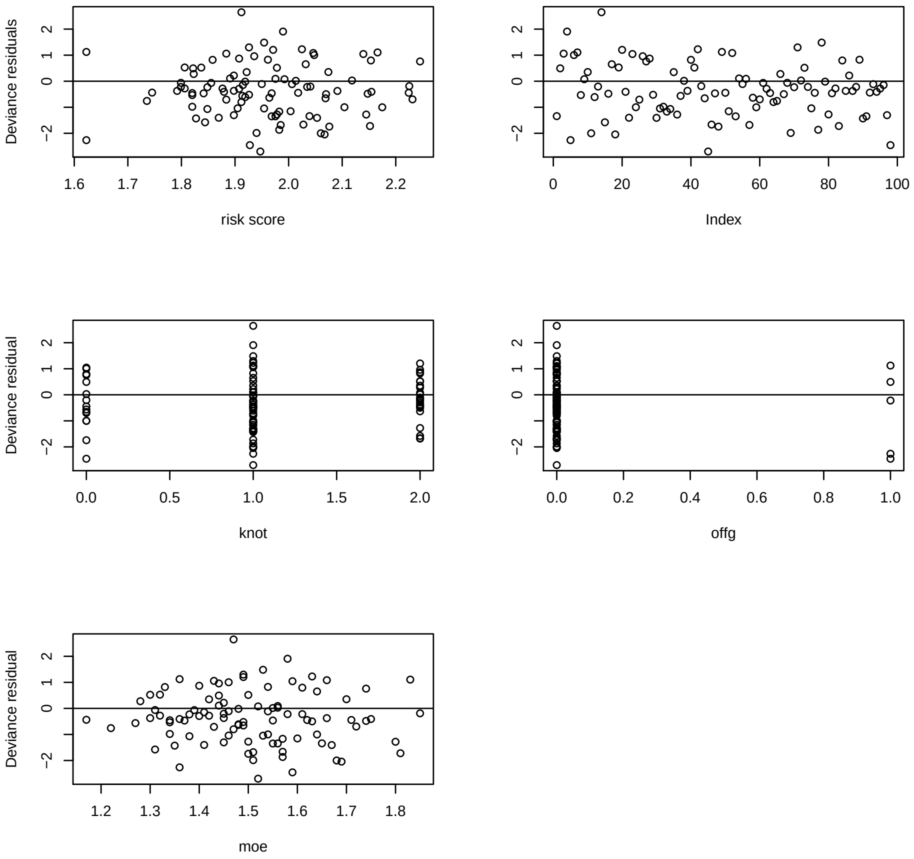


Figure 19: Deviance residuals against the risk score, index and covariates.



of less than 0.04 are of little concern. These plots indicate the influence of individual observations on the estimated regression coefficients for each covariate. Most of the changes in the regression coefficients are less than 0.02 se.'s of the coefficients and all others are less than 0.03 se.'s. Therefore data sets where the influence plot is tightly clustered around zero indicate an absence of influential observations.

6.5 Interpretation of results

Finally, we can fit the Weibull AFT model with only statistically significant covariates – 'knot', 'offg' and 'moe' or in symbolic form:

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

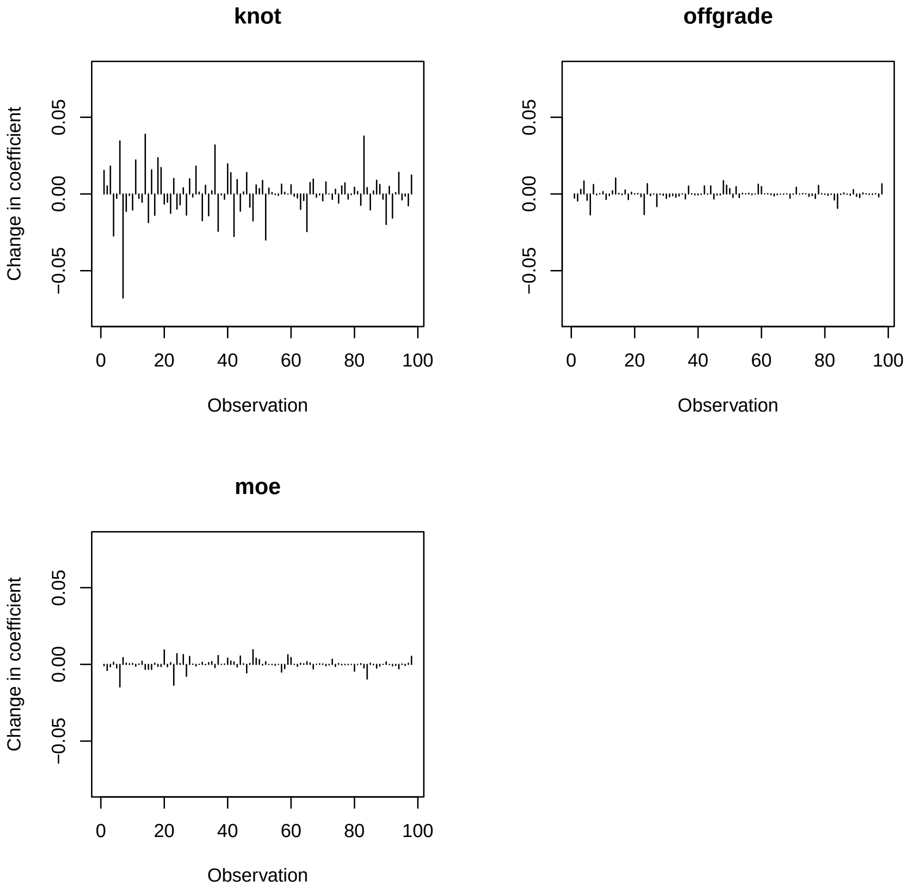
The summary for this model is given in Table 29:

Table 29: 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				

We conclude that the acceleration factor η for 'offg' is 0.81 (less than 1), which indicates that a smaller survival load is more likely for off–grade lumber. The η's for 'knot' are also less than 1 implying 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 an even smaller acceleration factor. The acceleration factor η for 'moe' is 2.03 (more than 1), indicating that a larger survival load is more likely for a piece of lumber with a higher MOE value. These conclusions are consistent with the ones drawn from application of the Cox PH model.

Figure 20: Influence statistics.



6.6 Simulation Study

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.

The 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 were 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, based on the error component of the model. We 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 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 a total of 4×4 (16) different settings for the simulation we conducted. Each simulation involved 1000 replications with a sample size 98.

Simulation results. 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 30 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 30: 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 has estimates of this coefficient, which are closer to the true values than those for the other distributions. The patterns of the 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.

6.7 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 21 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.

Thus overall, the AFT model seems promising for appllcation in this context. Thus in Section 7 we add a Bayesian version of the AFT to enable prior information to be incorporated.

7 A Bayesian version of AFT

Let Y_I denote the event time. The following regression model is assumed

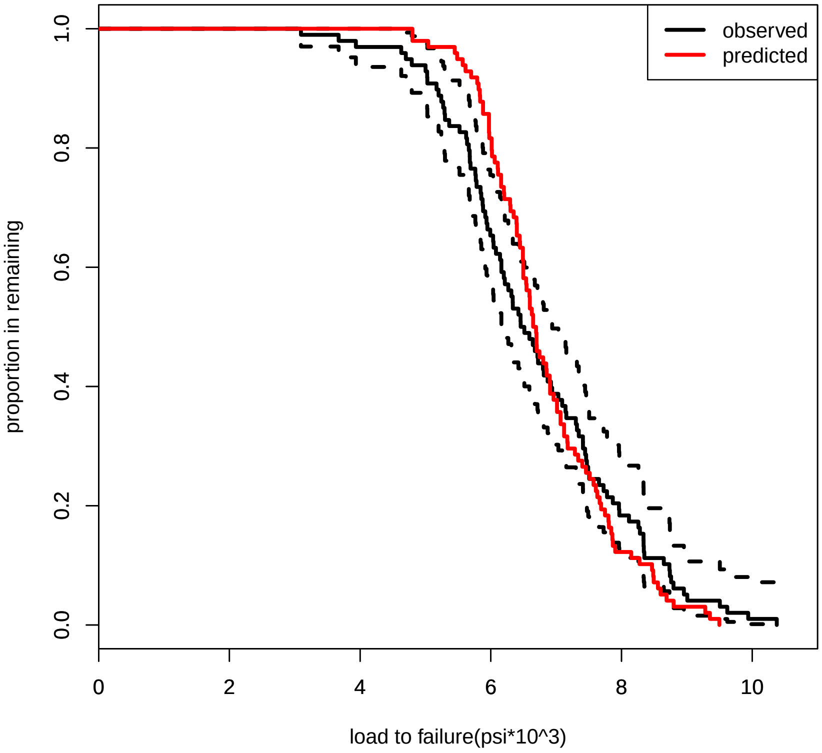
$$log(Y_i) = \beta' x_i + \epsilon_i,$$

where β is the vector of the corresponding regression coefficients and x_i is the vector of covariates. The error term ϵ_i are assumed to be independent and identically distributed with a univariate density $g_\epsilon(e)$. This density is expressed as a mixture of Bayesian G-splines (normal densities with equidistant means and constant variances). [13] specify the error density $g_\epsilon(e)$ as a shifted and scaled penalized Gaussian mixture (PGM), which is expressed as

$$g_\epsilon(e) = \gamma^{-1} \sum_{j=-K}^{K} \omega_j(a) \varphi\{\tau^{-1}(e - \alpha) | \mu_j, \theta\}, \tag{8}$$

where α and τ are the intercept and scale parameter, respectively, μ_j is a fine grid of equidistant knots centered around zero ($\mu_0 = 0$) and θ^2 is a fixed

Figure 21: Comparing observed and predicted survival curves.



basis variance, common for all mixture components. The mixture weights are given by

$$\omega_j(a) = \frac{\exp(a_j)}{\sum_{k=-K}^{K} \exp(a_k)},\ j = -K, \ldots, K.$$

Here $0 < \omega_j < 1,\ j = -K, \ldots, K$ and $\sum_{j=-K}^{K} \omega_j = 1$ to ensure that g_ϵ is a density. For details, see [13].

7.1 Results

For the analysis of the bending and tension data, the AFT and Bayesian AFT model were fitted. In both models, we included the following covariates: ‘knot’, ‘offg’ and ‘moe’. A Bayesian approach with the MCMC methodology is used to estimate the model parameters. The computation was performed using the bayesSurv library in R. The estimates for the regression coefficients to be quite close to each other for the bending data. However for tension data the estimates do differ especially for the covariate ‘knot’.

Comparison of the two approaches. For the bending data set we applied the reduced Weibull AFT model given by

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

where μ and α_i are regression parameters, σ is a scale parameter, and ϵ is the random error. The R output for the reduced model appears below:

```
survreg(formula = Surv(mor) ~ factor(knot) + offg + moe, data = eortc, dist
    = "weibull")
                Value Std. Error      z         p
(Intercept)    1.0174     0.1796   5.67  1.47e-08
factor(knot)1 -0.1428     0.0412  -3.46  5.31e-04
factor(knot)2 -0.0978     0.0471  -2.08  3.79e-02
offg          -0.2118     0.0661  -3.20  1.35e-03
moe            0.7057     0.1119   6.31  2.86e-10
Log(scale)    -1.9688     0.0786 -25.04 2.00e-138
Scale= 0.140
```

Next we give the R output for Bayesian Weibull AFT model:

```
                 Mean      SD  Naive SE Time-series SE
factor.knot.1 -0.16031 0.04631 0.0002929       0.002302
factor.knot.2 -0.08622 0.05023 0.0003177       0.001883
offg          -0.21302 0.08625 0.0005455       0.005014
moe            0.67881 0.10831 0.0006850       0.013821
```

Note that the Bayesian approach does not generate p-values like the non-Bayesian one as the two paradigms for inference are quite different. However the coefficient estimates from these two approaches are similar.

Turning to the tension data we get the following listing for the R output for weibull AFT model:

```
survreg(formula = Surv(mot) ~ factor(knot) + offg + moe, data = eortc,dist =
     "weibull")
                Value Std. Error        z        p
(Intercept)   -0.0352     0.3170   -0.111 9.11e-01
factor(knot)1 -0.0878     0.0713   -1.232 2.18e-01
factor(knot)2  0.0148     0.0953    0.155 8.77e-01
offg          -0.4927     0.1449   -3.402 6.70e-04
moe            1.0643     0.1945    5.471 4.48e-08
Log(scale)    -1.4140     0.0763  -18.528 1.24e-76

Scale= 0.243
```

The R output for Bayesian Weibull AFT model is given by:

```
                Mean     SD  Naive SE Time-series SE
factor.knot.1 -0.07334 0.0833 0.0005268       0.005585
factor.knot.2  0.08435 0.1118 0.0007073       0.007004
offg          -0.39563 0.1611 0.0010188       0.004488
moe            1.22327 0.2154 0.0013622       0.030248
```

The discrepancy between the two approaches seems somewhat larger for tension than for bending.

Predictive distribution for the Bayesian AFT model. The Bayesian approach also enables us to develop a predictive distribution for lumber strengths given the covariates. The one used here is provided in the R function predictive2 of the bayesSurv library is based on the Bayesian AFT model. It differs from the to the hierarchical Bayes method developed by [20]. The two approaches are to be compared in future work.

The function *predictive*2 computes predictive densities, survivor and hazard curves for specified combinations of covariates. To compute the predictive survival functions, we need to specify the combinations of covariates: ‘knot’, ‘offg’, and ‘moe’, for which the survival functions are computed as follows:

```
eortc.pred <- data.frame(mor=c(1, 1), knot=eortc$knot,offg=eortc$offg,moe=
    eortc$moe)
```

Computation of the values of predictive survival function on the equidistant grid of 5 time values from 1 to 10 is then performed using the following code:

```
pred <- predictive2(Surv(mor)˜factor(knot)+offg+moe, grid=seq(1, 10, length
    =5), Gspline=list(dim=1, K=15),quantile=c(0.025, 0.975), only.aver=FALSE
    ,  predict=list(Surv=TRUE, density=FALSE, hazard=TRUE, cum.hazard=FALSE)
    ,data=eortc.pred)
```

We give a complete listing of the output for all the items in the test samples in Appendix C. Here we provide just the first ten predictions for bending test and tension samples. The first ten lines of R output for bending data: predictive survival probabilities The R output for bending data: predictive survival probabilities given the relevant combinations of the covariates 'knot', 'offg', and 'moe' for each piece:.

```
> pred$grid
[1]  1.00  3.25  5.50  7.75 10.00

           [,1]      [,2]      [,3]        [,4]         [,5]
 [1,] 0.9999985 0.9986256 0.9100139 0.291686169 0.0084855182
 [2,] 0.9999940 0.9931019 0.6262136 0.047264637 0.0013653316
 [3,] 0.9999956 0.9958788 0.7360959 0.052465177 0.0009051875
 [4,] 0.9999979 0.9981359 0.8711861 0.189736563 0.0034501664
 [5,] 0.9999765 0.9541225 0.1857986 0.002057685 0.0002222578
 [6,] 0.9999990 0.9988219 0.9267654 0.366402726 0.0184144255
 [7,] 0.9999994 0.9993006 0.9654483 0.583086212 0.0748428375
 [8,] 0.9999935 0.9928616 0.6137014 0.019852839 0.0005825212
 [9,] 0.9999986 0.9984618 0.8987351 0.257551609 0.0075251075
[10,] 0.9999988 0.9988708 0.9308221 0.373069890 0.0165692096
[
```

Thus for piece # 1, for example, the probability of the piece surviving to a load of 7.75 is about 0.29. In contrast for piece #8 that probability is just 0.02. **What was the actual failure load in these cases?**

The first ten lines of R output for tension data: predictive survival probabilities given the relevant combinations of the covariates 'knot', 'offg', and 'moe' for each piece:

```
$grid
[1]  1.00  3.25  5.50  7.75 10.00

           [,1]       [,2]        [,3]         [,4]         [,5]
 [1,] 0.9998054 0.47894740 0.031226112 0.0014119742 9.035891e-05
 [2,] 0.9999960 0.88981999 0.277166112 0.0342113185 3.458908e-03
 [3,] 0.9999992 0.99252661 0.786548845 0.3670187402 1.150359e-01
 [4,] 0.9996624 0.35901186 0.012683602 0.0003876112 2.030179e-05
 [5,] 0.9998526 0.51468008 0.037815390 0.0018173140 1.192733e-04
 [6,] 0.9999741 0.69778485 0.086562849 0.0050160783 3.371247e-04
 [7,] 0.9998313 0.44885159 0.021959031 0.0007690396 4.226916e-05
 [8,] 0.9999986 0.96408794 0.528730459 0.1312943332 2.233619e-02
 [9,] 0.9999670 0.70129198 0.099240712 0.0070185758 5.497349e-04
[10,] 0.9999907 0.83292569 0.204574367 0.0219364470 2.153119e-03
```

8 Summary and concluding remarks

This study is based on the wood strength data collected in a FPInnovations (FPI) laboratory. However, the sample sizes for both bending and testing were small making all our findings provisional, bending confirmation with larger datasets.

We employed survival analysis methods in a very different context from that in which they were first developed, namely to model the load–to–failure in manufactured lumber. 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. The association between MSRC and FC was examined by using 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 situations. Moreover, the AFT model has a more realistic interpretation in terms of an effect on the expected load–to–failure. It also provides more informative results. In this context the AFT model has explanatory advantage over the Cox PH alternative, in that the covariates have a direct effect on load to failure rather on hazard functions as in the Cox PH model. This leads us to recommend subject to confirmation with a larger sample, the AFT alternative over the Cox PH model. Subject to the same caveat, we see the final, Weibull AFT model as being suitable for changing the grading rules in the future should that prove necessary. As well, this model could be used to develop machine grading rules as new technologies come on stream.

The finding reported in the previous paragraph led us to further develop the AFT approach within a Bayesian context. The results for the data from the tension tests unlike that for bending, differed to quite a degree, in that even small discrepancies on the log scale can turn into large one on the original scale. Finally we demonstrated the value of the Bayesian approach by developing a predictive distribution for the load–to–failure.

The Cox PH model and the Weibull AFT model yield the same significant covariates - ‘knot’, ‘off-grade’ and ‘moe’, indicating these three are the most important predictors for modeling reliability. 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. Finally the survival load increases log–linearly as a function of ‘moe’ .

These significant covariates can be used in conjunction with the final AFT model to match pieces of lumber in describing the relationships among strength properties. Obtaining matched pairs in this way can be used to address the challenging problem of estimating the degree of association between bending and tension survival loads. This is another potential benefit to be realized from the work reported here. Moreover it may have application in duration of load studies currently underway.

In practice the model relating the strength of a piece of lumber to its covariates cannot be known and we explore through simulation studies the inferential effect of mis-specifying 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 in another way.

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.

Acknowledgements. We thank Roy Abbott for professionally grading the lumber test samples used in this study, and the staff at FPInnovations for assisting with the testing procedure. We thank Samuel Wong for his major role in conducting the experiments at FPInnovations as well as for early discussions relating to the topic of this paper. Thanks as well to Yilan Zhu for her assistance in those experiments.

A Appendix A: R code for analysis of bending data

```
require(bayesSurv)
library(coda)
library(lattice)
library(smoothSurv)
# specifiy the prior
prior.error <- list(K=15, c4delta=1.5, order=3, prior.intercept="normal", mean.int
var.intercept=100,prior.scale="gamma", shape.scale=1, rate.scale=0.005,prior.lamb
shape.lambda=1, rate.lambda=0.005)
prior.betaGamma <- list(mean.prior=rep(0, 4), var.prior=rep(100, 4))
prior.b <- list(prior.D = "inv.wishart", df.D = 2, scale.D = 0.002*c(1,0,1))
# intital value for the regression parameters
library(survival)
table<-read.table("bending.txt", header = TRUE)
eortc <-data.frame(table)
# Reduced AFT model
fit0 <-survreg(Surv(mor)~factor(knot)+offg+moe,dist="weibull",data=eortc)
summary(fit0)
beta.init <- fit0$coeff[-(1:2)]
gamma.init <- fit0$coeff["moe"]
init <- list(beta = c(gamma.init, beta.init), D = c(1, 0, 1), lambda = 100,
intercept = fit0$coeff["(Intercept)"],
scale = fit0$scale, gamma = 0, sigma = 0.2)
# MCMC sampling
library(bayesSurv)
sample <- bayessurvreg2(Surv(mor)~factor(knot)+offg+moe, prior=prior.error, init=i
prior.beta=prior.betaGamma, prior.b=prior.b, nsimul=list(niter=125000, nthin=5, n
store=list(b=TRUE), data=eortc)
library(coda)
betaGamma <- read.table("beta.sim", header=TRUE)
betaGamma <- mcmc(betaGamma)
summary(betaGamma)
HPDinterval(betaGamma)
exp.betaGamma <- mcmc(exp(betaGamma))
summary(exp.betaGamma)
```

```
HPDinterval(exp.betaGamma)

# Compute predictive quantities based on a Bayesian survival regression model fitt
# This function computes predictive densities, survivor and hazard curves
# for specified combinations of covariates.
eortc.pred <- data.frame(mor=c(1, 1), knot=eortc$knot,offg=eortc$offg,moe=eortc$mo
pred <- predictive2(Surv(mor)~factor(knot)+offg+moe, grid=seq(1, 10, length=5),
Gspline=list(dim=1, K=15),quantile=c(0.025, 0.975), only.aver=FALSE,
predict=list(Surv=TRUE, density=FALSE,  hazard=TRUE, cum.hazard=FALSE),data=eortc.
pred$grid
pred$Surv
pred$hazard
plot(pred$Surv[1,],type="l",ylab="Survivor")
```

B Appendix B: R code for analysis of tension data

```
require(bayesSurv)
library(coda)
library(lattice)
library(smoothSurv)
# specifiy the prior
prior.error <- list(K=15, c4delta=1.5, order=3, prior.intercept="normal", mean.int
var.intercept=100,prior.scale="gamma", shape.scale=1, rate.scale=0.005,
prior.lambda="gamma", shape.lambda=1, rate.lambda=0.005)
prior.betaGamma <- list(mean.prior=rep(0, 4), var.prior=rep(100, 4))
prior.b <- list(prior.D = "inv.wishart", df.D = 2, scale.D = 0.002*c(1,0,1))
# intital value for the regression parameters
library(survival)
table<-read.table("tension.txt", header = TRUE)
eortc <-data.frame(table)
###Reduced AFT model
fit0 <-survreg(Surv(mot)˜factor(knot)+offg+moe,dist="weibull",data=eortc)
summary(fit0)
beta.init <- fit0$coeff[-(1:2)]
gamma.init <- fit0$coeff["moe"]
init <- list(beta = c(gamma.init, beta.init), D = c(1, 0, 1), lambda = 100,
intercept = fit0$coeff["(Intercept)"],
scale = fit0$scale, gamma = 0, sigma = 0.2)
# MCMC sampling
library(bayesSurv)
sample <- bayessurvreg2(Surv(mot)˜factor(knot)+offg+moe, prior=prior.error, init=i
prior.beta=prior.betaGamma, prior.b=prior.b, nsimul=list(niter=125000, nthin=5, nb
store=list(b=TRUE), data=eortc)
library(coda)
betaGamma <- read.table("beta.sim", header=TRUE)
betaGamma <- mcmc(betaGamma)
summary(betaGamma)
HPDinterval(betaGamma)
exp.betaGamma <- mcmc(exp(betaGamma))
summary(exp.betaGamma)
```

```
HPDinterval(exp.betaGamma)
eortc.pred <- data.frame(mot=c(1, 1), knot=eortc$knot,offg=eortc$offg,moe=eortc$mo
pred <- predictive2(Surv(mot)~factor(knot)+offg+moe, grid=seq(1, 10, length=5),
Gspline=list(dim=1, K=15),quantile=c(0.025, 0.975), only.aver=FALSE,
predict=list(Surv=TRUE, density=FALSE, hazard=TRUE, cum.hazard=FALSE),data=eortc.p
pred$grid
pred$Surv
pred$hazard
```

C Appendix C: bayesSurv predictions

- The R output for bending data: predictive survival probabilities given the relevant combinations of the covariates ‘knot’, ‘offg’, and ‘moe’ for each piece:

```
> pred$grid
[1]  1.00  3.25  5.50  7.75 10.00

           [,1]      [,2]      [,3]        [,4]         [,5]
 [1,] 0.9999985 0.9986256 0.9100139 0.291686169 0.0084855182
 [2,] 0.9999940 0.9931019 0.6262136 0.047264637 0.0013653316
 [3,] 0.9999956 0.9958788 0.7360959 0.052465177 0.0009051875
 [4,] 0.9999979 0.9981359 0.8711861 0.189736563 0.0034501664
 [5,] 0.9999765 0.9541225 0.1857986 0.002057685 0.0002222578
 [6,] 0.9999990 0.9988219 0.9267654 0.366402726 0.0184144255
 [7,] 0.9999994 0.9993006 0.9654483 0.583086212 0.0748428375
 [8,] 0.9999935 0.9928616 0.6137014 0.019852839 0.0005825212
 [9,] 0.9999986 0.9984618 0.8987351 0.257551609 0.0075251075
[10,] 0.9999988 0.9988708 0.9308221 0.373069890 0.0165692096
[11,] 0.9999987 0.9987805 0.9231058 0.340026654 0.0127028066
[12,] 0.9999966 0.9969031 0.7902563 0.084992486 0.0012787989
[13,] 0.9999990 0.9987765 0.9228106 0.349938051 0.0165193321
[14,] 0.9999964 0.9967262 0.7802083 0.077487592 0.0011847014
[15,] 0.9999960 0.9953560 0.7217239 0.050658317 0.0010662659
[16,] 0.9999995 0.9993386 0.9685954 0.615640313 0.0855803066
[17,] 0.9999984 0.9985678 0.9052124 0.276051746 0.0074214827
[18,] 0.9999987 0.9988268 0.9270599 0.356492251 0.0145164328
[19,] 0.9999929 0.9919269 0.5825357 0.015827700 0.0005393558
[20,] 0.9999984 0.9982422 0.8820113 0.214169702 0.0053299396
[21,] 0.9999969 0.9965556 0.7781415 0.079456422 0.0015190358
[22,] 0.9999952 0.9953564 0.7116372 0.042693883 0.0008068100
[23,] 0.9999995 0.9992514 0.9635842 0.580865439 0.0697211583
[24,] 0.9999994 0.9991135 0.9522343 0.500463925 0.0428560464
[25,] 0.9999956 0.9958788 0.7360959 0.052465177 0.0009051875
[26,] 0.9999980 0.9977671 0.8486286 0.151556204 0.0031158728
[27,] 0.9999998 0.9995616 0.9837820 0.774312822 0.2235565045
[28,] 0.9999975 0.9972487 0.8162455 0.111187539 0.0021249073
[29,] 0.9999967 0.9970679 0.7999239 0.093026114 0.0013858671
[30,] 0.9999986 0.9987317 0.9189522 0.323711518 0.0111068033
[31,] 0.9999973 0.9976235 0.8349459 0.130605350 0.0019932773
[32,] 0.9999935 0.9928616 0.6137014 0.019852839 0.0005825212
[33,] 0.9999978 0.9980465 0.8645571 0.176878438 0.0030674481
[34,] 0.9999945 0.9944246 0.6719481 0.030972883 0.0006927562
[35,] 0.9999977 0.9975265 0.8331147 0.130268517 0.0025584156
[36,] 0.9999993 0.9992190 0.9594581 0.537161880 0.0551032047
[37,] 0.9999976 0.9971649 0.8090804 0.115730107 0.0026489318
[38,] 0.9999988 0.9986451 0.9132915 0.304414723 0.0107248856
[39,] 0.9999960 0.9963326 0.7589458 0.064010383 0.0010283479
[40,] 0.9999963 0.9958868 0.7454982 0.060861344 0.0012180852
[41,] 0.9999958 0.9950623 0.7092285 0.046149151 0.0010016186
[42,] 0.9999983 0.9985065 0.9001778 0.260718610 0.0064977548
[43,] 0.9999997 0.9995634 0.9823701 0.754622956 0.2023974662
[44,] 0.9999992 0.9989444 0.9375292 0.416574907 0.0254424890
[45,] 0.9999972 0.9974986 0.8267209 0.120385950 0.0018085205
[46,] 0.9999989 0.9987508 0.9219078 0.337117378 0.0136069069
```

```
[47,] 0.9999942 0.9940702 0.6579371 0.027755646 0.0006614979
[48,] 0.9999992 0.9989813 0.9407753 0.433427557 0.0282986428
[49,] 0.9999765 0.9541225 0.1857986 0.002057685 0.0002222578
[50,] 0.9999998 0.9995094 0.9809746 0.742309266 0.1839880784
[51,] 0.9999981 0.9982982 0.8835843 0.216869891 0.0044090836
[52,] 0.9999985 0.9986802 0.9145910 0.307585291 0.0097076635
[53,] 0.9999975 0.9978491 0.8503901 0.152681073 0.0024525144
[54,] 0.9999958 0.9961138 0.7477211 0.058005599 0.0009632071
[55,] 0.9999982 0.9979764 0.8628674 0.175032417 0.0038355381
[56,] 0.9999977 0.9979511 0.8576284 0.164520908 0.0027373896
[57,] 0.9999986 0.9983931 0.8934134 0.242661510 0.0066985895
[58,] 0.9999983 0.9981593 0.8759118 0.200621903 0.0047664595
[59,] 0.9999997 0.9993685 0.9722549 0.654392850 0.1083577920
[60,] 0.9999998 0.9995274 0.9819632 0.753351889 0.1967459155
[61,] 0.9999947 0.9947565 0.6855730 0.034519402 0.0007270842
[62,] 0.9999975 0.9972487 0.8162455 0.111187539 0.0021249073
[63,] 0.9999935 0.9928616 0.6137014 0.019852839 0.0005825212
[64,] 0.9999964 0.9967262 0.7802083 0.077487592 0.0011847014
[65,] 0.9999899 0.9853489 0.4147551 0.005256465 0.0003846893
[66,] 0.9999953 0.9944137 0.6830334 0.038202177 0.0008905710
[67,] 0.9999992 0.9990084 0.9432247 0.438664995 0.0274348844
[68,] 0.9999926 0.9914183 0.5665135 0.014128269 0.0005199462
[69,] 0.9999970 0.9973649 0.8181480 0.110717244 0.0016479562
[70,] 0.9999945 0.9944246 0.6719481 0.030972883 0.0006927562
[71,] 0.9999967 0.9970679 0.7999239 0.093026114 0.0013858671
[72,] 0.9999994 0.9991718 0.9571365 0.533262700 0.0523207496
[73,] 0.9999985 0.9983200 0.8878419 0.228191364 0.0059706047
[74,] 0.9999989 0.9987995 0.9259140 0.353787910 0.0153228628
[75,] 0.9999962 0.9965364 0.7697733 0.070498620 0.0011017376
[76,] 0.9999982 0.9984413 0.8949011 0.245721960 0.0056977330
[77,] 0.9999978 0.9980465 0.8645571 0.176878438 0.0030674481
[78,] 0.9999973 0.9976235 0.8349459 0.130605350 0.0019932773
[79,] 0.9999966 0.9969031 0.7902563 0.084992486 0.0012787989
[80,] 0.9999985 0.9983200 0.8878419 0.228191364 0.0059706047
[81,] 0.9999975 0.9978491 0.8503901 0.152681073 0.0024525144
[82,] 0.9999929 0.9919269 0.5825357 0.015827700 0.0005393558
[83,] 0.9999993 0.9992471 0.9615664 0.552727076 0.0612050492
[84,] 0.9999996 0.9993003 0.9673397 0.611244019 0.0836654736
[85,] 0.9999924 0.9908810 0.5502335 0.012612965 0.0005017490
[86,] 0.9999960 0.9963326 0.7589458 0.064010383 0.0010283479
[87,] 0.9999993 0.9991124 0.9516823 0.489345513 0.0383103672
[88,] 0.9999900 0.9895592 0.5351473 0.024459781 0.0006533109
[89,] 0.9999974 0.9977401 0.8428325 0.141372543 0.0022063407
[90,] 0.9999937 0.9932895 0.6287999 0.022217631 0.0006066941
[91,] 0.9999977 0.9979511 0.8576284 0.164520908 0.0027373896
[92,] 0.9999926 0.9889995 0.5145798 0.013107458 0.0005376674
[93,] 0.9999987 0.9985876 0.9086654 0.288463381 0.0095242196
[94,] 0.9999995 0.9993628 0.9702309 0.630091796 0.0936612401
[95,] 0.9999954 0.9956268 0.7240681 0.047368529 0.0008533286
[96,] 0.9999976 0.9973926 0.8248544 0.120453201 0.0023281923
[97,] 0.9999960 0.9963326 0.7589458 0.064010383 0.0010283479
[98,] 0.9999970 0.9970999 0.7922647 0.146679365 0.0051763286
```

- The R output for tension data: predictive survival probabilities given the relevant combinations of the covariates ‘knot’, ‘offg’, and ‘moe’ for each piece:

```
$grid
[1]  1.00  3.25  5.50  7.75 10.00

            [,1]       [,2]        [,3]         [,4]         [,5]
 [1,] 0.9998054 0.47894740 0.031226112 0.0014119742 9.035891e-05
 [2,] 0.9999960 0.88981999 0.277166112 0.0342113185 3.458908e-03
 [3,] 0.9999992 0.99252661 0.786548845 0.3670187402 1.150359e-01
 [4,] 0.9996624 0.35901186 0.012683602 0.0003876112 2.030179e-05
 [5,] 0.9998526 0.51468008 0.037815390 0.0018173140 1.192733e-04
 [6,] 0.9999741 0.69778485 0.086562849 0.0050160783 3.371247e-04
 [7,] 0.9998313 0.44885159 0.021959031 0.0007690396 4.226916e-05
 [8,] 0.9999986 0.96408794 0.528730459 0.1312943332 2.233619e-02
 [9,] 0.9999670 0.70129198 0.099240712 0.0070185758 5.497349e-04
[10,] 0.9999907 0.83292569 0.204574367 0.0219364470 2.153119e-03
[11,] 0.9998530 0.46731392 0.024430717 0.0008815748 4.898079e-05
[12,] 0.9999939 0.85385709 0.216696567 0.0219202425 1.961960e-03
[13,] 0.9999950 0.87280652 0.245992785 0.0274789845 2.609960e-03
[14,] 0.9999945 0.86357884 0.231097776 0.0245632078 2.263867e-03
[15,] 0.9999877 0.78482826 0.141577992 0.0107245182 8.193739e-04
[16,] 0.9999914 0.82165495 0.176644852 0.0154319699 1.271301e-03
[17,] 0.9999823 0.74344473 0.111599843 0.0073695973 5.261592e-04
[18,] 0.9999026 0.52291212 0.033406365 0.0013259340 7.630768e-05
[19,] 0.9998654 0.52572013 0.039657108 0.0018530985 1.170063e-04
[20,] 0.9965939 0.23053058 0.009752753 0.0004425914 2.916773e-05
[21,] 0.9999989 0.97922374 0.632656644 0.2012564002 4.246216e-02
[22,] 0.9999955 0.88154986 0.261357843 0.0306880639 3.006127e-03
[23,] 0.9999772 0.71345331 0.094377534 0.0057078479 3.911148e-04
[24,] 0.9999903 0.80989399 0.164391349 0.0136876622 1.098703e-03
[25,] 0.9999823 0.74344473 0.111599843 0.0073695973 5.261592e-04
[26,] 0.9999706 0.68169597 0.079259640 0.0044043018 2.905565e-04
[27,] 0.9999985 0.96077868 0.510771980 0.1214693828 1.994818e-02
[28,] 0.9996346 0.52515735 0.061121328 0.0051553143 4.966681e-04
[29,] 0.9999258 0.55971371 0.040893222 0.0017376665 1.026399e-04
[30,] 0.9999877 0.78482826 0.141577992 0.0107245182 8.193739e-04
[31,] 0.9999619 0.64835845 0.066118867 0.0033874231 2.157944e-04
[32,] 0.9998883 0.50438839 0.030134592 0.0011576106 6.581177e-05
[33,] 0.9999433 0.59590526 0.049781003 0.0022732569 1.381338e-04
[34,] 0.9999665 0.66521133 0.072451075 0.0038639981 2.504043e-04
[35,] 0.9999565 0.63116756 0.060243524 0.0029675607 1.859689e-04
[36,] 0.9998063 0.43049773 0.019715712 0.0006707382 3.648536e-05
[37,] 0.9999711 0.71647762 0.107440621 0.0079022823 6.312245e-04
[38,] 0.9997062 0.37651266 0.014182407 0.0004446272 2.349741e-05
[39,] 0.9996853 0.41965024 0.022165693 0.0008671941 5.108058e-05
[40,] 0.9999258 0.55971371 0.040893222 0.0017376665 1.026399e-04
[41,] 0.9999931 0.84363312 0.202809713 0.0195301266 1.698958e-03
[42,] 0.9999741 0.69778485 0.086562849 0.0050160783 3.371247e-04
[43,] 0.9999848 0.78589203 0.156139954 0.0140922738 1.253901e-03
[44,] 0.9997775 0.41229548 0.017682953 0.0005849070 3.150028e-05
[45,] 0.9999991 0.98356851 0.683767714 0.2520205617 6.301111e-02
[46,] 0.9999433 0.59590526 0.049781003 0.0022732569 1.381338e-04
[47,] 0.9997443 0.39428688 0.015843845 0.0005099922 2.720283e-05
[48,] 0.9999963 0.89762931 0.293387816 0.0380698340 3.975545e-03
[49,] 0.9999974 0.92932560 0.384977181 0.0669352848 8.670512e-03
[50,] 0.9999844 0.75773286 0.121031548 0.0083608889 6.100636e-04
[51,] 0.9999891 0.79761763 0.152701139 0.0121238075 9.490359e-04
```

```
[52,] 0.9999772 0.71345331 0.094377534 0.0057078479 3.911148e-04
[53,] 0.9999799 0.72867956 0.102718815 0.0064889524 4.536836e-04
[54,] 0.9999799 0.72867956 0.102718815 0.0064889524 4.536836e-04
[55,] 0.9999619 0.64835845 0.066118867 0.0033874231 2.157944e-04
[56,] 0.9988301 0.21907043 0.004486177 0.0001127682 5.510785e-06
[57,] 0.9999862 0.77153081 0.131022366 0.0094748652 7.071388e-04
[58,] 0.9999974 0.92932560 0.384977181 0.0669352848 8.670512e-03
[59,] 0.9999503 0.61367144 0.054804611 0.0025980741 1.602714e-04
[60,] 0.9999957 0.89591535 0.300252785 0.0415080869 4.605245e-03
[61,] 0.9999741 0.69778485 0.086562849 0.0050160783 3.371247e-04
[62,] 0.9999978 0.93995269 0.420526246 0.0800782200 1.107736e-02
[63,] 0.9999980 0.94061762 0.415516736 0.0761939514 1.015331e-02
[64,] 0.9999706 0.68169597 0.079259640 0.0044043018 2.905565e-04
[65,] 0.9999665 0.66521133 0.072451075 0.0038639981 2.504043e-04
[66,] 0.9999565 0.63116756 0.060243524 0.0029675607 1.859689e-04
[67,] 0.9999967 0.90499127 0.309990452 0.0422847269 4.564020e-03
[68,] 0.9999880 0.80706679 0.173160654 0.0159886534 1.396356e-03
[69,] 0.9999565 0.63116756 0.060243524 0.0029675607 1.859689e-04
[70,] 0.9999741 0.69778485 0.086562849 0.0050160783 3.371247e-04
[71,] 0.9999665 0.66521133 0.072451075 0.0038639981 2.504043e-04
[72,] 0.9998883 0.50438839 0.030134592 0.0011576106 6.581177e-05
[73,] 0.9997775 0.41229548 0.017682953 0.0005849070 3.150028e-05
[74,] 0.9999989 0.97243893 0.577954055 0.1602413190 3.000493e-02
[75,] 0.9999950 0.87280652 0.245992785 0.0274789845 2.609960e-03
[76,] 0.9999916 0.84105330 0.211128872 0.0222396330 2.091078e-03
[77,] 0.9999804 0.75948974 0.135089844 0.0112158064 9.539373e-04
[78,] 0.9999827 0.77292609 0.145350003 0.0125789158 1.093920e-03
[79,] 0.9998063 0.43049773 0.019715712 0.0006707382 3.648536e-05
[80,] 0.9999260 0.60286509 0.059798590 0.0033893705 2.390340e-04
[81,] 0.9999981 0.94868665 0.463793009 0.1015015038 1.618479e-02
[82,] 0.9999894 0.81888826 0.185291522 0.0178693742 1.598415e-03
[83,] 0.9999979 0.94474179 0.438500371 0.0873552707 1.249622e-02
[84,] 0.9999914 0.82165495 0.176644852 0.0154319699 1.271301e-03
[85,] 0.9999975 0.92454040 0.361723917 0.0572751809 6.852269e-03
[86,] 0.9999351 0.57790629 0.045151132 0.0019879844 1.190647e-04
[87,] 0.9996624 0.35901186 0.012683602 0.0003876112 2.030179e-05
[88,] 0.9999962 0.90407587 0.324383072 0.0497298621 6.121239e-03
[89,] 0.9999823 0.74344473 0.111599843 0.0073695973 5.261592e-04
[90,] 0.9999891 0.79761763 0.152701139 0.0121238075 9.490359e-04
[91,] 0.9999772 0.71345331 0.094377534 0.0057078479 3.911148e-04
[92,] 0.9997443 0.39428688 0.015843845 0.0005099922 2.720283e-05
[93,] 0.9999026 0.52291212 0.033406365 0.0013259340 7.630768e-05
[94,] 0.9999914 0.82165495 0.176644852 0.0154319699 1.271301e-03
[95,] 0.9999891 0.79761763 0.152701139 0.0121238075 9.490359e-04
[96,] 0.9999916 0.84105330 0.211128872 0.0222396330 2.091078e-03
[97,] 0.9999613 0.67987018 0.087737982 0.0056118463 4.068430e-04
[98,] 0.9819035 0.07413874 0.001294146 0.0000372117 1.918832e-06
```

References

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

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

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

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

[5] F.S.C.F.S. Chapin III, A.D.M.G.A.D. McGuire, R.W.R.R.W. Ruess, T.N.H.T.N. Hollingsworth, M.C.M.M.C. Mack, J.F.J.J.F. Johnstone, E.S.K.E.S. Kasischke, E.S.E.E.S. Euskirchen, J.B.J.J.B. Jones, M.T.J.M.T. Jorgenson, et al. Resilience of alaska's boreal forest to climatic change this article is one of a selection of papers from the dynamics of change in alaska's boreal forests: Resilience and vulnerability in response to climate warming. *Canadian Journal of Forest Research*, 40(7):1360–1370, 2010.

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

[7] N.C. Coops and R.H. Waring. Estimating the vulnerability of fifteen tree species under changing climate in northwest north america. *Ecological Modelling*, 2011.

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

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

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

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

[12] R.A. Johson, W.E. James, and D.W. Green. Some Bivariate Distributions for Modeling the Strength Properties of Lumber. *Forest Products Laboratory Research Paper*, 1999.

[13] A. Komarek, E. Lesaffre, and C. Legrand. Baseline and treatment effect heterogeneity in disease free survival between centers in a random effects accelerated failure time model with flexible error distribution. *Statistics in Medicine*, 26:726–745, 2007.

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

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

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

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

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

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

[20] S. Wong, C. Lum, L. Wu, and J.V. Zidek. A bayesian framework for relating the strength of lumber to its covariates. 2012.