To the Graduate Council:

I am submitting herewith a thesis written by David Joseph Edwards entitled "An Applied Statistical Reliability Analysis of the Internal Bond of Medium Density Fiberboard." I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Statistics.

Frank Guess, Co-chair

Timothy Young, Co-chair

We have read this thesis
and recommend its acceptance:

Halima Bensmail

Accepted for the Council:

Vice Chancellor and
Dean of Graduate Studies

An Applied Statistical Reliability Analysis of the Internal Bond of Medium Density Fiberboard

Masters Thesis
Presented for the
Master of Science Degree
The University of Tennessee, Knoxville

David Joseph Edwards
May 2004

Dedication

This masters thesis is dedicated to my wife:

Elizabeth,

who has provided constant love and support

throughout this thesis' research and writing process.

Without her compassion, understanding, and spark for life,

progress toward the goal of finishing this thesis

would not have been easy.

She is truly a gift from God.

To my family who has always been there for me and

who have always believed I could accomplish anything.

To my friends who make me laugh and remind me that

life is an adventure to be enjoyed.

To our pet Snoopy for being a part of my support structure.

Her listening ear and love for attention has made life less stressful

and more enjoyable.

Acknowledgements

The work on this thesis was funded by Associate Research Professor Timothy Young from the United States Department of Agriculture, Special Wood Utilization Research Grants, which is administered by The University of Tennessee Agricultural Experiment Station and the Tennessee Forest Products Center.

I would like to thank my thesis committee co-chairs, Dr. Frank Guess and Timothy Young, for their guidance, persistence, and attention to details. They have helped to provide considerable growth for future graduate work and my professional career as a statistician. In addition, I would like to thank the other member of my committee, Dr. Halima Bensmail, who has provided needed support and suggestions to enhance my understandings.

Also, I appreciate Dr. Hamparsum Bozdogan and Dr. William Seaver for reading over portions of this thesis and providing their helpful comments and changes. Dr. Bozdogan read and commented on Chapter 4 on information criteria, while Dr. Seaver did Chapter 5 on bootstrapping. I appreciate some additional comments from Dr. Russell Zaretzki on his reading of Chapter 5 on bootstrapping. Thanks also to the other members of the faculty and staff of the Department of Statistics and the Tennessee Forest Products Center. Working together and networking has certainly aided in the completion of this thesis.

Abstract

The forest products industry has seen tremendous growth in recent years and has a huge impact on the economies of many countries. For example, in the state of Maine in 1997, the forest products industry accounted for 9 billion U.S. dollars for that year. In the state of Tennessee, for example in 2000, this figure was 22 billion U.S. dollars for that year. It has, therefore, become more important in this industry to focus on producing higher quality products. Statistical reliability methods, among other techniques, have been employed to help monitor and improve the quality of forest products. With such a large focus on quality improvement, data is quite plentiful, allowing for more useful analyses and examples.

In this thesis, we demonstrate the usefulness of statistical reliability tools and apply them to help assess, manage, and improve the internal bond (IB) of medium density fiberboard (MDF). MDF is a high quality engineered wood composite that undergoes destructive testing during production. Workers test cross sections of MDF panels and measure the IB in pounds per square inches. IB is a key metric of quality since it provides a direct measurement for the strength of MDF, which is important to customers and the manufacturers.

Graphical procedures such as histograms, scatter plots, probability plots, and survival curves are explored to help the practitioner gain insights regarding the distributions of IB and strengths of different MDF product types. Much information can be revealed from a graphics approach.

Though useful, probability plots can be a subjective way to assess the parametric

distribution of a data set. Insightful developments in information criteria, in particular Akaike's Information Criteria and Bozdogan's Information Complexity Criteria, have made probability plotting more objective by assigning numeric scores to each plot. The plot with the lowest score is deemed the best among competing models. In application to MDF, we will see that initial intuitions are not always confirmed. Therefore, information criteria prove to be useful tools for the practitioner seeking more clarity regarding distributional assumptions. We recommend more usage of these helpful information criteria.

Estimating lower percentiles in failure data analysis can provide valuable assistance to the practitioner for understanding product warranties and their costs. Since data may not be plentiful for the lower tails, estimation of these percentiles may not be an easy task. Indeed, we stress times to not even try to estimate the lowest percentiles. If samples are large and parametric assumptions are weak or not available, asymptotic approximations can be utilized. However, unless the sample size is sufficiently large, such approximations will not be accurate.

Bootstrap techniques provide one solution for the estimation of lower percentiles when asymptotic approximations should not be utilized. This computer intensive resampling scheme provides a method for estimating the true sampling distribution of these percentiles, or any population parameter of interest. This can be used for various parametric models or for nonparametric settings, when the parametric model might be imperfect or misspecified. The empirical bootstrap distribution can then be used for inferences such as determining standard errors and constructing confidence intervals. Helpful applications of the bootstrap to the MDF data show this procedure's advantages

and limitations in order to aid the practitioner in their decision-making. Graphics can readily warn the practitioner when even certain bootstrap procedures are not advisable.

To be able to say that improvements have been made, we must be able to measure reliability expressed in percentiles that allow for statistical variation. We need to make comparisons of these reliability measures between products and within products before and after process improvement interventions. Knowing when to trust confidence intervals and when not to trust them are crucial for managers and users of MDF to make successful decisions.

Table of Contents

Chapters **Page**

List of Tables

Page

List of Figures

Page

Chapter 1

Introduction

Meeker and Escobar (2004) stress the importance of reliability for modern products and point out that "manufacturing industries have gone through a revolution in the use of statistical methods for product quality. Tools for process monitoring and experimental design are much more commonly used today to maintain and improve product quality." Consumer expectations and demands for higher quality products are growing almost as fast as the technology that makes product development possible. Meeker and Escobar (1998) indicate "customers expect purchased products to be reliable and safe" as well as to "perform their intended function under usual operating conditions, for some specified period of time."

In order to narrow the focus here, it should be mentioned that this growth in both technology and customer expectations is certainly familiar to the forest products industry. This industry has seen tremendous growth and has an impact on the economies of countries around the world plus many states in the United States, e.g., Tennessee, Maine, etc. Forest products are present in furniture, shelving, flooring, structural applications, and unquestionably many others.

According to Williams (2001) and Young and Winistorfer (1999), raw materials and labor costs were relatively low and inexpensive during the early 20th century. Instead, technology proved to be a major limitation for the enhancement of production. The quality of final products was of little concern to most forest products industries. This has changed drastically with the dawn of the information age and newer technologies.

Restrictions and regulations have been enforced that limit the availability of raw materials, thus driving prices for these materials higher. Stronger interest has been placed on producing more reliable products. Approaches in statistical quality and process management have been employed to help improve the quality of forest products.

In this thesis, as a case study, statistical reliability ideas and tools are applied to help assess, manage and improve the strength of a particular forest product, Medium Density Fiberboard or simply MDF. This particular product has undergone tremendous market growth, and international demands for MDF are steadily increasing.

During the manufacturing process of MDF, the product undergoes extensive monitoring where many process variables such as fiber-mat weight, line speed, MDF width and thickness, etc., are collected automatically via sensors. In particular, the real-time data warehouse used in this thesis stored 2,850 process variables and was obtained from a world-class North American MDF manufacturer making 100 million lineal feet of MDF per year. A smaller subset of 230 process variables was used throughout this study. The data warehouse we used was from March 19th to September 10th, 2002 containing 1,478 records (data rows). See Young and Guess (2002) for additional details. Compare, also, Guess, Edwards, Pickrell, and Young (2003).

Furthermore, the MDF product also undergoes destructive testing at different time periods of production in order to determine the strength of MDF. If this strength complies with the specifications of quality set forth by the manufacturer and consumer, the product is shipped. Workers perform this testing by sampling cross sections of MDF panels over time. A special measuring device is utilized that pulls the cross section apart and measures the strength of internal bond (IB) in pounds per square inches (psi) until

failure. In engineering settings, this is often called tensile strength. Internal bond is a key metric of quality and unfortunately is not determined via sensors. It is not automatically available, because it requires human labor. This makes reliability improvements in MDF more complicated.

Throughout this thesis, different statistical tools will be presented for analyzing and interpreting reliability data for the purposes of improving product quality. As just mentioned, the data pertaining to the internal bond of medium density fiberboard will serve as a useful example. It is the intent of this author, not to reprove the well-known theoretical underpinnings of the statistical methodologies used throughout this thesis, but rather to apply them appropriately. This approach will provide a helpful set of tools that can be used by practitioners seeking to improve the quality and reliability of their respective products.

Chapter 2 of this thesis is a required literature review that serves the reader as brief introductions to the major ideas of the thesis. This chapter also provides a useful set of literature that not only develops further and expounds on the concepts offered, but may lead the reader to other research insights that will be informative. The literature review begins with extensive background and work pertaining to the internal bond of medium density fiberboard. It is assumed that the reader (especially those from outside the forest products industry) has had little exposure to MDF, its properties, uses, and current research around the world. Therefore, key background will be presented here in the hopes of creating a more comfortable sense of familiarity for what is to come in the rest of this thesis. Next, literature review pertaining to statistical reliability methods and studies are presented in the context of demonstrating the practicality of analyzing failure-

time and strength data. The following segment of the literature review provides an approach of using information criteria for the purposes of selecting the "best" underlying statistical model for data.

The final segment of the literature review summarizes the computer intensive resampling method known as the bootstrap, its background, and current work utilizing this tool. Again, it is assumed that the reader has had little or modest exposure to the statistical methodologies presented. Greater efforts in Chapter 2 will be made to ensure clear and proper understanding of the background material.

Chapter 3 applies simple statistical tools to aid in the understanding and exploration of the reliability of medium density fiberboard. Further background on MDF will be provided along with the method used to categorize different types of MDF used in the study. Descriptive statistics and histograms are utilized along with probability plots and survival plots (Kaplan-Meier estimates) as a methodology for obtaining more information from reliability data. This method also allows for greater ease in interpretability. Probability plots help demonstrate how a data set conforms to a particular distribution. Survival plots are nonparametric plots that can be utilized when parametric models are not justified.

This thesis demonstrates that graphical exploration is a helpful way for the practitioner to understand differences in different MDF product types, e.g., product types of MDF may have different densities and thickness. The thesis also explores the sources of variation that may influence internal bond.

Chapter 4 summarizes information criteria that are helpful in objectively identifying the underlying parametric distribution of different product types of MDF. It

is common to use the models of Weibull, lognormal, and normal in various reliability settings on the original data or the transformed data. See Meeker and Escobar (1998). In particular, Akaike's Information Criteria (AIC) and further work by Bozdogan (2000) with his Information Complexity Criterion (ICOMP) will be presented. The identification of a parametric model is essential for many statistical tests and allows for better ease in estimating desired population parameters such as percentiles.

Recall that in Chapter 3, probability plots will be introduced as a graphical approach for aiding in the identification of a parametric model. In many cases, they can easily identify an underlying statistical distribution for a particular data set. However, this method is subjective and thus makes any such model determination more difficult. The theory supporting AIC and ICOMP makes probability plotting more objective by accounting for the likelihood of the underlying model. They both create a numeric "score" for each probability plot. The model with the lowest score is picked as the "best" fit of the data.

In Chapter 5, different bootstrap methods are presented for the purposes of constructing confidence intervals for model parameters. In particular, interest lies in obtaining confidence intervals for the extreme lower percentiles for the internal bond of MDF. In reliability studies, it is generally of most interest to estimate the lower percentiles. These lower numbers are more crucial for estimating percent fall out during warranty, early failures during normal usage, as well as percent falling out of the specification limits. The idea behind bootstrapping is to simulate the sampling process a specified (usually large) number of times and obtain an empirical bootstrap distribution for the desired parameter. The bootstrap distribution is then used to acquire

characteristics about the population parameter such as bias, standard error, and confidence intervals. One bootstrap method is completely nonparametric and requires no assumptions about the underlying distribution of the data.

Other methods using the bootstrap do require parametric assumptions. As different methods of bootstrap sampling exist, so do different methods for constructing bootstrap confidence intervals. In light of many different possibilities, these methods will be explored and compared. Draft recommendations will be provided regarding the circumstances that dictate which bootstrap methods are most appropriate. Overall, the objective of this portion of the thesis is to illustrate the usefulness of the bootstrap tool as an alternative to other statistical methods available such as the normal large sample approximate confidence intervals and/or the likelihood-based interval.

The final chapter of the thesis, Chapter 6, summarizes the entire thesis and also has concluding remarks. Suggestions for possible future work are referred to and explored in this chapter. The work presented in any thesis is by no means complete and/or definitive. As with many aspects of life, new ideas and technologies become available and work builds upon itself. The bootstrap and its aggressive growth is a classical example of this. The reader should note that the author plans future work in this field during his career and more work is forthcoming.

Chapter 2

Literature Review

We now begin a brief literature review that describes other research and current issues involving the topics of this thesis. In particular, the subject matter to be discussed includes:

(1) the background and uses of medium density fiberboard with respect to its internal bond,

(2) reliability data analysis,

(3) using important information criteria such as Akaike's Information Criteria (AIC) and Information Complexity Criterion (ICOMP) to alleviate subjectivity in probability plotting,

and finally

(4) bootstrapping and its applications.

Since readers outside of the forest products industry may have little prior knowledge of medium density fiberboard, substantial emphasis and background are presented first. It is assumed that the reader has some prior knowledge of statistical methods and applications. Therefore, there will be less focus in this chapter on literature that pertains to research in reliability analysis, information criteria, and bootstrapping. Rather, many of the details for this will be presented in Chapters 3, 4, and 5 of this thesis.

2.1 MEDIUM DENSITY FIBERBOARD

The financial impact of the forest products industry on worldwide economies is overwhelming. For example, in the state of Maine, the forest products industry accounted for approximately 9 billion U.S. dollars per year and in the state of Tennessee in 2000, this figure was approximately 22 billion U.S. dollars per year according to English, Jensen, and Menard (2004). There are many examples of forest products being used in furniture, cabinets, shelving, flooring, paneling, and molding. One product called medium density fiberboard (MDF), which is a wood composite, will be the focus of this section of the literature review.

Specifically, MDF is a high quality engineered timber product offering superior qualities of consistency of finish and density, freedom from knots and natural irregularities, as well as having the characteristics of strength, durability, and uniformity are not always found in natural timber. MDF is used by home building and furniture manufacturing industries worldwide and is considered an industry leader in quality and productivity.

In fact, international demands for this product are increasing. China is an aggressive growth area for MDF and produces the largest amount in the world. For example, ten new MDF manufacturing plants are opening in China. Also, to illustrate the increasing demands for MDF, one such large producer of MDF has a capacity in excess of 100 million lineal feet per year. Suchsland and Woodson (1986) and Maloney (1993) cover manufacturing practices of MDF. Figure 2.1 illustrates a comparison between MDF and common particleboard. Other illustrative comparisons can be found in Chapter 3 of this thesis.

Figure 2.1. A comparison of common particle board to MDF.

A brief literature search on MDF in Web of Science database yielded 160+ references. We then focused on MDF papers that dealt with the key variable of internal bond (IB), which yielded a more modest 21 references. IB is a key metric of quality and serves as a measure of strength for MDF. It is through destructive testing on MDF that IB information is obtained. We will briefly describe these references and others, which help provide an indication as to why studying MDF and its IB, are important. Many of the journal articles listed below are from forest products researchers from countries as varied as Japan, China, Denmark, Sweden, New Zealand, and the United States of America. It is important to note that this literature review is not exhaustive and many more useful and informative references are available on this topic.

We begin with Wang, Winistorfer and Young (2004) who investigate the "formation characteristics of the vertical density profile of MDF... . Results of laboratory studies indicate the vertical density profile of MDF is formed from a combination of actions that occur both during compaction and also after the press has reached final position." They assert that methodologies for the formation of density profiles for oriented strandboard (OSB) discussed in Wang and Winistorfer (2000a) apply, also, to the formation of density profiles for MDF. It was determined that "high-density surface layers are easier to create in MDF than in OSB."

Widsten, Laine, Tuominen and Qvintus-Leino (2003) study IB being improved by higher defibration temperatures. It was seen that other improvements in MDF properties were also obtained by this approach.

We note that van Houts, Winistorfer and Wang (2003) comment, "Acetylation is a treatment known to reduce the swelling and water absorption behavior of wood." They

found IB reliability was reasonable for acetylation treated products.

Tsunoda, Watanabe, Fukuda and Hagio (2002) "examined the resistance of medium density fiberboard treated with zinc borate to fungal and termite attack." They observed that this treatment led to no significant loss in IB and other MDF properties.

Rials, Kelley and So (2002) used near infrared spectroscopy for predicting various characteristics of MDF samples. One of the characteristics was IB. This would provide a useful process improvement tool for manufacturing in the future. That is, this prediction method may help alleviate some of the destructive testing that is currently necessary to obtain IB information.

Young and Winistorfer (2001) use simple autocorrelation time series and process improvement techniques on MDF thickness, rather than IB. Wang, Winistorfer, Young and Helton (2001) study MDF produced in the laboratory. They used a technique called the "step-closure pressing schedule" which helped increase IB in analyzed samples. They also observed, "greater core density did not result in higher internal bond strength."

"Internal bond (IB) strength ... was closely related with carbonate types and level used" according to Park, Riedl, Hsu and Shields (2001). They conducted a study to "optimize hot pressing time and adhesive content for the manufacture of three-layer medium density fiberboard (MDF) through the cure acceleration of phenol-formaldehyde (PF) adhesives... ." In particular, they used three carbonates (propylene carbonate, sodium carbonate, and potassium carbonate) in their study.

Han, Umemura, Zhang, Honda and Kawai (2001) study fiberboard manufactured from reed straw or from wheat straw. These MDF's had IB ten times higher than particleboard. The thickness swelling, however, of the wheat MDF did not meet industry

fiberboard standards although all other properties were met. The IB of both reed and wheat MDF's did meet industry standards.

Also, van Houts, Bhattacharyya and Jayaraman (2000) contend that due to "the moisture and temperature gradients developed during hot pressing of medium density fibreboard (MDF), residual stresses occur within the board as it equilibrates to room conditions." They study the measurement of these residual stresses and show how these can affect MDF properties. In particular, they studied the effect of residual stresses on internal bond.

In van Houts, Bhattacharyya and Jayaraman (2001a) it was shown that a method known as the "Taguchi method of experimental design can be utilized to investigate methods for relieving the residual stresses present in medium density fibreboard (MDF)." The Taguchi method involves subjecting various MDF panels to varying degrees of heat, moisture, and pressure.

Furthermore, van Houts, Bhattacharyya and Jayaraman (2001b) report on the "Taguchi analysis of the internal bond strength, surface layer tensile modulus, surface layer tensile strength and thickness swell of the treated specimens. These properties were measured to indicate whether the treatments had any effect on panel strength and dimensional stability." They find that moisture and heat had little influence on IB for 8 mm MDF. Heat increases IB in 17 mm board but moisture had little effect.

Chow, Bao, Youngquist, Rowell, Muehl and Krzysik (1996) studied the "effects of fiber acetylation, resin content, and wax content on mechanical properties of dry-process hardboard made from aspen and pine... ." The results from the investigation revealed that "Tensile stress parallel to face and internal bond (IB) were generally higher

for untreated boards than for acetylated boards." Increases in resin content increased IB while increases in wax content showed a decrease in IB.

Hsu (1993) employed a self-sealing steam press system for the effective production of phenol-formaldehyde-bonded fiberboard. The relationship between the press system and board properties was the focus of the study. It was determined that "the most significant factor affecting...internal bond is resin content.... Although mat consolidation time before steam injection affects other board properties...it has no significant effect on...internal bond."

Gomez-Bueso, Westin, Torgilsson, Olesen and Simonson (2000) report on "Lignocellulosic fibers of different origins ... acetylated in large batches. The fibers used were of commercial, medium density fiberboard (MDF) pulp quality produced from softwood, beech, waste wood (low quality residue from an intermediate forest cutting) and wheat straw, respectively. Fiber from de-inked, semi-bleached, recycled paper was also included in the study." This composite fiber MDF had great properties such as thickness swelling being decreased around 90% and mechanical characteristics were modestly enhanced. They performed what is called in reliability circles an "elephant test." It involved "cyclic testing according to EN 321, (three cycles, each comprising 72 h water immersion, 24 h freezing at -18 degrees C and 72 h drying at 70 degrees C) show that more than 90% of the internal bond, IB, remained after the testing. This value can be compared with the corresponding value of 30-40% obtained for fiberboard's made from unmodified fibers."

Wang, Chen and Fann (1999) investigated "a compression shear device for easy and fast measurement of the bonded shear strength of wood-based materials to replace

the conventional method used to evaluate internal bond strength (IB)." They found that measuring strength of MDF or particleboard by the suggested compression shear strength and by the conventional approach of internal bond strength were significantly correlated. This provides an alternative approach to measuring strengths of materials.

Young and Guess (2002) present modern high technology approaches to managing the manufacturing of MDF and related data with real time process data feedback. They employ a useful regression model to predict the MDF strength of internal bond by using knowledge on more than 230 process variables.

Guess, Edwards, Pickrell and Young (2003) apply statistical reliability tools to manage and seek improvements in the internal bond of MDF. As a part of the MDF manufacturing process, the product undergoes destructive testing at various intervals to determine compliance with customer's specifications. Workers perform these tests over sampled cross sections of the MDF panel to measure the IB in pounds per square inches until failure. They explore both graphically and statistically this "pressure-to-failure" of MDF.

For other references and work that connect to MDF and internal bond among many others, see Park, Riedl, Hsu and Shields (1998), Ogawa and Ohkoshi (1997), Xu, Winistorfer and Moschler (1996), Yusuf, Imamura, Takahashi and Minato (1995), Xu and Winistorfer (1995), Hashim, Murphy, Dickinson and Dinwoodie (1994), Labosky, Yobp, Janowiak and Blankenhorn (1993), Chow and Zhao (1992), Butterfield, Chapman, Christie and Dickson (1992), and Rowell, Youngquist, Rowell and Hyatt (1991).

For more general and specific information, see also the following websites: http://web.utk.edu/~tfpc/ and http://www.spcforwood.com.

2.2 RELIABILITY

We now do a brief review of reliability literature. Reliability data analysis along with studying the IB of MDF is the core of the rest of this thesis. Currently, the best single source of statistical reliability analysis is Meeker and Escobar (1998). This "excellent and comprehensive book on reliability methods and their applications" was listed as number one in the top five recommended books for statisticians by Ziegel (2003) in the September 2003 edition of Amstat News.

According to Meeker and Escobar (1998), "Reliability is often defined as the probability that a system, vehicle, machine, device, and so on will perform its intended function under operating conditions, for a specified period of time." They also list some of many reasons for collecting reliability data such as "assessing characteristics of materials over a warranty period or over the product's design life", "predicting product warranty costs", and "tracking the product in the field to provide information on causes of failure and methods of improving product reliability." Meeker and Escobar (1998) provides an excellent and thorough treatment of the Weibull distribution, other reliability functions, and validating/exploring graphically these models.

For reliability studies and in particular for the case of analyzing strengths of materials, it is common to first consider the standard Weibull distribution as the underlying model. According to Cox and Oakes (1984), it was Fisher and Tippett (1928) that introduced the Weibull distribution when working with the extreme value distribution. In fact, even Weibull himself analyzed strengths of materials during the 1930's and found that the standard normal distribution did not fit his examples well. See Weibull (1939) as well as Chapter 3 of this thesis for more information on this famous

distribution. Also, see Cox and Oakes (1984) for more on the Weibull distribution as well as the analysis of reliability data in general. Extreme values have been studied by the Russians also, independently of Weibull.

Guess, Edwards, Pickrell, and Young (2003) shows that in the case of analyzing the strength of the internal bond of medium density fiberboard, it was natural to first consider the Weibull distribution. Although it sometimes fit parts of the IB data, it was surprisingly not a valid model for the total range of the internal bond. It was determined that other parametric distributions for strengths be investigated and ultimately, a nonparametric approach was needed. For more on specific parametric and/or nonparametric reliability models, see Meeker and Escobar (1998). Hollander and Wolfe (1973) provide an insightful and thorough coverage of nonparametrics in general.

Young and Guess (2002) "focuses on how modern data mining can be integrated with real-time relational databases and commercial data warehouses to improve reliability in real-time" for forest products. In particular, interest lies in improving reliability in the manufacturing of medium density fiberboard. This improvement is called for since the "cost of unacceptable MDF was as large as 5% to 10% of total manufacturing costs" and "prevention can result in annual savings of millions of U.S..."

In Walker and Guess (2003), the reliability of the bursting strengths of two designs for polyethylene terephthalate bottles were compared. It was determined that neither design was more reliable than the other. Furthermore, they stress "(1) the need of operational clear definitions for reliability, (2) the need of graphical exploratory analysis to discover anomalies in the data, (3) the value of nonparametric methods, and (4) the problems of using parametric techniques when the assumptions are violated."

Urbanik (1998) presents "multiple load levels" for corrugated fiberboard and "related them to the probability of time to failure." A reliability analysis was conducted for the logarithm of failure time data varying with load level. The "results were used to (a) quantify the performance of two corrugated fiberboards having significantly different components and (b) show that a safe-load-level test using multiple load levels and cyclic humidity is more sensitive to material strength differences than a dynamic edgewise compression test at standard atmospheric conditions."

Kim, Guess and Young (2004) discuss the usage of data mining tools in reliability applications. Caution is given for the use of decision trees in such applications and a new tool known as GUIDE is utilized for the purposes of comparison to regression techniques. They present a case study that focuses on predicting the internal bond of medium density fiberboard based on product specifications such as density, thickness, and width.

Many other excellent sources of information on reliability and its applications exist. Guess, Walker and Gallant (1992) focus on how different measures of reliability such as means, medians, percentiles, etc. can be interpreted and used. For a thorough treatment of the underlying theory of reliability and life testing, see Barlow and Proschan (1965, 1974, and 1981). See also and compare texts by Lawless (2003), O'Connor (1985), and Mann, Schafer and Singpurwalla (1974).

2.3 INFORMATION CRITERIA

More details and specifics regarding information criteria will be presented in Chapter 4 of this thesis. The focus is not broad and we present only several of many excellent sources on information criteria. Recall that probability plots provide a

graphical demonstration of how a particular data set conforms to a specific probability distribution. That is, the data are ordered and then plotted against the theoretical order statistics for the desired distribution. If the data set "conforms" to that particular distribution, the points will form roughly a straight line. More information on probability plotting can be found in Chapter 6 of Meeker and Escobar (1998).

Many statistical techniques, such as probability plotting just described, are very subjective and allow for decisions to be made based on someone's own personal assessment. However, choosing a model based on information criteria is much more objective and helps relieve much of the ambiguity that is present when looking solely at a probability plot. See the excellent review article of Bozdogan (2000) and the lecture notes of Bozdogan (2001). For additional comments, also compare for example Bozdogan and Bearse (2003), Urmanov, Gribok, Bozdogan, Hines and Uhrig (2002), and Bozdogan and Haughton (1998). See these papers and their extensive references plus Professor Bozdogan's helpful website where his lecture notes are available: http://web.utk.edu/~bozdogan/Stat563_2003 .

This approach greatly helps remove the subjectivity in analyzing plots. It allows an objective score where the model with the lowest score wins as the best. We urge this as an important additional tool for practitioners and engineers in deciding on a parametric model. In our Chapter 4, we will discuss these in greater detail.

2.4 BOOTSTRAPPING

According to Efron and Tibshirani (1993), "the bootstrap is a data-based simulation method for statistical inference.... The use of the term bootstrap derives from

the phrase *to pull oneself up by one's bootstrap*." Efron's bootstrap is a Monte Carlo simulation that requires no parametric assumptions about the underlying population from which the data is drawn. It is a computationally intensive statistical method that can require a large number of iterations and hence usually requires the use of the computer.

More recently, Chernick (1999) is another excellent book that provides a thorough and insightful treatment of many bootstrap methods and their applications. A student, practitioner, or others beginning to learn more about bootstrapping would do well to start here. Chernick (1999) has an extensive helpful bibliography, also. We follow the approach of Meeker and Escobar (1998) who use bootstrapping to estimate percentiles while others might consider cross-validation or jackknifing as in Giudici (2003).

Hall (2003) provides a brief and useful prehistory of the bootstrap. The idea is to further explore past connections with the bootstrap that may not be as well known. That is, "the relationship of bootstrap techniques to certain early work on permutation testing, the jackknife and cross-validation is well understood. Less known, however, are the connections of the bootstrap to research on survey sampling for spatial data in the first half of the last century or to work from the 1940s to the 1970s on subsampling and resampling."

For further readings on bootstrap methodology, theory, and applications, including some of Efron's earlier work, see Efron (2003), Efron and Tibshirani (1991), Efron and Gong (1983), and Diaconis and Efron (1983) among many others.

Meeker and Escobar (1998) present bootstrapping related directly to reliability data analysis. Also, see the helpful insights and comments on limitations of the bootstrap in Ghosh, Parr, Singh and Babu (1984). In addition, compare Parr (1983, 1985a, and

1985b).

A completely nonparametric bootstrap, or Efron's bootstrap is conducted as follows according to Martinez and Martinez (2002):

1. Beginning with a random sample denoted by **x**, calculate an estimate for some parameter, θ.
2. Sample with replacement from **x** to obtain $\mathbf{x}^{*b}$, where b represents the b^{th} bootstrap replicate.
3. Using $\mathbf{x}^{*b}$, calculate an estimate for θ.
4. Repeat steps 2 and 3 a large number of times.
5. Use the distribution of the estimates for θ to obtain desired characteristics such as standard error, bias, and confidence intervals.

Other forms of the bootstrap and methods of implementation have since emerged from Efron's earlier work. The above nonparametric method, other bootstrap methods, and the general bootstrap methodology for constructing bootstrap confidence intervals will be described in greater detail in this thesis' Chapter 5.

Chapter 9 of Meeker and Escobar (1998) provides a useful treatment of bootstrap methods and applications for reliability data. They present two methods of bootstrap sampling, (1) the fully parametric bootstrap that includes parametric sampling for parametric inference and (2) nonparametric bootstrap sampling for parametric inference. In both cases, it is necessary to first determine the underlying parametric distribution of the data. Applications of these methods to the construction of confidence intervals are presented in large detail and in particular, the bootstrap-t method for constructing confidence intervals is dealt with thoroughly. For an excellent treatment of the

construction of confidence intervals and other statistical intervals in general, see Hahn and Meeker (1991).

Pages 217-220 of Chapter 9 in Meeker and Escobar (1998) reviews the same fully nonparametric bootstrap method as presented in Efron and Tibshirani (1993) as it applies to analyzing reliability data. They capture some of the limitations and present warnings for the use of this fully nonparametric bootstrap. Recall, also, Ghosh, Parr et al. (1984).

For example, Meeker and Escobar (1998) contend that the "justification for the bootstrap is based on large-sample theory. Even with large samples, however, there can be difficulties in the tails of the sample. For the nonparametric bootstrap, there will be a separate bootstrap distribution at each time for which there were one or more failures in the original sample." This would not pose a problem outside the tails of the original data where the bootstrap distribution will be approximately continuous. However, in the extreme tails of the original data, there may be only a small number of failures or outcomes. In this case, the bootstrap distribution may be anything but continuous. As can be seen by the examples presented, when the extreme tails are of interest (as is often the case in reliability studies), the standard fully nonparametric bootstrap methods are not as useful as other bootstrap methods. Rather the standard bootstrap methods have a place when estimating parameters such as the quartiles (25^{th} or 75^{th} percentiles).

Chapter 3 of Chernick (1999) mentions several different methods for the construction of bootstrap confidence intervals. These include the standard percentile method, the bias corrected and accelerated percentile method, the iterated bootstrap, and the bootstrap-t. The aforementioned methods are described concisely and their ranges of

applicability as well as limitations are discussed. The last chapter provides information on the overall limitations of the bootstrap and when the bootstrap fails. These include, but are not limited to, (1) a sample size that is too small and (2) when attempting to estimate extreme values. Also, compare with Efron and Tibshirani (1993) and Davison and Hinkley (1997) as they provide further and extensive information on bootstrap theory, methods, and limitations.

Recall that the bootstrap is a computer intensive statistical method. Martinez and Martinez (2002) devote Chapter 6 to the bootstrap, including estimation of standard error, bias, and confidence intervals, using MATLAB. Several routines are provided along with examples to aid the reader in getting started. Other Monte Carlo techniques such as the jackknife are also discussed.

Lunneborg (2000) shows how to construct bootstrap confidence intervals using Resampling Stats and/or S-PLUS statistical programming packages. This thorough work provides step-by-step algorithms for implementation of many resampling methods for the purposes of analyzing data.

DiCiccio and Efron (1996) is devoted to bootstrap confidence intervals and provides a useful survey of many such intervals (standard, percentile, bootstrap-t, etc.). Its focus is to "improve by an order of magnitude upon the accuracy of the standard intervals...in a way that allows routine application even to very complicated problems." Examples for each method are provided and the underlying theory is also presented.

Polansky (1999) shows that bootstrap confidence intervals constructed using percentile methods "have bounds on their finite sample coverage probabilities. Depending on the functional of interest and the distribution of the data, these bounds can

be quite low." It is said that the "bounds are valid even for methods that are asymptotically second-order accurate."

Boos (2003) gives his own thoughts on bootstrapping. In particular, he contends that "the real reason the bootstrap was so path-breaking and has remained so popular is that Efron described it mainly in terms of creating a 'bootstrap world,' where the data analyst knows everything." In a sense, any population parameter of interest can be estimated simply through simulation. For example, "if the variance of a complicated parameter estimate in this world is desired, just computer generate B replicate samples (bootstrap samples or resamples), compute the estimate for each resample and then use the sample variance of the B estimates as an approximation to the variance." Thus, "In effect this bootstrap world simulation approach opened up complicated statistical methods to anybody with a computer and a random number generator."

Chapter 3

Exploring Graphically and Statistically the Reliability of Medium Density Fiberboard

This chapter is a slightly revised version of a paper by the same name published in the *International Journal of Reliability and Applications* in 2003 by Frank M. Guess, David J. Edwards, Timothy M. Pickrell, and Timothy M. Young:

Guess, F. M., Edwards, D. J., Pickrell, T. M., and Young, T. M. (2003). Exploring Graphically and Statistically the Reliability of Medium Density Fiberboard. *International Journal of Reliability and Applications*, 4(4), 97-110.

My primary contributions to this paper include (1) all of the computer work (charts, graphs, etc.), (2) many of the interpretations presented, and (3) portions of the writing.

3.1 INTRODUCTION AND MOTIVATION

Medium Density Fiberboard (MDF) is used internationally in a host of building needs and furniture construction. It is a superior engineered wood product with great strength, reliability and grooving ability for unique designs. In addition, MDF offers superior qualities on consistency of finish and density, plus freedom from knots and natural irregularities.

MDF has characteristics of strength, durability and uniformity not always found in natural timber or standard particleboard. It has excellent machinability due to its homogenous consistency and smaller variation in needed characteristics compared to natural wood. These features make MDF particularly suited for use in flooring, paneling, and manufacturing of furniture, cabinets, and moldings. It, also, has environmentally friendly properties of using wood waste to manufacture useful byproducts. This does not happen as easily with traditional wood products.

Figure 3.1 and Figure 3.2 demonstrate the marked differences between

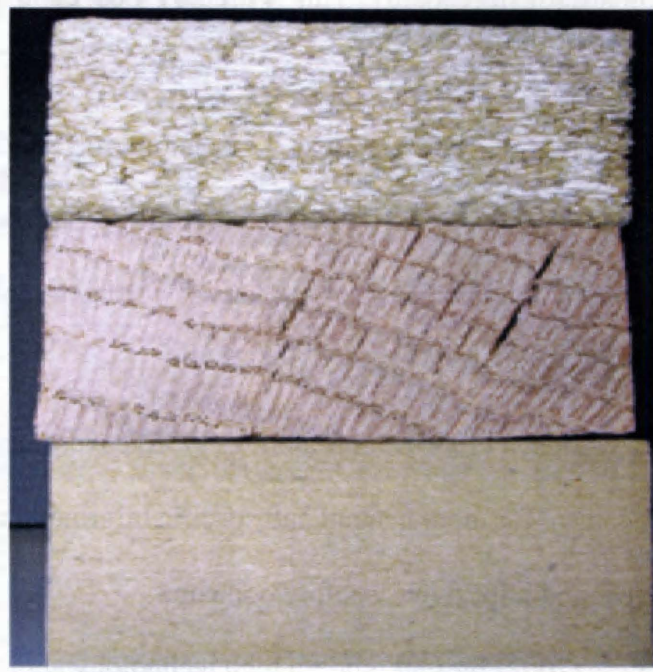


Figure 3.1. Cross sections of forest products from width view. Top is particleboard, middle is natural wood, and bottom is MDF.

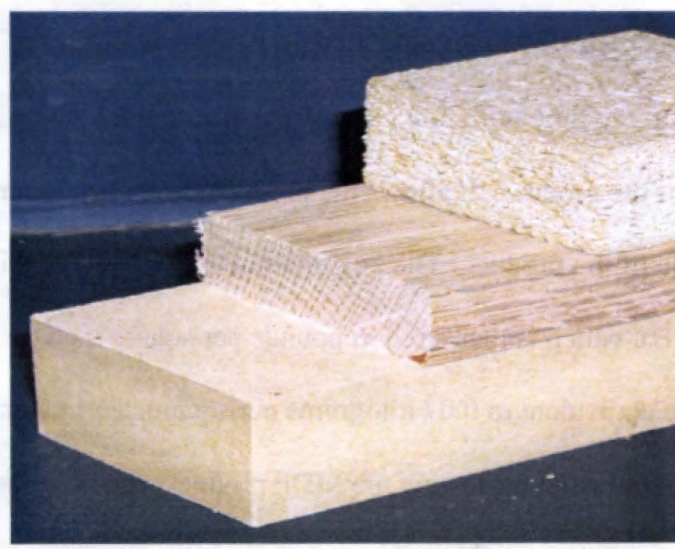


Figure 3.2. Cross sections of forest products from diagonal view. Top is particleboard, middle is natural wood, and bottom is MDF.

particleboard, natural timber, and MDF. We have observed the common usage in writing particleboard or fiberboard (or fibreboard). This is used by most companies or timber associations, which we sampled. Our sample ranged from Georgia-Pacific, Sabah Timber Industries Association of Malaysia, to an Association of New Zealand Forestry Companies. In some settings, however, you will see it written as two words "particle board." We followed the more typical industrial usage of writing as one word "particleboard" or "fiberboard."

Suchsland and Woodson (1986) and Maloney (1993) cover manufacturing practices of MDF. See, also, for more general and specific information: http://web.utk.edu/~tfpc/ and http://www.spcforwood.com.

Young and Guess (2002) present modern high technology approaches to managing the manufacturing of MDF and related data with real time process data feedback. They employ regression prediction of MDF strengths using knowledge on more than 230 process variables.

Here, we are interested in the statistical reliability properties of *the strength to failure* as opposed to the time to failure. The strength to failure data will give us a clear idea of the utility of the product. It allows the producer to make assurances to customers about the useful life of the product. The key measure of MDF's reliability and quality is its internal bond (IB), which is measured in pounds per square inch (psi) until breaking. Note that one psi is equivalent to 0.07 kilograms per square centimeter.

For a number of reasons, testing the MDF product types over any extended time, under a variety of operating conditions is not economically feasible. Instead of an elaborate procedure of designing and implementing life tests, a "pull apart" destructive

approach provides an instantaneous measurement of the IB. Destructive "pull apart" test samples are taken periodically to determine IB. Due to the cost and loss of products from destructive testing, manufacturers obviously want to keep such tests to a minimum.

Another advantage to destructive testing is immediate feedback into the manufacturing process leading to rapid process improvement. Also, it can help modify or stop the manufacturing process; thus, preventing great waste of materials. The cost of unacceptable MDF was as large as 5% to 10% of total manufacturing costs, which can result in 10 to 15 million dollars per plant per year. In 2003 to 2004, ten such high production plants are anticipated to be built in Asia.

We present new results on different IB data regarding the statistical distributions for various product types of IB. This is important for studying potential warranty issues, understanding wearing over time, failure of products under misuse, and variation in IB between product types. This is, also, needed within particular product types of MDF. Our data covers the time period from March 19, 2002 to September 10, 2002.

We explore graphically and statistically the distributions of the strengths of this material. It would be natural to consider first the standard Weibull model for strengths of materials. Indeed, the researcher Weibull himself first analyzed strengths of different materials, ranging from cotton to metal. From his data sets, he found the primary available distribution of the normal did not fit his examples well in the 1930's. The alternative parametric model he originally proposed is what we now call the "three parameter" Weibull. See Weibull (1939 and 1951).

The original three parameter Weibull is often reduced and written today as a two parameter distribution. Recall this two parameter Weibull density function can be written

in the following parameterization as

$$f(x) = \lambda\beta x^{\beta-1} e^{(-\lambda x^{\beta})} \tag{3.1}$$

where $x \geq 0$ (and $f(x) = 0$, for $x < 0$), while the reliability function is

$$\overline{F}(x) = e^{(-\lambda x^{\beta})} \tag{3.2}$$

where $x \geq 0$ (and $\overline{F}(x) = 1$ for $x < 0$). Another common reason for modeling data with a Weibull distribution is that it may be suitable for either increasing, constant (i.e., an exponential) or decreasing hazard functions. For MDF subject to destructive "pull apart" tests, we would conjecture an increasing failure rate. This would lead us to hypothesis, a priori, that the shape parameter $\beta > 1$ for any MDF product type that might be Weibull.

See, for example, the excellent book of Meeker and Escobar (1998) for a thorough treatment of the Weibull, other reliability functions and validating/exploring graphically these models. Also, compare texts by O'Connor (1985), Barlow and Proschan (1981), etc.

Although the Weibull can sometimes fit parts of our IB data for some categories of MDF, it is surprisingly not a valid model for this data for the total IB range. Other parametric distributions of strengths are employed and a nonparametric approach is needed. See Meeker and Escobar (1998) for reliability parametric/nonparametric models and the insightful Hollander and Wolfe (1999) on nonparametrics, in general.

The spirit of this chapter is that of an exploratory analysis via graphs, descriptive statistics, and tests. See the excellent overview on graphics by Scott (2003) and his references. Section 3.2 covers the types of MDF and ways these types are determined. Section 3.3 explores both graphically and statistically particular types of MDF, while

Section 3.4 provides concluding comments.

3.2 CATEGORIZING TYPES OF MDF

We begin the analysis by sorting the IB data by three key characteristics:

- density (lbs/ft^3)
- thickness (inches)
- and width (inches).

These three characteristics differentiate the MDF's for various applications. Since MDF in this particular study was produced in continuous length of sheets, length was not a crucial variable for our purposes. Further, for the purpose of analysis, the MDF was separated into two main groups:

- Group I: standard density
- Group II: high density.

The *high density* type is MDF with densities on the upper end of the scale 47-48 pounds per cubic foot (752.86-768.88 kg/m^3). The *standard density* type is the MDF with densities ranging from 45-46 pounds per cubic foot (720.83-736.85 kg/m^3).

Within each group the IB was measured in accordance with classification by density, thickness and width. The type numbers (with density, thickness, and width after each type number) are listed in Table 3.1.

Since there were a number of types in each group, we select the primary types, which sold the most, for a more detailed analysis. These were Types 1 and 3 from Group I (standard density) and Types 2 and 5 from Group 2 (high density). See Table 3.1 for more details.

Table 3.1. Group and type numbers for different MDF products with (A, B, C) where A = density, B = thickness, and C = width where for example Type 1 in Group I represents A = 46 pounds per cubic foot, B = 0.625 inches thickness, and C = 61 inches (or the equivalent metric units).

Group I: standard density Type #	Group II: high density Type #
1 (46,0.625,61)	2 (48,0.75,61)
3 (46, 0.75,49)	5 (48,0.625,61)
4 (46,0.75,61)	9 (48,0.75,49)
6 (45,1,61)	10 (48,0.375,61)
7 (46,0.625,49)	11 (48,0.5,61)
8 (45,1,49)	14 (48,0.5,49)
12 (46,0.688)	15 (48,0.625,49)
13 (46,0.688,49)	16 (47,1,61)
17 (45,1.125,61)	18 (48,0.4379,49)
20 (46,0.875,61)	19 (48,0.375,49)
22 (45,1.12549)	21 (48,0.563,61)
	23 (48,0.4379,61)
	24 (48,0.688,61)

Another reason behind our splitting into two distinct groups was that the destructive testing of the MDF is concerned mainly with the IB strength. A priori, this is reasonably hypothesized to be mainly a function of density. Furthermore, from the nature of the destructive testing, which involved the cutting of many cross sections from different pieces of MDF, the lengths and widths were not the major factors effecting strengths. In the next section graphs and statistical tests will demonstrate strikingly this hypothesis to be true.

3.3 EXPLORING GRAPHICALLY AND STATICALLY IB IN TYPES OF MDF

The initial analysis began with the assessment of the underlying distribution of the internal bond strengths, categorized by the density, thickness and width measurements. We want to first understand means, medians, and percentiles of the strengths of MDF. See, for example, Guess, Walker, and Gallant (1992) for more on these measures.

Table 3.2 provides a descriptive statistics comparison of product Types 1 and 2. These numbers have been rounded to one decimal place. Note that both the mean and median in Type 1 are 120.2, while for Type 2 the mean and median are close at 180.0 and 179.0. Type 1 has less variation as measured by the IQR and standard deviation, but

Table 3.2. Descriptive statistics comparison of Types 1 and 2.

	Type 1 IB (psi)	Type 2 IB (psi)
Mean	120.2	180.0
Median	120.2	179.0
Std. Dev.	9.9	12.3
IQR	12.3	17.6
Min	87.2	140.6
Max	164.5	214.5

Type 1 has a bigger range of 77.3 when compared to Type 2 being 73.9. This bigger range for Type 1 can be understood by its outliers, boxplots, and the histograms in Figures 3.3 and 3.4.

From the histogram in Figure 3.3, we see that the distribution of the primary product, Type 1, is approximately normal. Recall the mean and median being the same. Figure 3.4 suggests that we explore the reasons behind the weakness of the units in the 140 to 150 psi bins, plus understand the much better strengths in the higher bins overall, especially for the 190 to 210+ psi bins, in order to improve the reliability.

Recall the exploratory flavor of Tukey of examining many views of the same data. See, also, Scott (2003). Figure 3.5 is an overlay plot that gives another look at the differences and similarities between Types 1 and 2. Notice it can be a little misleading when compared to the actual raw data or the histogram. The plot shows quite a distinction between the two product types, providing evidence that Type 2 is much stronger than Type 1. That is, heavier products or products requiring more load bearing strength, such as shelving, would make use of Type 2 MDF. Type 1, with less strength, would be used more extensively in products not requiring large strength, such as picture frames.

Probability plots were used extensively in this analysis because they give a clear demonstration of how a particular data set conforms to a specific candidate probability distribution. The data are ordered and then plotted against the theoretical order statistics for a desired distribution. If the data set "conforms" to that particular distribution, the points will form a straight line. Simultaneous confidence bands provide objective bounds of deviation from the line or not. Those data points outside the confidence bands are shown to deviate from the candidate probability distribution in question. See

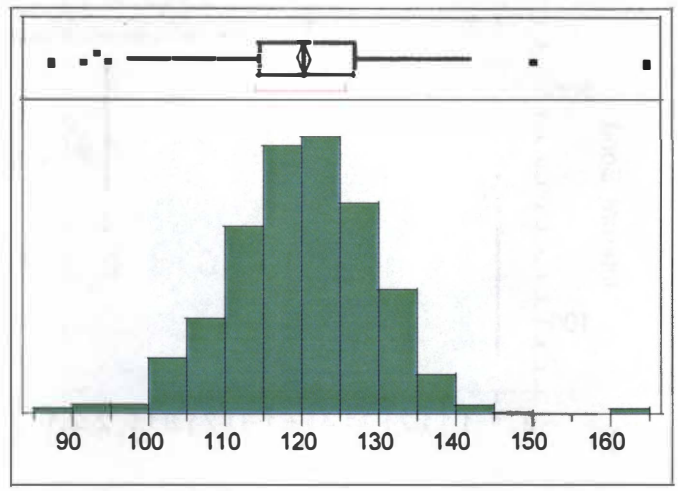


Figure 3.3. Histogram and Boxplot of Primary Product (Type 1) from JMP.

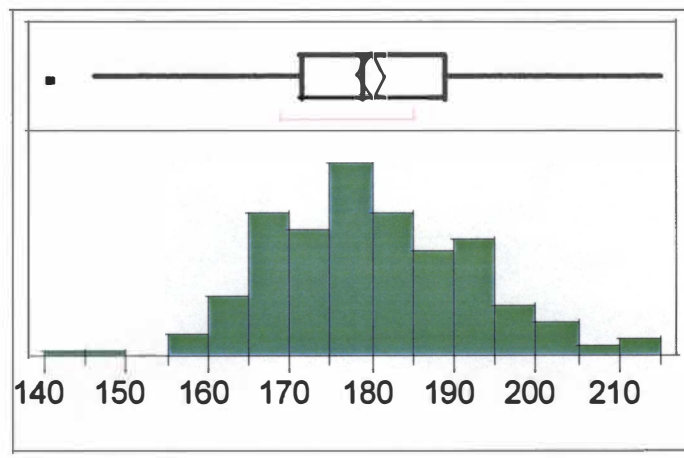


Figure 3.4. Histogram and Boxplot of Type 2 from JMP.

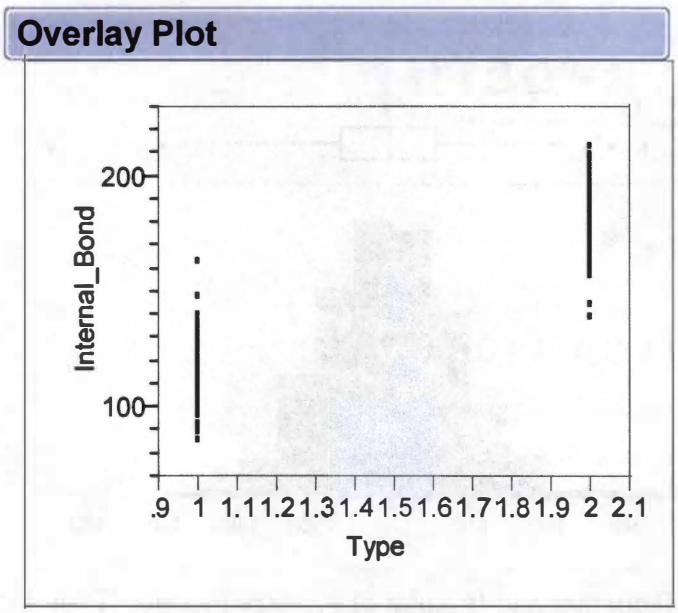


Figure 3.5. Overlay Plot of Types 1 and 2 from JMP.

Chapter 6 of Meeker and Escobar (1998) for further information. Normal and Weibull probability plots were produced for Types 1 and 2 as shown in Figures 3.6 and 3.7.

In Figure 3.6(a) for Type 1, there is clear departure from the Weibull in the upper tail, but appears to be following this distribution in the middle and the lower tail. Fitting in the lower tail can be important for estimating percent fall out of specification limits. Figure 3.6(b) for Type 2 shows clear departure from the Weibull distribution overall. The snake-like meandering is a systemic pattern that strongly suggests the Weibull does not fit at all for Type 2.

Recall that the histogram of Type 1 as well as the mean and median being the same provides some evidence for normality. Figure 3.7(a) shows a normal plot with points that fall mostly within the simultaneous bounds, except for some outliers. There is some clear departure in the tails that may not be following so perfectly a normal distribution. Again, recall the lower tail is important in estimating lower percentiles. Thus, as shown in Figure 3.6(a), the Weibull may prove to be a better model for estimating these lower percentiles for Type 1. However, this is only a conjecture and we will see in Chapter 4 that such subjective conjectures may not always hold.

Figure 3.7(b) shows less departure from normality and certainly appears to be a better fit of the data than the corresponding Weibull distribution for Type 2. In fact, as we will see later, large p-values will not allow for normality to be rejected for Type 2. Overall, neither model appears to be the best, thus; a nonparametric approach may be more appropriate.

As seen, the probability plots have been a very visual and indeed subjective

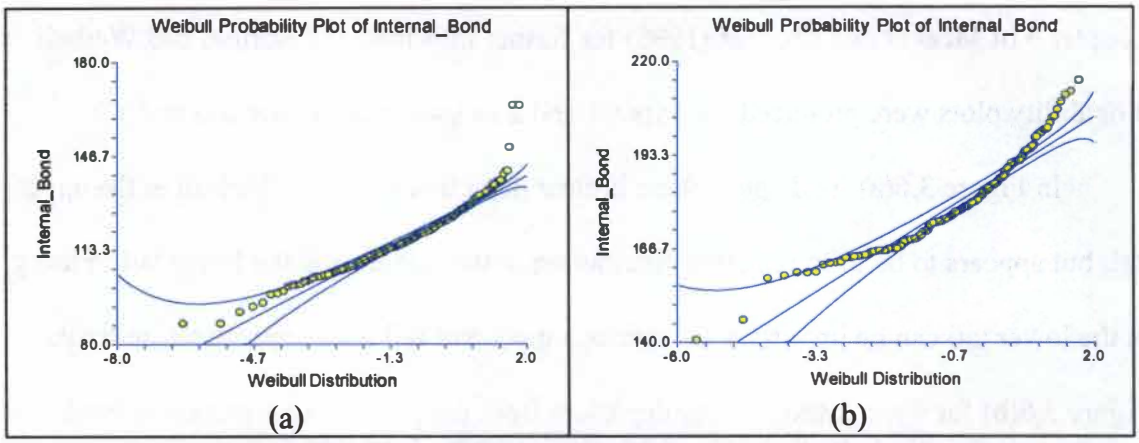


Figure 3.6. NCSS Weibull Probability Plots. (a) Weibull plot for Type 1; (b) Weibull plot for Type 2.

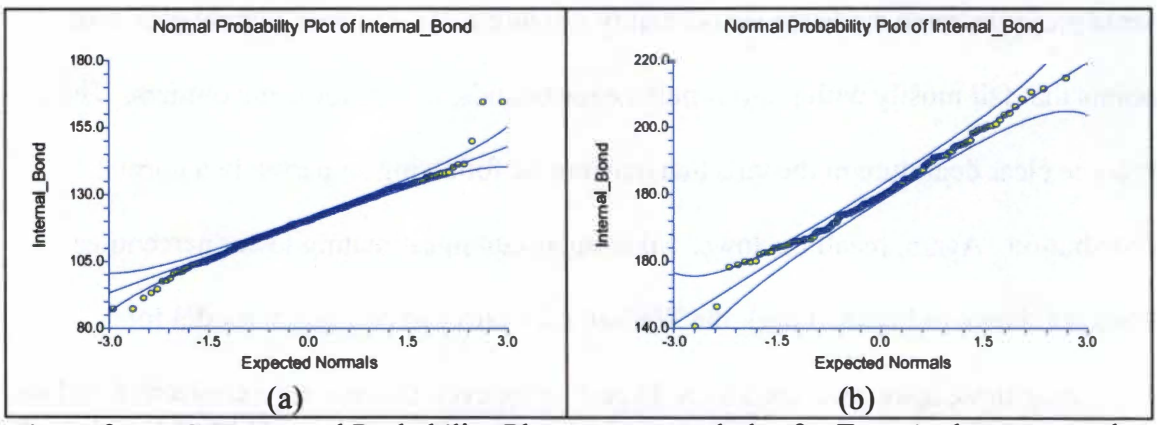


Figure 3.7. NCSS Normal Probability Plots. (a) Normal plot for Type 1; (b) Normal plot for Type 2.

method for assessing the underlying distribution for the different product types. In particular, the normal distribution was determined to be the reasonable fit compared to Weibull for some product types. We do not show all types here to save space in the thesis. Therefore, it is natural to ask if the data truly follows a normal distribution and if this is statistically significant by testing. Tests such as the Shapiro-Wilk, Kolmogorov-Smirnov, and others exist to help answer these questions more objectively. These tests will produce different p-values, as seen in the Tables 3.3 and 3.4.

For Type 1, we clearly reject normality using the Shapiro-Wilk and Anderson-Darling test for alpha level of 0.05. The Kolmogorov-Smirnov test has bigger p-values, but recall it tends to have low power. Notice that four different software packages (SAS, JMP, NCSS, and Minitab) were used for checking the consistency of the tests statistics and p-values. Table 3.4 with its larger p-values for all three tests shows we can not reject normality for any reasonable alpha levels. Still we may want to seek better understanding by other plots. Walker and Guess (2003) stress the need for more nonparametric plots and analysis, when the parametric models may be weak or not the strongest. Nonparametric plots known as Kaplan-Meier estimators, survival plots, or reliability plots will now be shown for various product types.

Immediately, one should notice the large gap present between Type 1 and 2 in the Kaplan-Meier nonparametric survival plots in Figure 3.8. Recall further that the medians are very different. That is, 120.3 and 179.0 psi for Types 1 and 2, respectively. Based on this, Type 2 has even more evidence of being significantly stronger than Type 1. A two sample t-test was conducted with variances assumed unequal. This assumption was based on a test for unequal variances provided by SAS, which yielded a p-value of

Table 3.3. Normality test comparisons for Type 1.

		Test Statistic / p-value	
Software / Test	**Shapiro-Wilk**	**Kolmogorov-Smirnov**	**Anderson-Darling**
JMP	0.97947 / <0.0001	N/A	N/A
SAS	0.97947 / <0.0001	0.0357 / >0.15	0.80213 / 0.0393
Minitab	0.9880 / <0.01	0.035 / >0.15	0.802 / 0.038
NCSS	0.97947 / 0.00002	0.03317 / N/A	0.80213 / 0.03787

Table 3.4. Normality test comparisons for Type 2.

		Test Statistic / p-value	
Software / Test	**Shapiro-Wilk**	**Komogorov-Smirnov**	**Anderson-Darling**
JMP	0.990482 / 0.2514	N/A	N/A
SAS	0.990482 / 0.2514	0.044945 / >0.15	0.485528 / 0.2311
Minitab	0.9947 / >0.1	0.045 / >0.15	0.485 / 0.224
NCSS	0.99048 / 0.25142	0.0449 / N/A	0.48526 / 0.2268

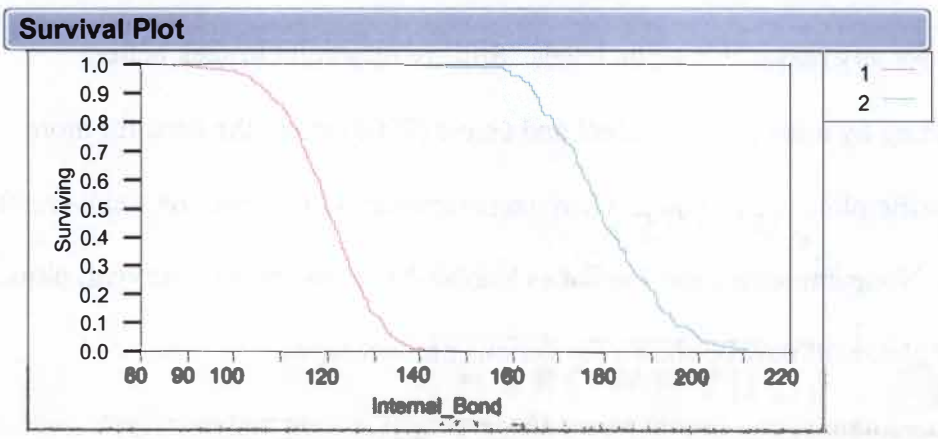


Figure 3.8. Survival Plot of Types 1 and 2 from JMP.

p=0.0004. This is quite a significant result and allows us to proceed with the stated t-test. In particular, the t-test gives a small p-value less than 0.0001, which is also highly statistically significant and allows for the conclusion that Types 1 and 2 are significantly different and thus, Type 2 is significantly stronger than Type 1.

It is appropriate to take a moment to provide the practitioner unfamiliar with survival plots with an explanation of the interpretability of these curves. Consider Figure 3.8 showing Types 1 and 2. The survival plot has the internal bond strength shown on the horizontal axis and the percentage of product surviving along the vertical axis. If we are interested in the internal bond strength where 50% of Type 1 MDF is surviving (or equivalently, where 50% have failed), simply find 0.5 on the vertical axis and move horizontally until reaching the survival curve for Type 1. Reading the horizontal axis at this point on the curve gives 120.3 psi, which is the median internal bond for Type 1 and what we would expect to obtain. Other examples follow similarly. Chapter 3 of Meeker and Escobar (1998) provides a thorough and helpful treatment of the construction and interpretation of survival curves.

Suppose that interest lies in comparing two product types of the same thickness, but with a different density. Here, we compare product Types 1 and 5. That is, Type 1 has a smaller density of 46 lbs/ft^3 while Type 5 has a higher density of 48 lbs/ft^3. However, they both have a thickness of 0.625 inches. The survival plot comparing these two products is shown in Figure 3.9. As with Figure 3.8, notice the large gap separating the two product types allowing for evidence that Type 5 is a much stronger product. Thus, we are seeing that product types of a higher density appear to be stronger than those at a lower density.

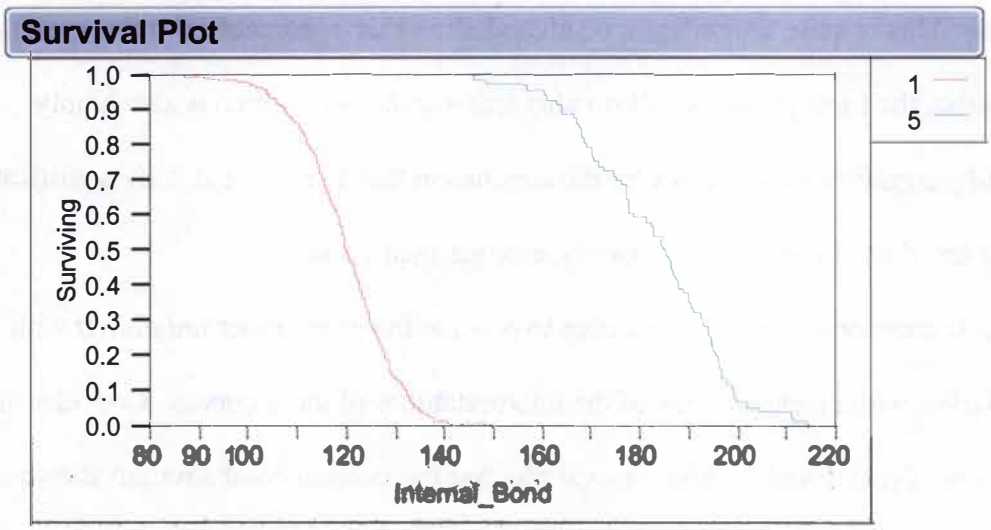


Figure 3.9. Survival Plot for Types 1 and 5 from JMP.

Instead, suppose that interest lies in comparing product types of the same density, but with a different thickness. Then, in this case, comparison is between product Types 1 and 3. Types 1 and 3 both have a density of 46 lbs/ft^3, but Type 1 has a thickness of 0.625 inches and Type 3 has a thickness of 0.75 inches. The survival plot showing this comparison is shown in Figure 3.10.

Notice that the gap we have been seeing in the plots is no longer present. This provides evidence that there are no differences among these two product types. That is, when density is held constant, thickness does not appear to have any effect on IB. However, it is important to verify this statistically. Figure 3.11 is an overlay plot of Types 1 and 3. A two-sample t-test was conducted (again, assuming unequal variances) and a p-value of 0.1988 was obtained. Thus, our suspicions are confirmed and it can be concluded that there are no statistically significant differences between Types 1 and 3 at particular levels.

A summary survival plot showing Types 1, 2, 3, and 5 is shown in Figure 3.12. From this plot, it is relatively easy to see which product types had the higher density and which had a lower density. However, it is not as obvious which product types had the higher or lower thickness making it clear that density is the main driver in determining MDF strength whereas thickness is not a large contributor to IB.

One noticeable attribute of the survival plot shown in Figure 3.12 is that the survival curves at the same density are crossing each other at some point. The explanation is quite simple. The significance of this crossing is that one product has a greater strength at lower pressures whereas the other product will surpass at higher pressures. For example, Type 2 starts out with a greater strength than Type 5 at the

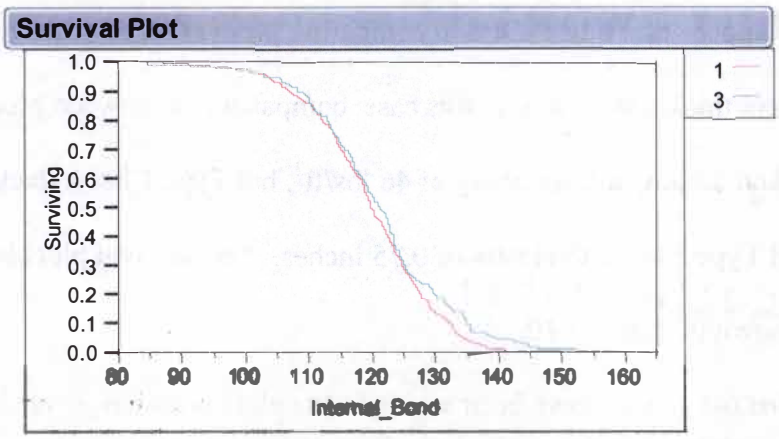


Figure 3.10. Survival Plot for Types 1 and 3 from JMP.

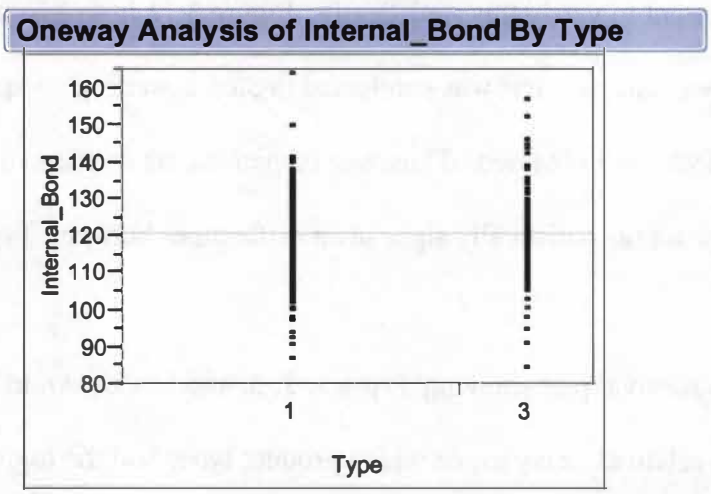


Figure 3.11. Overlay Plot of Types 1 and 3 from JMP.

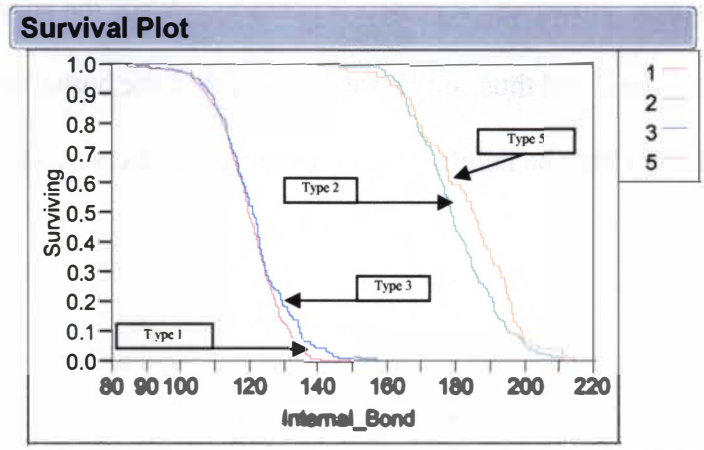


Figure 3.12. Summary Survival Plot for Types 1, 2, 3, and 5 from JMP.

extreme lower pressures. However, as pressure increases, we see the survival curve for Type 5 cross that of Type 2 and thus, surpass it in strength at the higher pressures. This crossing of survival plots may be helpful to note for manufacturers of MDF when developing new products.

3.4 SUMMARY

In conclusion, we find that exploring graphically and statistically the MDF's reliability as measured by IB means, medians, and other percentiles readable from survival plots are helpful ways for understanding each product type better. Recall Type 1 had more outliers, which suggests more need for process improvements there. Density is a key driver in improving IB average. In fact, it was the key source of variation in IB. Changes in thickness (or width) do not affect IB as much as changes in the density.

One should be aware that quality and reliability are more than just one number (not just the mean or median). We need to explore these and other descriptive statistics as well as graphs of the data. Also, be careful of potential software differences on some tests, which may be mild or sometimes severe in certain instances. Validation with a different software package than the first software analysis might be advisable. Besides histograms, survival curves are a very helpful and insightful way to view your data. These different views may surprise you, suggesting places for real world process improvements. Compare Deming (1986 and 1993). Future work on estimating C.I.'s on the lower percentiles and other sources of variation will be explored later.

Chapter 4

Using Helpful Information Criteria to Improve Objective Evaluation of Probability Plots

4.1 INTRODUCTION AND MOTIVATION

In Chapter 3, the reliability of the internal bond (IB) of medium density fiberboard (MDF) was explored graphically and statistically comparable to the approach in Meeker and Escobar (1998). In particular, probability plots and survival (reliability function) plots were utilized to allow for greater ease in obtaining the most information from the IB data and for ease in interpretability. Probability plots were discussed as a method for determining the underlying distribution of a particular data set. Recall that if the data set "conforms" to a distribution, the points on the plot will form a straight line. A shortcoming of this method, however, is the extreme subjectivity, i.e., for any given probability plot, different people may have conflicting conclusions. Chapter 6 of Meeker and Escobar (1998) provides examples of probability plots based on repeated samples of the same size. Their graphs can serve as a strong illustration of the subjectivity of plots alone.

Consider the following comparison example as shown in Figures 4.1-4.3. For this example, we make use of Type 1 MDF data. Recall that Type 1 is the primary product and is therefore richer in data than other product types. Figure 4.1 fits the normal, lognormal, and Weibull distributions using JMP Statistical Discovery Software. Figures 4.2 and 4.3 fit these same distributions, but use respectively SAS and S-PLUS software packages.

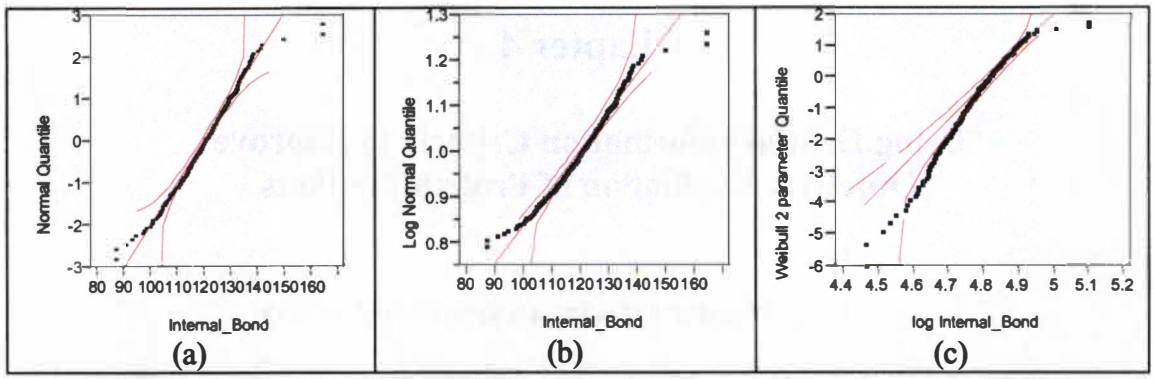


Figure 4.1. Comparing Probability Plots for Type 1 MDF using JMP. (a) Normal, (b) Lognormal, and (c) Weibull probability plots.

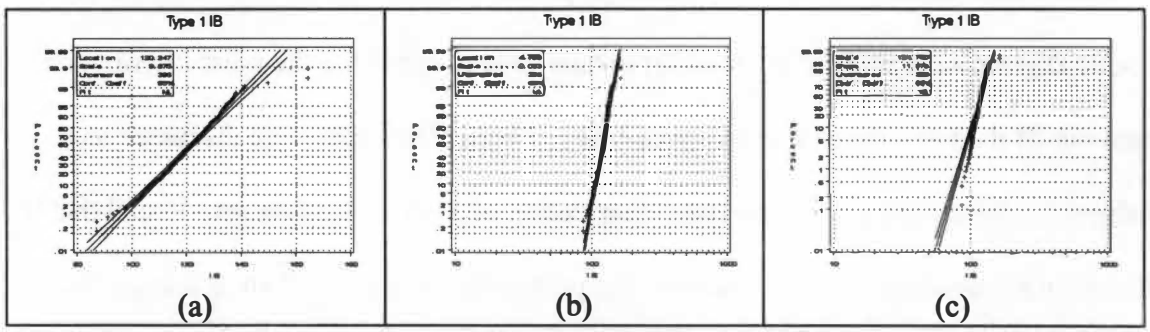


Figure 4.2. Comparing Probability Plots for Type 1 MDF using SAS. (a) Normal, (b) Lognormal, and (c) Weibull probability plots.

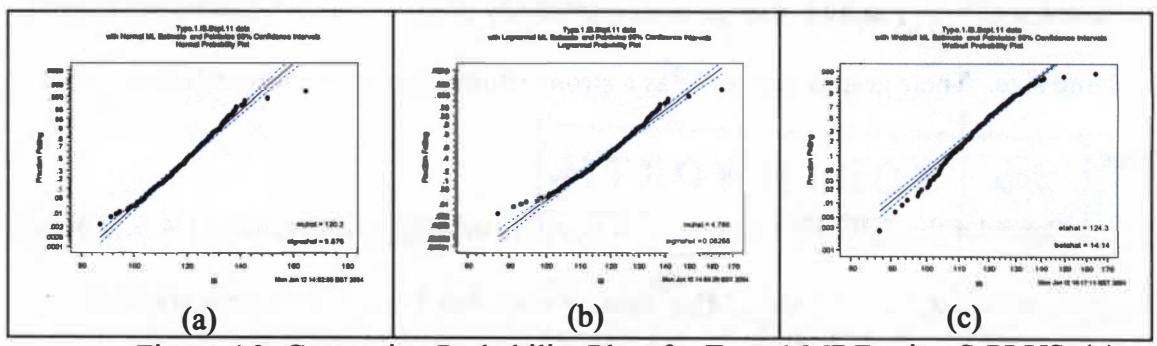


Figure 4.3. Comparing Probability Plots for Type 1 MDF using S-PLUS. (a) Normal, (b) Lognormal, and (c) Weibull probability plots.

Based solely on the plots shown in Figure 4.1-4.3, the choice between the normal, lognormal, and Weibull distributions for the IB of Type 1 MDF may be anything but straightforward. The comparisons of the packages shows that we must concern ourselves with comparing one set of probability plots within a particular statistical software package. In addition, we must examine the visual impressions between software packages. In particular, the placement of simultaneous confidence bands does not appear to agree for all three software packages. Again, we emphasize, as others, the subjectivity of graphical approaches without information criteria. Later in this section we will discuss these very helpful criteria.

Hypothetically, one may look at Figure 4.1 and suggest that the normal model provides the best fit. Likewise, a look at Figure 4.2 may suggest that the lognormal model is the best fit. Because of different scaling in the software packages, the placement of obvious outliers are different, which may also have an affect on which plot is chosen to best represent the data. It should be emphasized, also, that in many cases, probability plots are very useful and can easily identify an underlying statistical distribution for a particular data set. However, it is again further stressed that this method is subjective and thus makes model determination difficult.

Fortunately, modern developments have produced more objective approaches for determining the best candidate model for data than subjective probability plots, i.e., information criteria or model evaluation criteria. According to Bozdogan (2000), the "necessity of introducing the concept of model evaluation has been recognized as one of the most important technical areas, and the problem is posed on the choice of the best approximating model among a class of competing models by a suitable model evaluation

criteria given a data set. Model evaluation criteria are figures of merit, or performance measures, for competing models." In particular, Bozdogan (2000) reviews the basic ideas surrounding Akaike (1973) Information Criterion or AIC and then presents further work based on his Information Complexity Criterion (ICOMP) which is a new entropic model selection criteria. The theory supporting AIC and ICOMP makes probability plotting more objective by accounting for the likelihood of the underlying model, which creates a numeric "score" for each probability plot. The model with the lowest score is picked as the "best" fit of the data. See Bozdogan and Bearse (2003), Bozdogan and Haughton (1998), Urmanov, Gribok et al. (2002), Bozdogan (1990), and Bozdogan and Sclove (1984) who show how information criteria plays an important role in simple and multivariate regression analysis, cluster analysis, the detection of influential observations, etc.

The spirit of Chapter 4 of this thesis is to provide the reader with the essential background to wisely apply AIC and ICOMP. Also, we show how using these helpful information criteria can aid in the important selection of a better parametric model for the internal bond. Section 4.2 introduces and further develops the ideas behind AIC and ICOMP as it applies for use in probability plotting. Section 4.3 uses the IB data on MDF as a brief case study and shows probability plots along with their information criteria "scores" for the normal, lognormal, and Weibull distributions. It is the intent of this section to choose better underlying parametric distributions for the different MDF product types previously mentioned, i.e., Types 1, 2, 3, and 5. Section 4.4 provides concluding remarks, plus potential future work that may be conducted in the use of information criteria for quantile modeling.

4.2 A BRIEF SUMMARY OF AIC AND ICOMP

Bozdogan (2001) is an excellent place to start for very helpful background information on AIC, ICOMP, and their many applications. In this section, it is reviewed how these information criteria can be used with probability plotting to prevent the subjectivity when only using the plots.

Akaike's Information Criterion (AIC), like other model-selection methods, takes the form of a lack of fit term (such as minus twice the log likelihood) plus a penalty term. The penalty term is a "compensation for the bias in the lack of fit when the maximum likelihood estimators are used" according to Bozdogan (2001). AIC has the following form:

$$AIC = -2 \log L(\hat{\theta}) + 2k \tag{4.1}$$

where $L(\hat{\theta})$ is the maximized likelihood function for a particular population parameter θ (either scalar or vector valued) and k is the number of parameters in the model. For example, if we consider the normal model with the parameters μ and σ^2, then $k = 2$.

Recall that the model with these lowest information criteria score is chosen as the best fit of the data. In order to better understand why we take minus twice the log likelihood as the lack of fit term, we take a heuristic approach of reasoning by extremes and exponentials, in the spirit of Frank Proschan.

Consider an exponential model with failure rate parameter λ and a data set with the sole observation of x = 0. Here we have $X \sim Exp(\lambda)$ where the density is $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ and $\lambda > 0$ (and $f(x) = 0$ for $x < 0$). Our data has n = 1 and x = 0. This means our likelihood is simply $L(\lambda) = \lambda e^{-\lambda x}$ and then plugging in x = 0, we

get $L(\lambda) = \lambda e^{-\lambda(0)} = \lambda$.

Recall that the sample mean here is 0, and the failure rate is the reciprocal of the mean. Thus, for this Proschan style heuristic, allowing the failure rate to be infinity we have $\hat{\lambda} = +\infty$. This yields $L(\hat{\lambda}) = \infty$ and therefore, $-2\log L(\hat{\lambda}) = -\infty$. Then, AIC $= -\infty + 2(1) = -\infty$.

Thus, a model with infinite likelihood will obtain a score of negative infinity and prove to be the "best" underlying model (or at least tied for "best" model) for this given data set. Professor Frank Proschan would use such extreme heuristics to demonstrate the essentials of important concepts and methods in reliability and elsewhere.

In comparison to AIC, Bozdogan's ICOMP includes first the same lack of fit term. However, in contrast, the penalty term is substantially different. This penalty term takes into account the asymptotic properties of the maximum likelihood estimators as well as the "complexity" of the inverse Fisher information matrix, $\mathcal{F}^{-1}$, of the proposed model.

Basically, rather than twice the number of parameters in the model, ICOMP takes on a penalty term that is viewed as the "degree of interdependence among the components of the model" and has the goal of providing a "more judicious penalty term than AIC and other AIC-type criteria, since counting and penalizing the number of parameters in the model is necessary but by no means sufficient" according to Bozdogan (2000).

Since the complexity term takes into account the interdependence of the parameters in the model, ICOMP can only be used for models with two or more

parameters. As with AIC, the model with the lowest ICOMP score is considered the best among all competing models.

We now present the generalized formula for ICOMP, which is as follows:

$$\text{ICOMP}(\mathcal{F}^{-1}) = -2\log L(\hat{\theta}) + 2C_1(\hat{\mathcal{F}}^{-1}(\hat{\theta})) \tag{4.2}$$

where $C_1(\hat{\mathcal{F}}^{-1}(\hat{\theta}))$ is a measure of complexity of the inverse Fisher information matrix and is given by:

$$C_1(\hat{\mathcal{F}}^{-1}(\hat{\theta})) = \frac{r}{2}\log[\frac{trace(\hat{\mathcal{F}}^{-1}(\hat{\theta}))}{r}] - \frac{1}{2}\log\left|\hat{\mathcal{F}}^{-1}(\hat{\theta})\right| \tag{4.3}$$

where r is the rank of the inverse Fisher information matrix and $|\bullet|$ denotes the determinant. For those interested in the details of the theory underlying AIC and ICOMP, the reader should turn to Bozdogan (1987, 1988, and 1996) and Bozdogan and Haughton (1998), among others.

We now focus on the use of AIC and ICOMP in association with probability plotting for the purposes of choosing the best plot that represents the data. Chapter 6 of Bozdogan (2001) provides extensive information on quantile modeling and how to incorporate AIC and ICOMP into this graphical approach. Recall, first, that in probability plotting, the data are ordered and then plotted against the theoretical order statistics for a desired distribution. In order to make use of the information criterion previously discussed, it is necessary to fit a regression model through the plotted points. To do this, we make use of a first order approximation of the form:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, \ldots n \tag{4.4}$$

where y_i represents the ordered data, x_i represents the theoretical quantiles, and ε_i

corresponds to the error associated with using the first order approximation. We must also assume that the errors are approximately normally distributed with constant error variance in order to carry out least squares regression. That is, $\varepsilon_i \sim N(0,\sigma^2)$.

It is without question that another regression model other than that shown in (4.4) above may be considered for use here. Furthermore, a more complex model that does not assume a constant error variance can certainly be utilized and may prove to be more appropriate. In particular, further research in this area may be conducted later. However, given the usefulness of this methodology for the practitioner, we choose a rough and quick approximation in the spirit of Frank Wilcoxon in order to aid in the ease of carrying out the use of information criteria in quantile modeling. We employ the derived formulas of AIC and ICOMP that will be useful for this application as follows:

$$\text{AIC} = n\ln(2\pi) + n\ln(\hat{\sigma}^2) + n + 2(2) \tag{4.5}$$

where $k = 2$ since we are estimating β_0 and β_1 for the regression model. Some may argue that $k = 3$ since we also include the parameter σ^2 in the likelihood equation. However, letting $k = 2$ is the traditional approach in regression analysis, especially in the construction of confidence intervals. See our resident expert, Professor Bozdogan, and his papers cited in this chapter for helpful comments on this and other insights below.

For more on regression analysis in general, see Neter, Kutner, Nachtsheim and Wasserman (1996) and Montgomery, Peck and Vining (2001). These, among others, provide a thorough reference and would be helpful for the practitioner and those interested in learning more about the fundamentals and applications of linear regression models.

The derived formula for ICOMP is:

$$\text{ICOMP} = n\ln(2\pi) + n\ln(\hat{\sigma}^2) + n + 2C_1(\hat{\mathfrak{F}}^{-1}(\beta_0, \beta_1, \sigma^2)) \tag{4.6}$$

where

$$\hat{\mathfrak{F}}^{-1}(\beta_0, \beta_1, \sigma^2) = \begin{pmatrix} \dfrac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2} & 0 \\ 0 & \dfrac{2\hat{\sigma}^4}{n} \end{pmatrix}. \tag{4.7}$$

Without question, one may expect that since the inverse Fisher information matrix shown in (4.7) takes into account three parameters, then the form of $\hat{\mathfrak{F}}^{-1}(\beta_0, \beta_1, \sigma^2)$ would be a 3x3 matrix. However, recall that we estimate $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$ where $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ when we allow the intercept term, β_0, to not equal zero. Note that $\hat{\varepsilon}_i = y_i - \hat{\beta}_1 x_i$ when the intercept term is set equal to zero, i.e., $\beta_0 = 0$. Therefore, whether we are dealing with a linear model with an intercept term or not (i.e. $\beta_0 = 0$ or not) will have no affect on the form of the inverse Fisher information matrix since the intercept term is included in the estimation of $\hat{\sigma}^2$.

This is important since in some applications (e.g., engineering and industrial applications) the intercept term does need to be constrained to zero. Therefore, we simply fit the simple regression model that is the most appropriate and calculate the estimated error variance using the residuals of the fitted model. Then, substitute the calculated estimate of the error variance into the inverse Fisher information matrix given in (4.7). For more on the derivation of the formulas for AIC, ICOMP, and the inverse

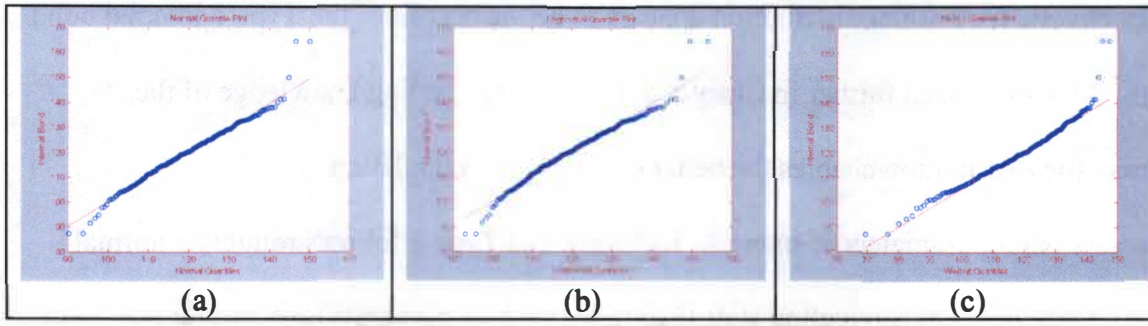
Fisher information matrix shown in formulas (4.5), (4.6), and (4.7) respectively, see the chapter on the theory of linear models in Bozdogan (2001). This chapter further illustrates the necessary technical details and shows how (4.7) is the same for both a model with or without an intercept term.

Let us next move to show how this can be used in application. In particular, we return to our data on the IB of MDF. Types 1, 2, 3, and 5 will be used and the best parametric distribution will be determined by scoring AIC and ICOMP.

4.3 USING AIC AND ICOMP WITH PROBABILITY PLOTS TO DETERMINE THE PARAMETRIC DISTRIBUTION OF THE INTERNAL BOND OF MEDIUM DENSITY FIBERBOARD

Using the helpful MATLAB programming language, a routine was constructed to calculate the quantiles, plot them, and score AIC and ICOMP for each of the normal, lognormal, and Weibull distributions. This was done for each of Type 1, 2, 3, and 5 MDF product types as an illustration and for useful comparison. Recall that product Types 1 and 3 have the same density of 46 lbs/ft^3 and a different thickness of 0.625 and 0.750 inches, respectively. Likewise, product Types 2 and 5 have the same density of 48 lbs/ft^3 and a different thickness of 0.750 and 0.625 inches, respectively. Recall Table 3.1.

We begin with Type 1 MDF. Figure 4.4 shows the three probability plots for the normal, lognormal, and Weibull distribution for Type 1 as produced by MATLAB. A close look at these plots clearly reveals their subjectivity. Two practitioners may not be able to agree on the best fit. Further, Table 4.1 shows the AIC and ICOMP scores for each plot. Based on the minimum values of AIC and ICOMP of 1428.6 and 1439.5,



(a) (b) (c)

Figure 4.4. Type 1 probability plots by MATLAB. (a) normal, (b) lognormal, and (c) Weibull probability plots.

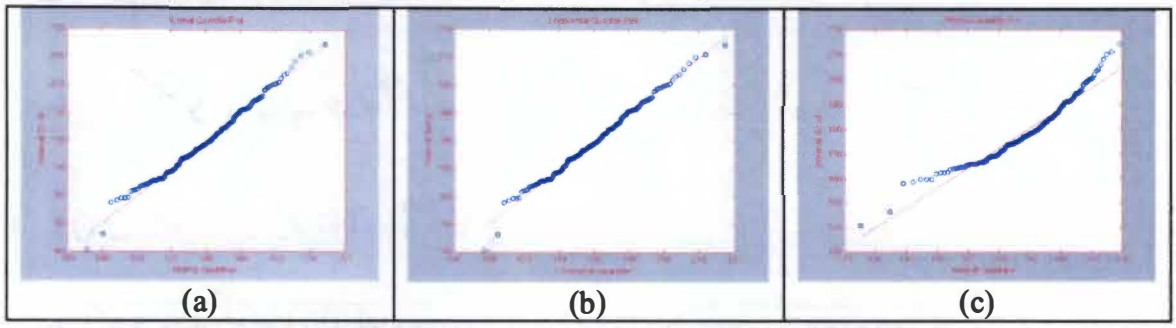
Table 4.1. AIC and ICOMP for Type 1 probability plots.

Distribution	AIC	ICOMP
Normal	**1428.6**	**1439.5**
Lognormal	1487.3	1498.1
Weibull	1678.1	1687.5

respectively, the normal distribution appears to be the best fit of the Type 1 internal bond data. As will be seen further in Chapter 5 of this thesis, having knowledge of the parametric distribution enables the better estimation of population characteristics/parameters of interest. Knowing that Type 1 follows roughly a normal distribution helps in conducting statistical tests (such as the analysis of variance (ANOVA)) where the assumption of normality is required.

Figure 4.5 shows the probability plots for Type 2 MDF as produced in MATLAB and Table 4.2 gives the AIC and ICOMP scores for each plot. Given, the minimum values of AIC and ICOMP of 571.12 and 584.11, respectively, choose the lognormal distribution as the best fit for the internal bond of Type 2 MDF. Recall that if $T \sim Lognormal(\mu, \sigma)$ then $Y = \log(T) \sim Normal(\mu, \sigma)$. In the case of the lognormal, we define the parameter μ as the mean of the logarithm of T and σ as the standard deviation of the logarithm of T. That is, the lognormal parameters are the mean and standard deviation of the transformed data. This distribution appears commonly in reliability data and falls into the location-scale family of distributions. For more information on the lognormal distribution, such as formulas for the expected value, variance, and quantiles, see Meeker and Escobar (1998).

Figure 4.6 and Table 4.3 give the probability plots and the AIC and ICOMP scores for Type 3 MDF, while Figure 4.7 and Table 4.4 show the probability plots and AIC and ICOMP scores for Type 5 MDF. Although the Type 3 AIC and ICOMP scores for the normal and lognormal were extremely close, the minimum value "wins," and thus, the normal distribution is chosen as the best fit for the internal bond of Type 3 MDF. For Type 5, we also will find that the normal distribution proves to be the best fit of internal

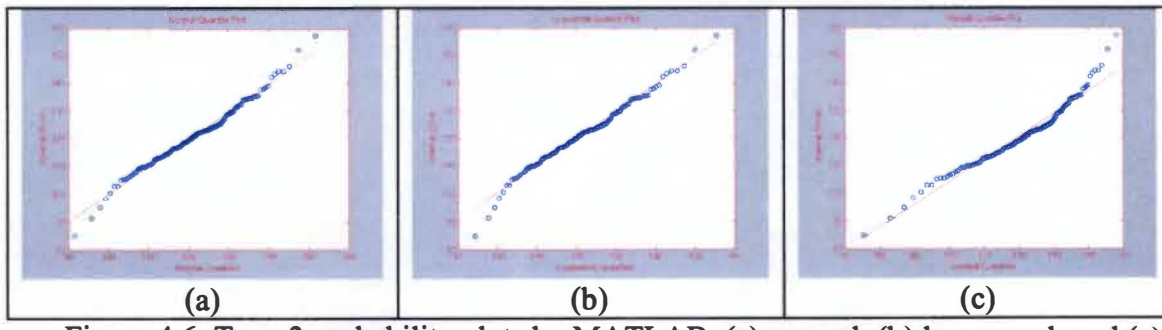


(a) (b) (c)

Figure 4.5. Type 2 probability plots by MATLAB. (a) normal, (b) lognormal, and (c) Weibull probability plots.

Table 4.2. AIC and ICOMP for Type 2 probability plots.

Distribution	AIC	ICOMP
Normal	614.47	627.24
Lognormal	**571.12**	**584.11**
Weibull	933.96	944.57

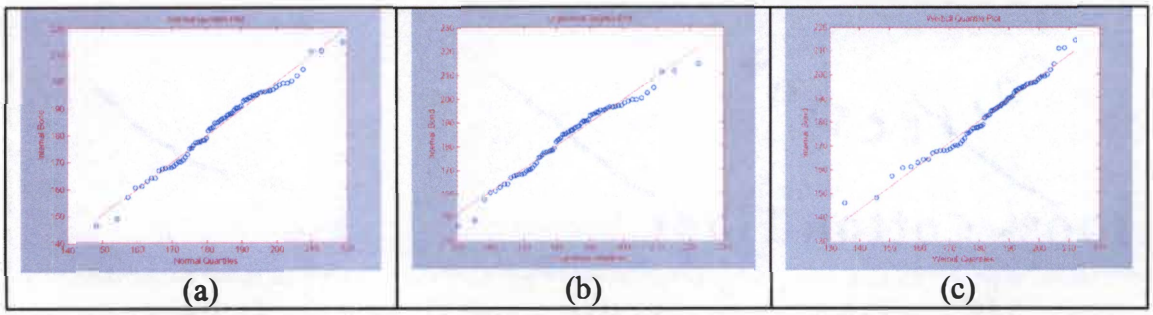


(a) (b) (c)

Figure 4.6. Type 3 probability plots by MATLAB. (a) normal, (b) lognormal, and (c) Weibull probability plots.

Table 4.3. AIC and ICOMP for Type 3 probability plots.

Distribution	AIC	ICOMP
Normal	**542.36**	**553.11**
Lognormal	543.14	553.84
Weibull	711.85	721.05



(a) (b) (c)

Figure 4.7. Type 5 probability plots by MATLAB. (a) normal, (b) lognormal, and (c) Weibull probability plots.

Table 4.4. AIC and ICOMP for Type 5 probability plots.

Distribution	AIC	ICOMP
Normal	**305.54**	**315.14**
Lognormal	342.47	351.57
Weibull	321.08	330.24

bond. There was no conclusive evidence to suggest that the Weibull distribution was the best underlying distribution for the top four MDF product types.

This might be surprising since the Weibull is often the first choice when studying the strengths of materials. Indeed, the researcher Dr. Weibull himself first analyzed strengths of different materials, ranging from cotton to metal. From his data sets, he found the primary available distribution of the normal did not fit his examples well in the 1930's. The alternative parametric model he originally proposed is what we now call the "three parameter" Weibull. Compare Weibull (1939 and 1951).

The internal bond of MDF is an example of how important the information criteria are to find the parametric distribution for a data set. After all, our data surprises us and our intuition may not always be confirmed. These criteria let the data speak objectively for the "best" model.

4.4 SUMMARY

It cannot be reiterated enough that probability plotting is an extremely useful way to aid in the determination of the parametric distribution of a particular data set. In many cases, when comparing plots of different candidate distributions, little ambiguity is present and the choice of distribution is not difficult. However, this method is subjective and when ambiguity among candidate distribution probability plots is present, two or more variant conclusions may be reached.

Akaike's Information Criterion (AIC) and Bozdogan's Information Complexity Criterion (ICOMP) rescues practitioners from such subjectivity. These two forms of information criteria, among others, have a lack of fit term (minus twice the log

likelihood) plus a penalty term that accounts for the number of parameters in the respective model. Building on the fundamental ideas behind probability plotting, AIC and ICOMP make it much more objective by creating a numeric score for each plot. The plot with the lowest score is considered to be the best fit of the data set under consideration. This is great to explain to practitioners, who may not have much statistical training.

When applied to the different product types of MDF, it was discovered that the normal distribution is technically the best fit for Types 1, 3, and 5, while the lognormal distribution is the best fit for Type 2. Of course for Type 3, the lognormal was extremely second which might also be considered in that specific case. Thus, even though the Weibull distribution might be an intuitive choice when studying the strengths of materials, the actual data analysis of the internal bond of MDF does not support this first intuition. As stated in Chapter 3, data has a way of producing unexpected results in the light of intuition, exploration and theory.

Note that future work in the area of information criteria might include using another model different than this helpful first order approximation assuming normal errors with constant variance. Also, information criteria are certainly not limited to applications in probability plotting. Other considerations include their use in regression analysis, statistical process control, etc.

Chapter 5

Applying Bootstrap Techniques for Estimating Percentiles of the Internal Bond of Medium Density Fiberboard

5.1 INTRODUCTION AND MOTIVATION

In reliability studies, it is generally of high interest to estimate percentiles. In particular, interest usually lies in the estimation of the lower percentiles. These lower numbers are helpful for warranty analysis, understanding early failures during normal usage, plus improving the specification limits. Meeker and Escobar (1998) observe that the "traditional parameters of a statistical model (e.g., mean and standard deviation) are not of primary interest. Instead, design engineers, reliability engineers, managers, and customers are interested in specific measures of product reliability or particular characteristics of a failure-time distribution (e.g., failure probabilities, quantiles of the life distribution, failure rates)." See, also, Meeker and Escobar (2004).

Nelson (1990) further mentions that with "life data work, one often wants to know low percentiles such as the 1% and 10% points, which correspond to early failure. The 50% point is called the median and is commonly used as the 'typical' life." The first and third quartiles are also useful in studying the life of a product. We note that for some lower percentiles, samples sizes need to be adequately large. If samples are small, the lower percentiles should be avoided and the quartiles should be used instead. Recall the median is the second quartile. Compare the comments from Polansky (1999) warning to have sufficient sample size for the lower percentiles.

To be able to say that improvements have been made, we must be able to measure

reliability expressed in percentiles that allow for statistical variation. We need to make comparisons of these reliability measures between products and within products before and after process improvement interventions. Knowing when to trust confidence intervals and when not to trust them are crucial for managers and users of MDF to make successful decisions.

Chapter 5 is devoted to the estimation of percentiles of IB and their respective confidence intervals. In particular, the bootstrap will be presented as a useful method for obtaining the aforementioned estimates and confidence intervals. We next provide the reader with some useful background information on percentiles, its consistency, and asymptotic distribution.

Serfling (1980) defines the 100 pth percentile or pth quantile as $t_p = \inf\{t : F(t) \geq p\}$ where F represents the distribution function. That is, the pth quantile is the greatest lower bound of the set of all values, t, such that $F(t)$ is greater than or equal to a specified value of p where $0 \leq p \leq 1$. In practice, we take the infimum of the set since it is possible for the distribution function, F, to yield a set where the minimum does not exist, but the "inf" does. For example, the open interval (0, 1) has an inf of 0, but the min does not exist.

For a sample of n observations, $\{t_1, t_2, \ldots t_n\}$ on F, the sample pth quantile, denoted by $\hat{t}_p$ is "defined as the pth quantile of the sample distribution function", or as $F^{-1}(p)$. Furthermore, it has been shown that $\hat{t}_p$ is a consistent estimator of t_p. That is, as the sample size increases, the estimate of the quantile gets closer and closer to the true value. This asymptotic property is "such a fundamental property that the worth of an

inconsistent estimator should be questioned (or at least vigorously investigated)" according to Casella and Berger (2002). Thus, consistency is certainly a desirable property for an estimator. Stated statistically, for any small $\varepsilon > 0$, it follows that $P(\sup|\hat{t}_p - t_p| > \varepsilon) \to 0$ as the sample size n approaches infinity. Equivalently, this can be written as $P(\sup|\hat{t}_p - t_p| \le \varepsilon) \to 1$ as n approaches infinity. In particular, it should be noted that this convergence rate is exponential.

Serfling (1980) also thoroughly examines the asymptotic distribution of the sample quantile. In particular, under mild requirements (i.e. smoothness of the distribution function), the sample quantiles are asymptotically normal. We state the following theorem and corollary from Serfling (1980) without proof, which for more extensive details, see his book.

Theorem: Assume that the left and right hand derivatives of F exist at t_p and that F is continuous at t_p. Then, if the left hand derivative, denoted by $F'(t_p-)$, is greater than 0, then for $t < 0$,

$$\lim_{n\to\infty} P\left(\frac{\sqrt{n}(\hat{t}_p - t_p)}{\sqrt{p(1-p)}/F'(t_p-)} \le t\right) = \Phi(t). \tag{5.1}$$

Furthermore, if the right hand derivative, denoted by $F'(t_p+)$, is greater than 0, then for $t > 0$,

$$\lim_{n\to\infty} P\left(\frac{\sqrt{n}(\hat{t}_p - t_p)}{\sqrt{p(1-p)}/F'(t_p+)} \le t\right) = \Phi(t). \tag{5.2}$$

Finally, for $t = 0$, (5.1) and (5.2) can be simplified as,

$$\lim_{n\to\infty} P(\sqrt{n}(\hat{t}_p - t_p) \le 0) = \Phi(0) = 0.5. \tag{5.3}$$

Corollary: Assuming F is differentiable at t_p and $F'(t_p) > 0$, then

$$\hat{t}_p \text{ is } AN(t_p, \frac{p(1-p)}{F'(t_p)^2 n}). \tag{5.4}$$

where AN stands for "asymptotically normal".

For further details and an extensive proof of the above theorem and corollary, see Serfling (1980). This is a useful result since by possessing asymptotic normality; we can construct asymptotic normal confidence intervals for the pth quantile of a distribution. Chapter 8 of Meeker and Escobar (1998) provides very helpful information regarding the construction of such intervals for the location-scale distributions used commonly in reliability data analysis. In particular, a normal approximate confidence interval for t_p is given by:

$$\hat{t}_p \pm z_{1-\alpha/2} \widehat{se}_{\hat{t}_p} \tag{5.5}$$

where $\widehat{se}_{\hat{t}_p}$ is the standard error of the estimate and is given by:

$$\widehat{se}_{\hat{t}_p} = \sqrt{\widehat{Var}(\hat{t}_p)} = \hat{t}_p \{\widehat{Var}(\hat{\mu}) + 2\Phi^{-1}(p)\widehat{Cov}(\hat{\mu},\hat{\sigma}) + [\Phi^{-1}(p)]^2 \widehat{Var}(\hat{\sigma})\}^{1/2} \tag{5.6}$$

which is derived using the delta method and Φ^{-1} represents the inverse of the cumulative standard normal distribution. $\widehat{Var}(\hat{\mu})$, $\widehat{Var}(\hat{\sigma})$, and $\widehat{Cov}(\hat{\mu},\hat{\sigma})$ are obtained from the variance-covariance matrix or inverse Fisher information matrix, $\mathcal{F}^{-1}$. To compute these intervals by hand can, without question, be very tedious and time consuming. Fortunately, statistical software packages such as SAS have the capabilities (i.e. PROC RELIABILITY) to produce these confidence intervals for desired quantiles.

When the sample size is sufficiently large, the asymptotic normal intervals

provide very good approximations. Even though these intervals are approximations, they are usually good enough for practice. However, for small samples, these intervals may not provide accurate approximations. It is in this case that another method, whether it be parametric or nonparametric, is necessary to obtain better confidence intervals for desired quantiles. Bootstrap methods provide one possibility for better estimation given reasonable sample sizes.

Bootstrapping is a computer intensive statistical method where the basic idea is to simulate the sampling process a specified (usually large) number of times and obtain an empirical bootstrap distribution for a desired population parameter. This empirical bootstrap distribution is then used to acquire characteristics about the population parameter. These include, but are not limited to, the standard error, an estimate of bias, and confidence intervals. Some bootstrap methods are nonparametric and therefore do not require any parametric assumptions regarding the underlying distribution of a particular data set. Other methods using the bootstrap are parametric. These methods will be discussed and compared in detail later on in this chapter.

According to Chernick (1999), the "bootstrap is a form of a larger class of methods that resample from the original data set and thus are called resampling procedures. Some resampling procedures similar to the bootstrap go back a long way.... However, it was [Bradley] Efron who unified ideas and connected the simple nonparametric bootstrap, which 'resamples the data with replacement' with earlier accepted statistical tools for estimating standard errors such as the jackknife and the delta method."

Boos (2003) describes the bootstrap as a technique that "has made a fundamental

impact on how we carry out statistical inference in problems without analytic solutions." Davison and Hinkley (1997) tell us that the bootstrap is called so since "to use the data to generate more data seems analogous to a trick used by the fictional Baron Munchausen, who when he found himself at the bottom of a lake got out by pulling himself up by his bootstraps." They further assert the necessity of careful reasoning and investigation of the problem at hand despite the usefulness of bootstrap methods. It is contended that "unless certain basic ideas are understood, it is all too easy to produce a solution to the wrong problem, or a bad solution to the right one."

Efron and Tibshirani (1993) is an excellent starting point and a way to get acquainted with the fundamental concepts and applications of the bootstrap. Much of their work is written without rigorous technical details in order to focus on ideas rather than justification. Those details can be found in some of their later chapters as well as other works.

DiCiccio and Efron (1996) is devoted to the construction of bootstrap confidence intervals. Here, different methods are presented as well as the theoretical underpinnings.

We adopt next the notation of Martinez and Martinez (2002), which is also similar to Efron and Tibshirani (1993). In general, the basic nonparametric bootstrap procedure (Efron's bootstrap) can be summarized as follows. For a given data set, $\mathbf{x} = (x_1, x_2, ..., x_n)$ of size n, we estimate a population parameter, say θ, by $\hat{\theta}$. We then sample with replacement from the original data set to obtain a bootstrap sample of size n denoted by $\mathbf{x}^{*b} = (x_1^{*b}, x_2^{*b}, ...x_n^{*b})$. This resampling with replacement is done a large number of times and for each bootstrap sample we calculate the estimate of θ, which is denoted by $\hat{\theta}^{*b}$

where b stands for the bth bootstrap estimate of a total of B bootstrap replications. The empirical bootstrap distribution of $\hat{\theta}^*$, is defined and used as an estimate to the true distribution of $\hat{\theta}$.

The fundamental idea behind the bootstrap is that the empirical bootstrap distribution provides an approximation to the theoretical sampling distribution of the desired population parameter as the sample size increases. In particular, as n approaches infinity, the bootstrap distribution becomes more normal for most cases. The bootstrap has a wide range of applications and has enjoyed more growth in use in recent years. However, as with any statistical method, the bootstrap does have its limitations.

Beran (2003) emphasizes that "Success of the bootstrap...is not universal. Modifications to Efron's definition of the bootstrap are needed to make the idea work for modern procedures that are not classically regular."

As also described in Chapter 2 of this thesis, Meeker and Escobar (1998) contend that the "justification for the bootstrap is based on large-sample theory. Even with large samples, however, there can be difficulties in the tails of the sample. For the nonparametric bootstrap, there will be a separate bootstrap distribution at each time for which there were one or more failures in the original sample." This would not pose a problem outside the tails of the original data where the bootstrap distribution will be approximately continuous. However, in the extreme tails of the original data, there may be only a small number of failures or outcomes. In this case, the bootstrap distribution may be anything but continuous. As can be seen by the examples presented, when the extreme tails are of interest (as is often the case in reliability studies), the fully

nonparametric bootstrap methods may not prove to be as useful. Rather the standard bootstrap methods have a place when estimating parameters such as the quartiles (25^{th}, 50^{th} or 75^{th} percentiles).

We will see the above limitation of the bootstrap when applied to the estimation of lower percentiles for the internal bond of medium density fiberboard. The sample size also plays a key role in the accuracy of bootstrap methods. Further limitations are described, for example, in Chernick (1999) and Ghosh, Parr et al. (1984).

In section 5.2 coming, different methods for constructing bootstrap confidence intervals will be introduced. These include the standard normal, bootstrap-t, percentile, and bias-corrected percentile intervals. However, for each method of creating bootstrap confidence intervals, there is also more than one way to create bootstrap samples. In particular we discuss the completely nonparametric bootstrap, the completely parametric bootstrap, and finally a "nonparametric" bootstrap for parametric inference as described in Meeker and Escobar (1998). Each of these methods for creating bootstrap samples will be presented along with the above methods for producing confidence intervals.

Section 5.3 will apply what was presented in section 5.2 to the MDF data for estimating the lower percentiles of the internal bond. The asymptotic normal intervals will be compared with the bootstrap confidence intervals. Furthermore, the bootstrap confidence intervals will be compared among themselves with respect to sampling and construction method. Section 5.4 provides a summary and concluding comments with loose recommendations for which situations dictate which bootstrap method to use. Also, possible future work will be presented here.

5.2 A BRIEF INTRODUCTION TO BOOTSTRAP SAMPLING METHODS AND BOOTSTRAP CONFIDENCE INTERVALS

5.2.1 Methods of Bootstrap Sampling

Already in section 5.1, we introduced the completely nonparametric bootstrap or Efron's bootstrap. That is, no assumptions are made about the underlying parametric distribution of a data set of size n. The desired population parameter is estimated nonparametrically from the initial data. Then, sampling is done with replacement (usually a large number of times). For each sample of size n obtained, we nonparametrically estimate the population parameter. These estimates of the desired parameter are used to form the empirical bootstrap distribution that will be useful for inference. This empirical distribution is a discrete distribution that assigns a probability of 1/n to each value of x. This method of sampling is helpful since it has the advantage of no distributional assumptions. When it is not possible or feasible to make such an assumption, the completely nonparametric bootstrap sampling method should be employed. The next two methods of sampling below do require a parametric distributional assumption.

The completely parametric bootstrap is described briefly in Efron and Tibshirani (1993), as well as Chernick (1999) and Meeker and Escobar (1998). A parametric distribution is assumed and the initial data of size n is utilized only to obtain maximum likelihood estimates of the model parameters. From there, one must simulate a specified number of samples of size n from the parametric distribution. The population parameter is then estimated parametrically from each of the simulated samples, which then helps us create the desired bootstrap distribution necessary for inference.

Chernick (1999) interestingly notes that the "parametric form of bootstrapping is equivalent to maximum likelihood. However, in parametric problems, the existing theory on maximum likelihood estimation is adequate and the bootstrap adds little or nothing to the theory. Consequently, it is uncommon to see the parametric bootstrap used in real problems." However, Efron and Tibshirani (1993) argue that when the fully parametric bootstrap is used, it "provides more accurate answers than textbook formulas, and can provide answers in problems where no textbook formulae exist...The parametric bootstrap is useful in problems where some knowledge about the form of the underlying population is available, and for comparison to nonparametric analyses."

Meeker and Escobar (1998) points out that the parametric bootstrap has a disadvantage in reliability data problems. That is, the complete censoring process must be specified given that we are simulating data. This may seem to be unproblematic in simple examples where such specification is easy. However, this can be "more difficult for complicated systematic or random censoring. Often the needed information may be unknown." An alternative to this method requiring parametric assumptions is described next.

Meeker and Escobar (1998) describe and illustrate applications of a "nonparametric" bootstrap sampling method for parametric inference, which we denote as NBSP for nonparametric bootstrap sampling for parametric models. They contend, "This method is simple to use and generally, with moderate to large samples, provides results that are close to the fully parametric approach." This sampling scheme does require parametric assumptions. However, rather than simulating data, we sample with replacement from the original data. For each sample of size n, maximum likelihood

estimates are obtained based on the assumed parametric model. Then, the MLEs are used to parametrically estimate the population parameter of interest.

The distribution of estimates allows us to conduct the desired inferences. See Chapter 9 of Meeker and Escobar (1998) for more details on this method of bootstrap sampling. We move now to see how the aforementioned sampling schemes and the resulting empirical distributions allow us to construct bootstrap confidence intervals for population parameters.

5.2.2. Bootstrap Confidence Intervals

Different algorithms/methods are available for constructing bootstrap confidence intervals for population parameters. These include, but are certainly not limited to, the bootstrap standard confidence interval, bootstrap-t confidence interval, bootstrap percentile interval, and bias-corrected bootstrap percentile interval. We describe these briefly here and omit much of the theoretical details here. For those interested in the theoretical underpinnings and additional topics, many good books and articles exist. See, among others, Efron and Tibshirani (1993), DiCiccio and Efron (1996), Davison and Hinkley (1997), and Polansky (1999).

The bootstrap standard confidence interval is by far the easiest to implement. Efron and Tibshirani (1993) and others use the phrase "bootstrap standard confidence interval" while is it also known as the normal approximation bootstrap confidence interval. These intervals are based on the following asymptotic result:

$$Z = \frac{\hat{\theta} - \theta}{\widehat{se}_{\hat{\theta}}} \sim N(0,1). \tag{5.7}$$

Then, the standard confidence interval is given by:

$$[\hat{\theta} - z^{(\alpha/2)}\widehat{se}_{\hat{\theta}} , \hat{\theta} + z^{(1-\alpha/2)}\widehat{se}_{\hat{\theta}}] \tag{5.8}$$

where $\widehat{se}_{\hat{\theta}}$ simply the standard deviation of the bootstrap is estimates of θ and $z^{(\alpha/2)}$ is the $\alpha/2th$ quantile of the standard normal distribution. That is, for example, $z^{(0.025)} =$ -1.96.

Although this method is easy to use, (5.7) "is only an approximation in most problems, and the standard interval is only an approximate confidence interval, though a very useful one in an enormous variety of situations" according to Efron and Tibshirani (1993). This interval can be used when the asymptotic normality is valid.

Another useful method is the bootstrap-t confidence interval. For each of B (some large number of choice for B; usually larger than 1000) bootstrap samples, we compute:

$$Z^{*b} = \frac{\hat{\theta}^{*b} - \hat{\theta}}{\widehat{se}_{\hat{\theta}^{*b}}} \tag{5.9}$$

where $\widehat{se}_{\hat{\theta}^{*b}}$ is the standard error of $\hat{\theta}^{*b}$ for a particular bootstrap sample. The difficulty arises in the computation of this estimate standard error. In many situations, a nice closed formula does not exist. To remedy this, a possible solution is to bootstrap each bootstrap sample and then take $\widehat{se}_{\hat{\theta}^{*b}}$ to be the standard deviation of the "bootstrapped" bootstrap sample. Basically, one performs a double bootstrap to obtain the desired estimate of the standard error.

The only problem with this is the amount of computer power required to perform

such a large number of bootstrap replications. For example, if we want to obtain 2000 bootstrap samples and bootstrap each of those 50 times to obtain the sample estimate's standard error, then this requires 100,000 iterations.

After obtaining the B values of Z^{*b}, order them and calculate the $\alpha/2th$ and $1-(\alpha/2)th$ quantiles of the distribution of Z^{*b} values which will be denoted by $\hat{t}^{(\alpha/2)}$ and $\hat{t}^{(1-\alpha/2)}$ respectively. This is finding the appropriate percentiles of the sampling distribution, which is to be distinguished clearly from the percentiles of the original population data. At this point, the bootstrap-t confidence intervals can be computed. Then, a 100(1- α) % bootstrap-t confidence interval is given by:

$$[\hat{\theta}-\hat{t}^{(1-\alpha/2)}\widehat{se}_{\hat{\theta}}\ ,\ \hat{\theta}-\hat{t}^{(\alpha/2)}\widehat{se}_{\hat{\theta}}]. \qquad (5.10)$$

The bootstrap-t confidence intervals are second-order accurate (error goes to zero at a rate of 1/n), which makes them a popular choice in practice.

However, Efron and Tibshirani (1993) warn that the "bootstrap-t can give erratic results, and can be heavily influenced by a few outlying data points. The percentile based methods...are more reliable." Polansky (2000) "investigates two methods for stabilizing the endpoints of bootstrap-t intervals in the case of small samples. In those cases, this would be an approach for others to use. Two of the percentile-based methods will now be discussed.

Perhaps one of the most obvious ways to construct a confidence interval is to base it on the quantiles of the bootstrap distribution of estimates. Constructing a bootstrap confidence interval in this manner is known as the standard percentile method. Martinez and Martinez (2002) and Efron and Tibshirani (1993) maintain that "this technique has

the benefit of being more stable than the bootstrap-t, and it also enjoys better theoretical coverage properties." In particular, this method works well when a monotone transformation, $\phi = g(\theta)$ exists such that $\hat{\phi} = g(\hat{\theta})$ possesses an approximate normal distribution with mean ϕ and a standard deviation, τ , which is constant. After obtaining B bootstrap samples and estimating the desired population parameter, calculate the $\alpha/2th$ and 1-$(\alpha/2)th$ quantiles of the distribution of $\hat{\theta}^*$ denoted by $\hat{\theta}^{*(\alpha/2)}$ and $\hat{\theta}^{*(1-\alpha/2)}$ respectively. Then, a $100(1-\alpha)\%$ confidence interval for θ is given by:

$$[\hat{\theta}^{*(\alpha/2)}, \hat{\theta}^{*(1-\alpha/2)}] \qquad (5.11)$$

For example, in order to construct a 95% confidence interval, we simply calculate the 2.5^{th} and 97.5^{th} percentiles of the bootstrap distribution for the parameter of interest. It is generally recommended that the number of bootstrap replications be equal to or greater than 1000 for this method to produce accurate results.

Though the standard percentile method is easy to implement, Chernick (1999) points out that "the percentile method works if exactly 50% of the bootstrap distribution of $\hat{\theta}^*$ is less than $\hat{\theta}$" which may certainly not always be the situation and that "in the case of small samples, the percentile method does not work well." Furthermore, a two-sided $100(1-\alpha)\%$ confidence interval should have the probability of not covering the true value of a parameter, either above or below, of $\alpha/2$. The standard bootstrap percentile intervals are first order accurate (error goes to zero at a rate of $n^{-1/2}$) which means that the error in getting exactly the desired $\alpha/2$ probability is an order of magnitude greater than that of the bootstrap-t intervals which, one can recall, are second

order accurate. Fortunately, there are methods that help improve on the standard percentile method and one such will be shown next.

The final method for constructing bootstrap confidence interval that will be presented here is called the bias-corrected percentile interval. This method was introduced in Efron (1981) and discussed further in Efron (1987). The method is described there in greater length along with the needed theoretical details. The bias-corrected percentile method (or BC) works best when a monotone transformation, $\phi = g(\theta)$, exists so that $\hat{\phi} = g(\hat{\theta})$ is roughly normal with mean of $\phi - z_0\tau$ where z_0 is the bias correction constant and τ is the constant standard deviation of $\hat{\phi}$.

Assuming, again, that the aforementioned transformation exists, Efron (1987) shows that the transformation leads to the "obvious confidence interval $(\hat{\phi} + \tau z_0) \pm \tau z^{(\alpha)}$ for ϕ, which can then be converted back to a confidence interval for θ by the inverse transformation $\theta = g^{-1}(\phi)$. The advantage of the BC method is that all of this is done automatically from bootstrap calculations, without requiring the statistician to know the correct transformation g."

The bias correction constant is defined as the amount of difference between the median of the bootstrap distribution of estimates and the estimate, $\hat{\theta}$ from the original sample. That is, if we take the bias to be $bias = \hat{\theta} - \hat{\theta}^*_{0.5}$, then $\hat{\theta}^*_{0.5} = \hat{\theta} - bias$. This is explained further in Chernick (1999). Explicitly, we define the estimate of the bias correction constant, denoted by $\hat{z}_0$, simply as:

$$\hat{z}_0 = \Phi^{-1}(F^*(\hat{\theta})) \tag{5.12}$$

where Φ^{-1} represents the inverse cumulative normal distribution and F^* is the cumulative bootstrap distribution for the parameter of interest.

Alternatively, in other words, we can express (5.12) as the inverse cumulative normal distribution of the number of bootstrap estimates, $\hat{\theta}^{*b}$, that are less than the original sample estimate, $\hat{\theta}$, divided by the number of bootstrap replicates, B. That is, we rewrite (5.12) as:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#(\hat{\theta}^{*b} < \hat{\theta})}{B}\right). \tag{5.13}$$

Then, a $100(1-\alpha)\%$ confidence interval for θ is given by:

$$[\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)}] \tag{5.14}$$

where α_1 and α_2 are the new quantities on which to base the percentile confidence interval endpoints. Martinez and Martinez (2002) explain that "instead of basing the endpoints of the interval on the confidence level of $1-\alpha$, they are adjusted using information from the distribution of bootstrap replicates." These quantities are the lower and upper bias-corrected cut-off percentages and are defined as:

$$\alpha_1 = \Phi(2\hat{z}_0 + z^{(\alpha/2)}) \tag{5.15}$$

and

$$\alpha_2 = \Phi(2\hat{z}_0 + z^{(1-\alpha/2)}) \tag{5.16}$$

where Φ is the cumulative standard normal distribution and $z^{(\alpha/2)}$ is the $\alpha/2th$ quantile of the standard normal distribution. The bias-corrected percentile intervals have been found to be second-order accurate, which certainly improves on the standard percentile interval. The method does have the drawback of not being monotone in coverage. That

is, if we decrease the confidence level, we do not necessarily get a shorter interval than that obtained at a higher level of confidence. In general, the percentile and bias-corrected percentile methods give more conservative confidence intervals than the bootstrap-t.

In summary, Martinez and Martinez (2002) point out that the "bootstrap-t interval has good coverage probabilities, but does not perform well in practice. The bootstrap percentile interval is more dependable in most situations, but does not enjoy the good coverage property of the bootstrap-t interval."

Recall, rather, that the percentile interval possesses good *theoretical* coverage properties, which may not actually hold in practice. The bias-corrected percentile interval helps to remedy this by being both dependable and having good coverage properties. Efron (2003) further points out that even though the bootstrap-t and the bias-corrected intervals are second-order accurate, they are not widely used in application. Instead, researchers and even seasoned statisticians "seem all too happy with the standard intervals" which may certainly be due to its theoretical simplicity and ease in construction.

Other possibilities for bootstrap confidence intervals, which will not be described here, include the iterated or double bootstrap and the bias-corrected and accelerated (BC_a) percentile interval, among others. The iterated bootstrap requires the user to bootstrap the B_1 bootstrap samples, B_2 times. The price to pay here is an extremely large increase in iterations. For example, if $B_1=B_2=B$, then one must go through B^2 iterations, which can obviously take up a large amount of computing power. The bias-corrected and accelerated interval is built upon the bias-corrected interval described above. It requires the calculation of an acceleration constant when the standard deviation of the

transformation is not independent of the transformation. This constant, however, can prove to be difficult to determine and the reader is directed to Efron (1987) and Efron and Tibshirani (1993) for more on the BC_a method.

After this degree of introduction, we move next (in section 5.3) to demonstrate how these bootstrap sampling methods and confidence intervals can be applied for the purposes of estimating percentiles (especially lower percentiles) of the internal bond of medium density fiberboard. It should be emphasized here again that the estimation of lower percentiles is very important in reliability studies for strengths of failure. This helps manufacturers gauge process improvements, warranties, proportion of product falling out of spec, etc. Also, there are clear economic advantages to exploring these lower percentiles.

5.3 EXPLORING BOOTSTRAP METHODS AND CONFIDENCE INTERVALS FOR PERCENTILES OF THE INTERNAL BOND OF MDF

In this section, we review the methods of bootstrap sampling and for constructing confidence intervals in the context of estimating percentiles for the internal bond of medium density fiberboard. We also observe how they weigh against each other. For each method of sampling, the standard normal, bootstrap-t, standard percentile, and bias-corrected percentile intervals will be constructed and compared for the 1^{st}, 10^{th}, 25^{th}, and 50^{th} percentiles for MDF product Types 1 and 5. These two types were chosen to aid in the illustration of the benefits and limitations of the bootstrap. Type 1 MDF has n=396 observations while Type 5 MDF has n=74. We will thus be able to see how a smaller sample size compares to that of a sample that is sufficiently large to obtain relatively

accurate results. For each method of sampling, B=2000 bootstrap samples of the same size as the original sample were created. In many cases, but not always, this should be a sufficient number of bootstrap samples to create the confidence intervals. The asymptotic normal confidence intervals will also be provided in order to compare with the bootstrap results.

Furthermore, along with each method of bootstrap sampling, histograms of the empirical bootstrap distribution will be shown for each percentile. The fully nonparametric intervals will be shown first, followed respectively by the fully parametric intervals and the NBSP intervals described by Meeker and Escobar (1998). This is important for practitioners who may not have the luxury of developing or assuming certain parametric distributions due to their intense time pressures. Also, this protects them more from misspecified parametric models.

MATLAB was utilized as the program of choice for this author in order to construct these intervals. Certainly other software packages exist with capabilities of producing bootstrap confidence intervals. The SPLIDA add-on for S-PLUS developed by William Meeker plus Resampling Stats have such capabilities.

We begin, as before, with Type 1 MDF. One should recall from Chapter 4 that through the use of information criteria in conjunction with probability plotting, it was determined that the underlying parametric distribution for Type 1 MDF is better modeled by the normal, than Weibull or lognormal. This assumption will not be necessary for the fully nonparametric bootstrap intervals. However, it will be essential for constructing the fully parametric confidence intervals and the NBSP method described by Meeker and Escobar (1998) that involves nonparametric sampling for the purposes of parametric

inference. Again, this particular product type has a large sample size and this will help alleviate some of the limitations of the bootstrap based on sample size. Table 5.1 provides the 95% asymptotic normal confidence intervals for Type 1 MDF, while Table 5.2 shows the fully nonparametric 95% bootstrap confidence intervals. In the tables that follow, LCL stands for lower confidence limit, while UCL stands for upper confidence limit. The units for the point estimates and confidence limits are pounds per square inches (psi) as the reader will recall are the units for measuring internal bond. This is followed by Figure 5.1, which displays the nonparametric empirical bootstrap sampling distribution for each of the four quantiles.

An initial look at the bootstrap sampling distributions shown in Figure 5.1 shows that the bootstrap distribution becomes narrower and more peaked as the percentiles increase from 1 to 50, reflecting the standard errors being smaller as the numbers get larger. This is what your intuition would expect. It is advantageous to note that based on the histograms and given the relatively large sample size; the bootstrap distributions for Type 1 MDF roughly appear continuous rather than discrete (i.e. no holes are present in the histogram). Recall that this problem was described above and does occur frequently with small sample sizes.

The intervals for the 1st percentile of Type 1 MDF are rather wide. They are, in fact, wider than the asymptotic normal intervals. This, again, is to be expected given the limited amount of data in the extreme lower tail of the IB data. Users might consider not using them. This is a healthy warning of the dangers of using the bootstrap without thinking!

When we employ the fully nonparametric bootstrap, which samples with

Table 5.1. 95% Asymptotic normal confidence intervals for Type 1 MDF.

p	$\hat{t}_p$ = quantile	LCL	UCL
.01	97.2746	95.4023	99.1470
.10	107.5921	106.2795	108.9046
.25	113.5868	112.5093	114.6644
.50	120.2475	119.2749	121.2201

Table 5.2. Fully nonparametric 95% bootstrap confidence intervals for Type 1 MDF.

p	$\hat{t}_p$ = quantile	Interval Type	LCL	UCL
.01	94.4307	Standard	87.2652	100.4228
		Bootstrap-t	88.4673	99.1082
		Percentile	87.2000	100.6300
		Bias-Corrected	87.2000	99.2800
.10	107.6854	Standard	105.6784	109.5216
		Bootstrap-t	106.1158	108.7768
		Percentile	105.9300	109.7600
		Bias-Corrected	105.9008	109.4300
.25	114.3420	Standard	113.4248	115.3752
		Bootstrap-t	113.7840	115.0314
		Percentile	113.4000	115.4000
		Bias-Corrected	112.8000	115.1500
.50	120.2993	Standard	119.1490	121.2510
		Bootstrap-t	119.2056	120.7486
		Percentile	119.4000	121.6500
		Bias-Corrected	119.3000	121.6000

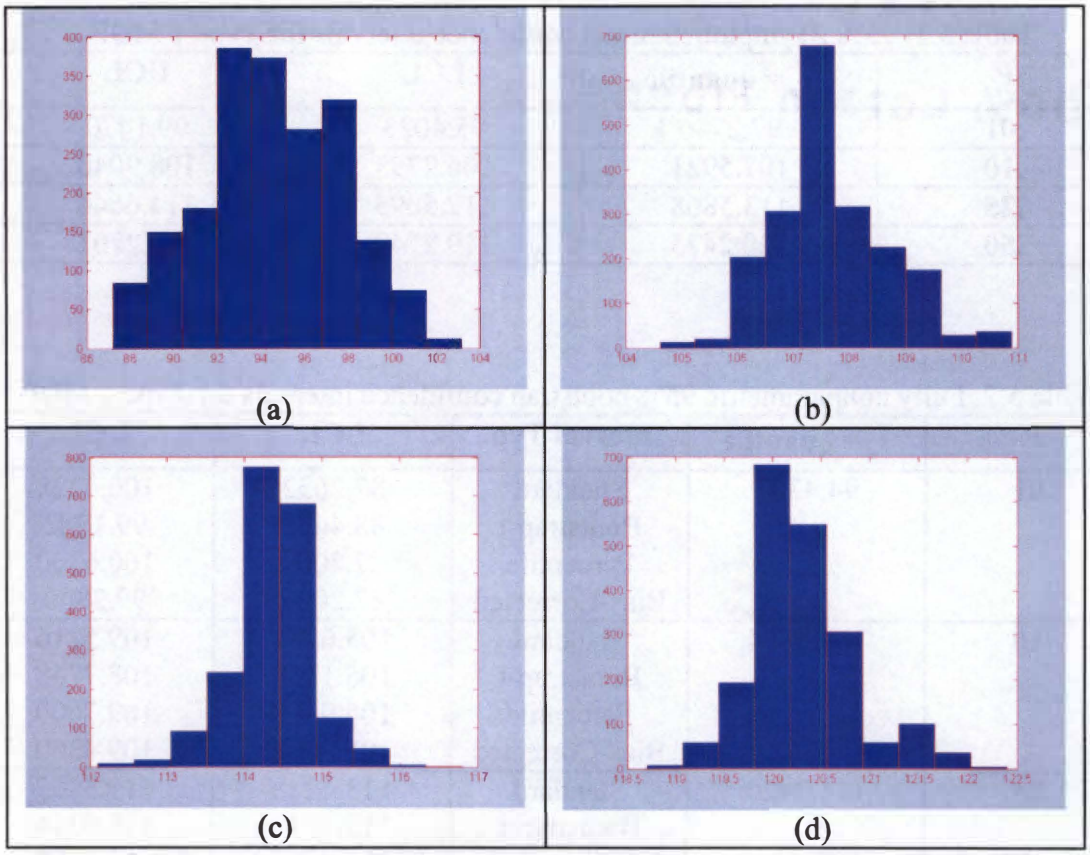


Figure 5.1. Sampling distribution of percentiles for Type 1 MDF under the fully nonparametric bootstrap sampling method. (a) 1st, (b) 10th, (c) 25th, and (d) 50th.

replacement, to obtain a new data set of n=396 observations, it may or may not select any of the few failures that occur in the extreme tails. This certainly proves to be an extreme limitation when our goal is to estimate the 1st percentile. We would expect only 4 out of about 400 to be below the 1st percentile.

On the other hand, the bootstrap is designed to simulate the sampling process and such variability present, as that in Figure 5.1(a) is certainly possible. Furthermore, these wide bootstrap intervals may be providing more useful information (warnings on uncertainty) to the engineer and/or practitioner regarding the variability present in the destructive sampling process. Note that the asymptotic normal intervals, which are theoretical, may be too narrow to capture all the information desired about the 1st percentile. I.e., they may be overly optimistic about the standard errors being smaller. For example, the asymptotic interval for the 1st percentile was [95.40 , 99.15] while the standard interval was [87.27 , 100.42]. This is quite an obvious difference!

As the percentiles increase and the "relative" IB data becomes more plentiful, the bootstrap confidence intervals are more closely matching the asymptotic intervals. For example, the standard bootstrap interval for the 50th percentile was [119.15 , 121.25] while the asymptotic interval was [119.27 , 121.22]. Also, it is useful to acknowledge that the different methods for constructing the bootstrap confidence intervals yielded very similar results. This supports that Figure 5.1 yields plots reasonably close enough to normality for all of these four intervals to be in agreement.

The reader need only briefly compare the intervals in Table 5.2 to see this result. This type of agreement can be expected when the sample size is sufficiently large since the bootstrap distributions will tend usually to be approximately normally distributed.

Thus, any of the proposed methods for the construction of bootstrap confidence intervals would prove to be useful here. As will be seen shortly, such agreement may not occur when the sample size is much smaller.

Table 5.3 and Figure 5.2 display the fully parametric bootstrap confidence intervals and the bootstrap sampling distributions for each percentile for Type 1 MDF, respectively. In this situation, the assumption of normality of Type 1 MDF is required. Here, the normal parameters were estimated from the original data and then used to simulate samples of size n=396.

A glance at the sampling distributions in Figure 5.2 reveals immediately its differences to the nonparametric sampling distributions of Figure 5.1. Notice that the distribution of the 1st percentile follows a normal distribution quite well in this larger sample case. Due to this, the intervals we obtain match very closely the asymptotically normal confidence intervals. Thus, rather than providing much more information about the IB percentiles, they help to confirm the accuracy of the asymptotic intervals. This can be a useful double check or potential warning, when needed. As with the nonparametric intervals for Type 1, there appear to be little differences between the methods for constructing bootstrap intervals.

Table 5.4 and Figure 5.3 show the confidence intervals and sampling distribution, respectively, for Type 1 MDF based on the NBSP sampling method described by Meeker and Escobar (1998). Recall that the sampling was done from the original data with replacement just as was done in the fully nonparametric method. The difference is that we must also assume the normal distribution as was done in the fully parametric case. Then, for each bootstrap sample, we estimate the normal parameters and use them to

Table 5.3. Fully parametric 95% bootstrap confidence intervals for Type 1 MDF.

p	$\hat{t}_p$ = quantile	Interval Type	LCL	UCL
.01	97.2830	Standard	95.3632	99.1280
		Bootstrap-t	95.8696	98.5501
		Percentile	95.4033	99.1887
		Bias-Corrected	95.3767	99.1410
.10	107.5917	Standard	106.2712	108.8809
		Bootstrap-t	106.6612	108.5103
		Percentile	106.2543	108.8704
		Bias-Corrected	106.2201	108.8598
.25	113.5573	Standard	112.4981	114.6588
		Bootstrap-t	112.8016	114.3739
		Percentile	112.4541	114.6764
		Bias-Corrected	112.5109	114.7343
.50	120.2372	Standard	119.3135	121.1804
		Bootstrap-t	119.6273	120.9105
		Percentile	119.2881	121.1449
		Bias-Corrected	119.2973	121.1502

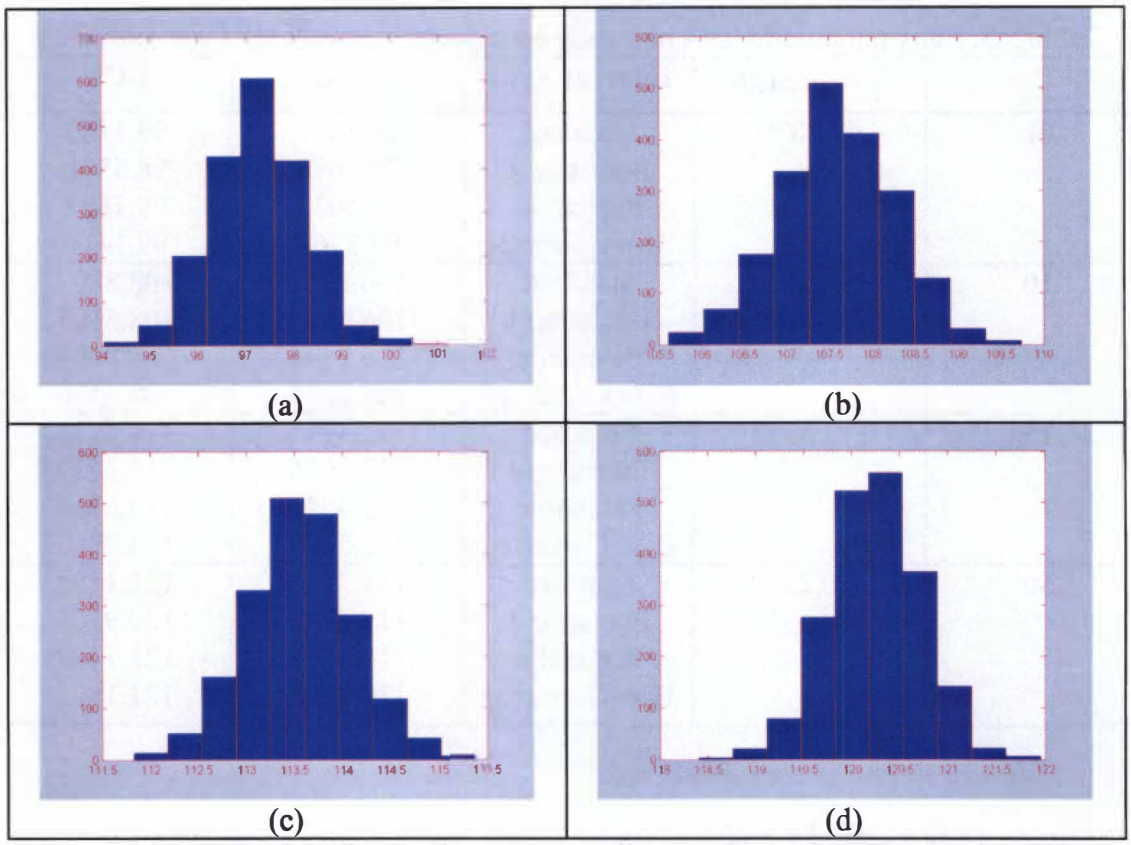


Figure 5.2. Sampling distribution of percentiles for Type 1 MDF under the fully parametric bootstrap sampling method. (a) 1st, (b) 10th, (c) 25th, and (d) 50th.

Table 5.4. NBSP 95% bootstrap confidence intervals for Type 1 MDF.

p	$\hat{t}_p$ = quantile	Interval Type	LCL	UCL
.01	97.3126	Standard	94.8629	99.6282
		Bootstrap-t	95.4981	98.9364
		Percentile	94.8999	99.6350
		Bias-Corrected	94.6364	99.4291
.10	107.6269	Standard	106.0747	109.0774
		Bootstrap-t	106.5334	108.6309
		Percentile	106.0586	109.1413
		Bias-Corrected	105.9957	109.0460
.25	113.6049	Standard	112.4488	114.7080
		Bootstrap-t	112.7936	114.3879
		Percentile	112.4553	114.6907
		Bias-Corrected	112.3902	114.6478
.50	120.2417	Standard	119.2569	121.2381
		Bootstrap-t	119.5950	120.9517
		Percentile	119.2832	121.2338
		Bias-Corrected	119.3141	121.3019

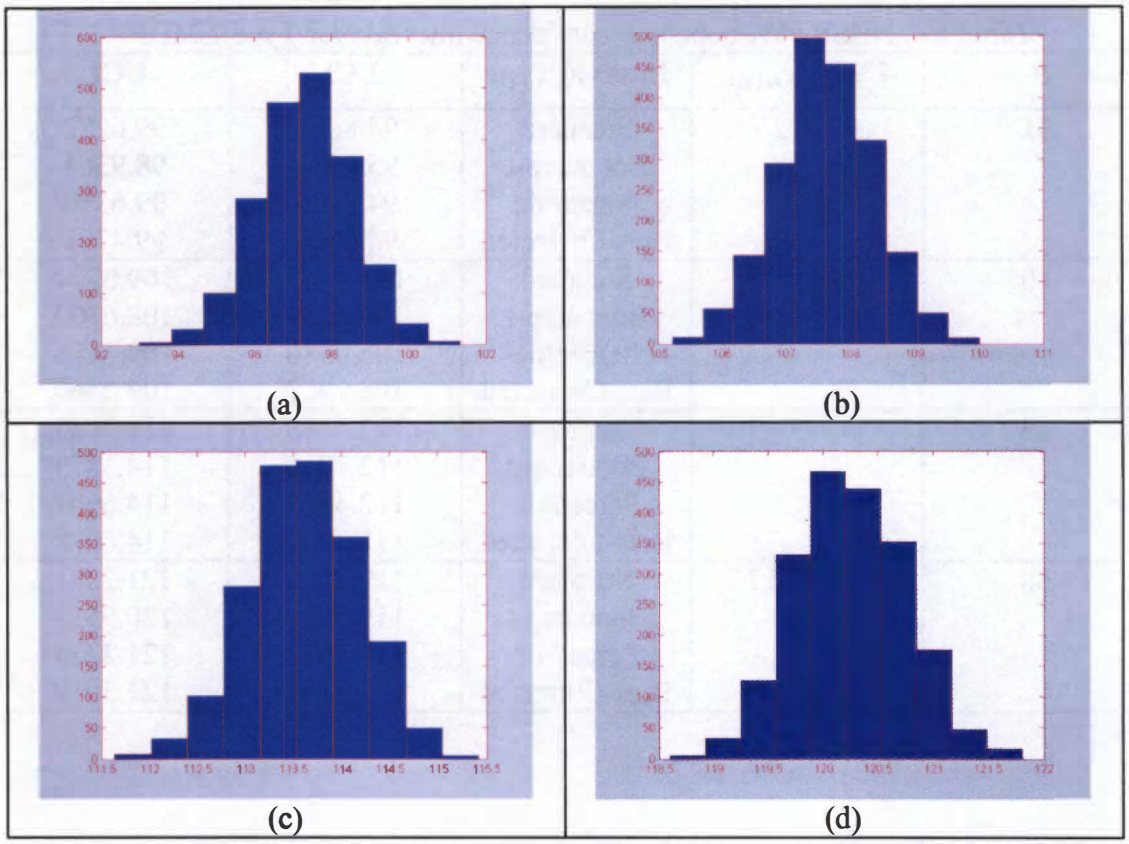


Figure 5.3. Sampling distribution of percentiles for Type 1 MDF under the NBSP method. (a) 1st, (b) 10th, (c) 25th, and (d) 50th.

parametrically estimate the percentiles of interest.

Therefore, this method proves to be a combination of the two aforementioned bootstrap methods and is in fact the method of choice for Meeker and Escobar (1998) for analyzing reliability data. As previously mentioned, they argue that a downside of the fully parametric bootstrap requires complete knowledge of the censoring mechanisms involved and must be taken into consideration when simulating the bootstrap samples.

Furthermore, the fully nonparametric method can lead to sampling distributions that are discrete, especially in the case of smaller sample sizes. Thus, they use the NBSP method, which does not require knowledge of the censoring mechanisms, and also does not give discrete sampling distributions with a reasonable sample size (i.e. n=7 or more).

The sampling distributions shown in Figure 5.3 are also normal for each of the percentiles. Again, we observe that the intervals are similar to the asymptotic intervals as well as similar among themselves.

Table 5.5 provides the 95% asymptotic normal intervals for Type 5 MDF. Table 5.6 and Figure 5.4 display the fully nonparametric intervals and sampling distributions, respectively, for Type 5 MDF percentiles. The sampling distributions shown provide an example of a limitation of the fully nonparametric bootstrap. When the sample size is relatively small, as is the case of Type 5 MDF, the sampling distributions are more discrete. Figure 5.4(a) and 5.4(b) show this clearly. Furthermore, the sampling distribution for the 1st percentile is very much skewed to the right. As the percentiles increase, the distribution becomes more symmetric (more normal) and has a more continuous appearance. Here, more differences among the various confidence intervals emerge.

Table 5.5. 95% Asymptotic normal confidence intervals for Type 5 MDF.

p	$\hat{t}_p$ = quantile	LCL	UCL
.01	150.4675	144.2243	156.7108
.10	165.3393	160.9623	169.7159
.25	173.9803	170.3872	177.5734
.50	183.5811	180.3380	186.8242

Table 5.6. Fully nonparametric 95% bootstrap confidence intervals for Type 5 MDF.

p	$\hat{t}_p$ = quantile	Interval Type	LCL	UCL
.01	149.3956	Standard	138.6630	155.0410
		Bootstrap-t	136.5856	147.2492
		Percentile	146.3000	161.1200
		Bias-Corrected	146.3000	150.7120
.10	165.3361	Standard	159.3674	169.4326
		Bootstrap-t	161.4890	166.6291
		Percentile	161.0000	168.9000
		Bias-Corrected	157.4000	168.2900
.25	172.8687	Standard	166.6393	177.9606
		Bootstrap-t	167.8679	175.4658
		Percentile	168.3000	177.8000
		Bias-Corrected	168.3000	177.6000
.50	185.2743	Standard	181.1404	189.9596
		Bootstrap-t	182.8249	190.2666
		Percentile	178.8000	189.0000
		Bias-Corrected	178.5433	188.8567

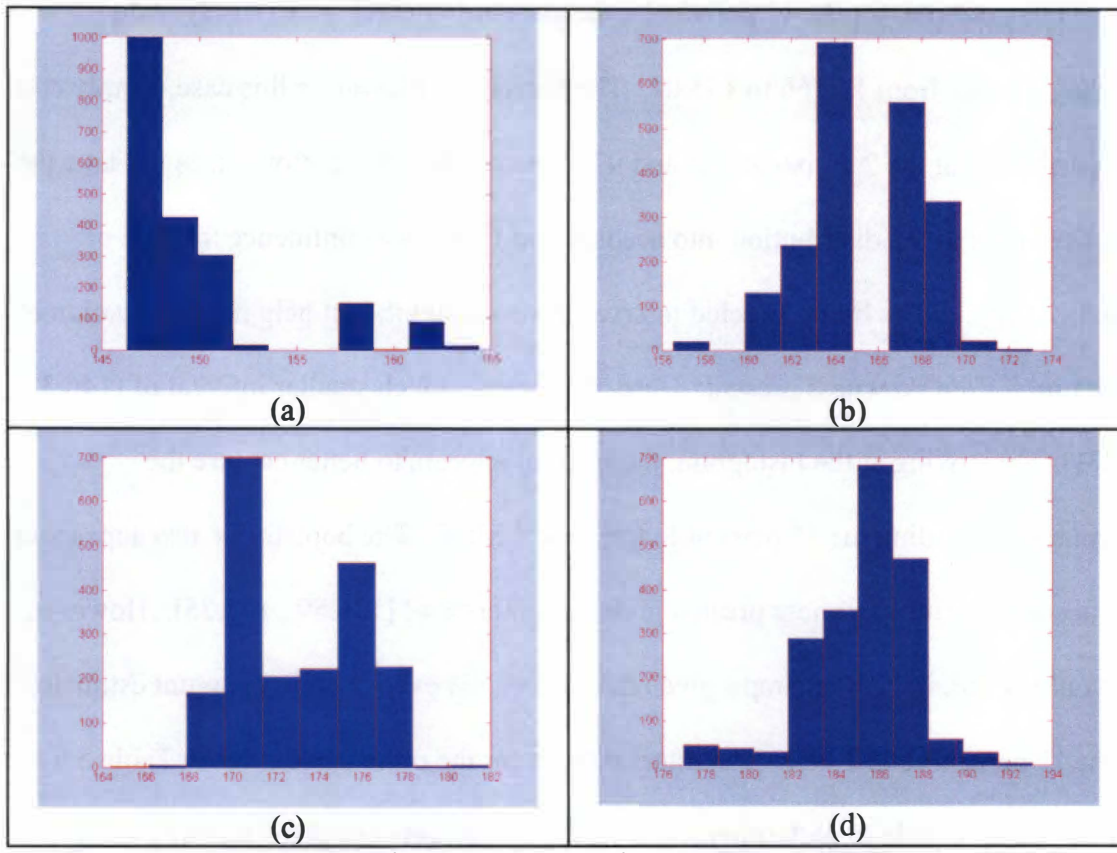


Figure 5.4. Sampling distribution of percentiles for Type 5 MDF under the fully nonparametric bootstrap sampling method. (a) 1st, (b) 10th, (c) 25th, and (d) 50th.

For example, for the 1st percentile, the standard interval is extremely wide covering a range from 138.66 to 155.04. The percentile interval, in this case, simply cuts the distribution at the 2.5th percentile and 97.5th percentile. It therefore, does not take the right skewness of the distribution into account and gives us a confidence interval of [146.3 , 161.12]. The bias-corrected interval provides significant help in this situation by correcting for the skewness present. Here, we obtain a much smaller interval of [146.3 , 150.71]. By looking at the histogram, the interval appears to better capture the information regarding the 1st percentile for Type 5 MDF. The bootstrap-t also appears to help account for the skewness present giving an interval of [136.59 , 147.25]. However, we would not trust the bootstrap-t given that it does not even contain the point estimate for the 1st percentile of 149.4. The interval types for the other percentiles in Table 5.6 are much more agreeable to each other. Workers and managers are given warning to not trust only one set of intervals in such cases. It provides a reality check. The histogram, also, present healthy warning signs to the user.

Practitioners are advised that when these histograms are discrete or appear "snaggle-toothed" to up the resampling size to, say, B=5000. If it no longer has a "snaggle-toothed" appearance, then the larger resampling size has helped. If it still, however, appears "snaggle-toothed" then practitioners are advised not to use the fully nonparametric approach for constructing bootstrap confidence intervals, at least not for the extreme lower percentiles. Instead, the three quartiles are likely safer or even just the median. Note well the warning of Polansky (1999) when the percentiles are very small such as 1% or 5% to not use bootstrap estimates.

In order to practice what we preach, since the histograms for the lower percentiles

in Figure 5.4 do appear "snaggle-toothed", the fully nonparametric bootstrap intervals were reconstructed using B=5000 bootstrap samples. What resulted were confidence intervals very similar to those in Table 5.6 and histograms that continued to have a "snaggle-toothed" appearance. Thus, even though the bias-corrected and bootstrap-t intervals have helped our situation a little, it is advised that the practitioner not use the lower percentile estimates here. It is very fortunate to have a graphical warning.

As more than likely suspected, the fully parametric intervals types shown in Table 5.7 are all agreeable to each other. Furthermore, they match very closely to the asymptotic normal intervals. Stated before, the usefulness of this would mainly be to assist in confirming the asymptotic intervals. Recall, from Chapter 4, that it was determined that Type 5 MDF follows a normal distribution. Figure 5.5 shows the sampling distributions, which are continuous and follow a normal distribution nicely. Placing a lot of faith in these parametric intervals may cause an incorrect inference about the Type 5 percentiles, especially if the distribution has been misspecified. It is essential to also point out that if the sample size had been larger, it is possible (and likely) that the bootstrap sampling distribution would approach a normal distribution making life in the bootstrap world much easier. The NBSP method intervals and sampling distributions are shown in Table 5.8 and Figure 5.6 respectively.

Comments analogous to the parametric intervals for Type 5 MDF described above apply to the NBSP method described by Meeker and Escobar (1998). That is, the intervals types are in greater agreement and the sampling distributions are normally distributed for each percentile. The plots are very helpful diagnostics.

Sampling using the NBSP method described by Meeker and Escobar (1998) may

Table 5.7. Fully parametric 95% bootstrap confidence intervals for Type 5 MDF.

p	$\hat{t}_p$ = quantile	Interval Type	LCL	UCL
.01	150.3468	Standard	143.9692	156.5137
		Bootstrap-t	145.6272	154.7383
		Percentile	143.8976	156.3403
		Bias-Corrected	143.2824	156.0068
.10	165.2401	Standard	160.9071	169.5224
		Bootstrap-t	161.9716	168.2961
		Percentile	161.0064	169.5044
		Bias-Corrected	161.0522	169.5169
.25	173.9953	Standard	170.2529	177.5766
		Bootstrap-t	171.2273	176.4632
		Percentile	170.4550	177.8767
		Bias-Corrected	170.3784	177.7188
.50	183.5092	Standard	180.3048	186.8574
		Bootstrap-t	181.2664	185.8846
		Percentile	180.2122	186.7233
		Bias-Corrected	180.3048	186.8574

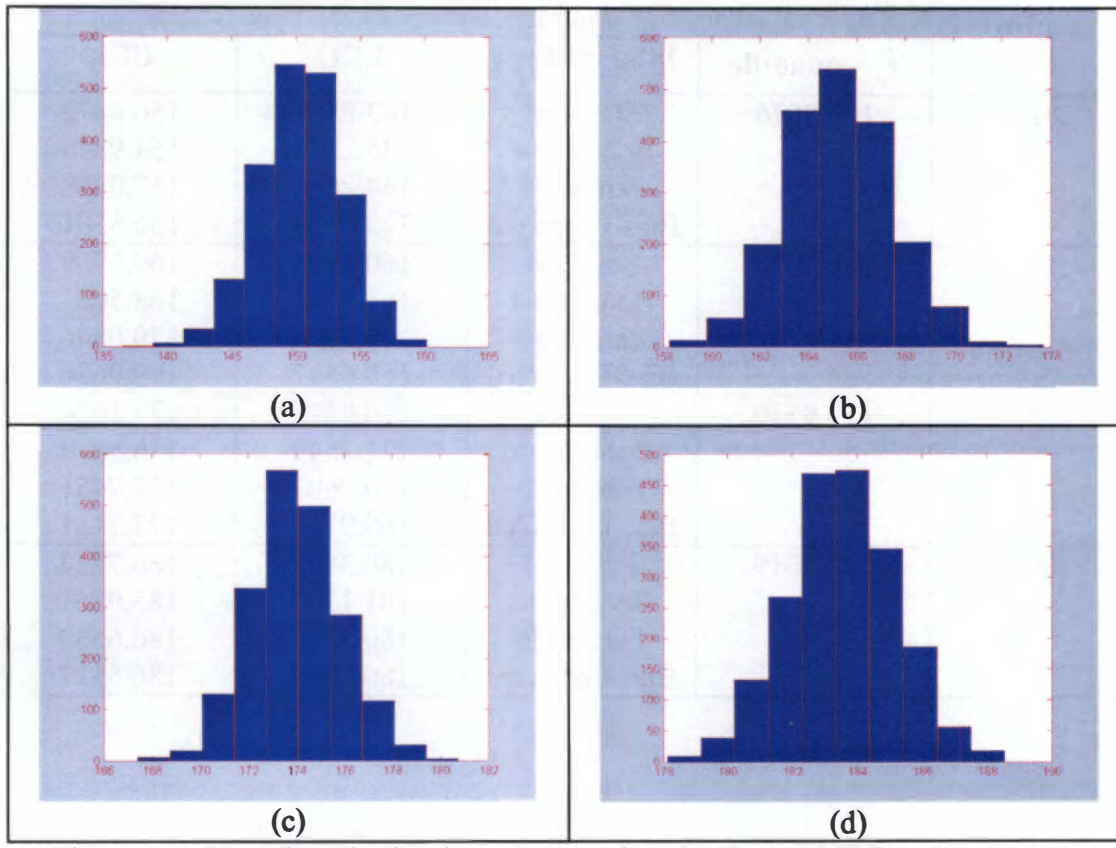


Figure 5.5. Sampling distribution of percentiles for Type 5 MDF under the fully parametric bootstrap sampling method. (a) 1st, (b) 10th, (c) 25th, and (d) 50th.

Table 5.8. NBSP 95% bootstrap confidence intervals for Type 5 MDF.

p	$\hat{t}_p$ = quantile	Interval Type	LCL	UCL
.01	150.5676	Standard	143.8357	156.6472
		Bootstrap-t	145.2324	154.9846
		Percentile	144.2618	157.0768
		Bias-Corrected	143.5739	156.5301
.10	165.3426	Standard	160.4306	169.9989
		Bootstrap-t	161.6941	168.5081
		Percentile	160.7296	170.0846
		Bias-Corrected	160.6547	169.9636
.25	173.9540	Standard	170.0271	177.8024
		Bootstrap-t	171.2549	176.5879
		Percentile	170.0607	177.7451
		Bias-Corrected	169.9757	177.7111
.50	183.5619	Standard	180.3897	186.7724
		Bootstrap-t	181.1307	185.9261
		Percentile	180.2257	186.6669
		Bias-Corrected	180.1838	186.5517

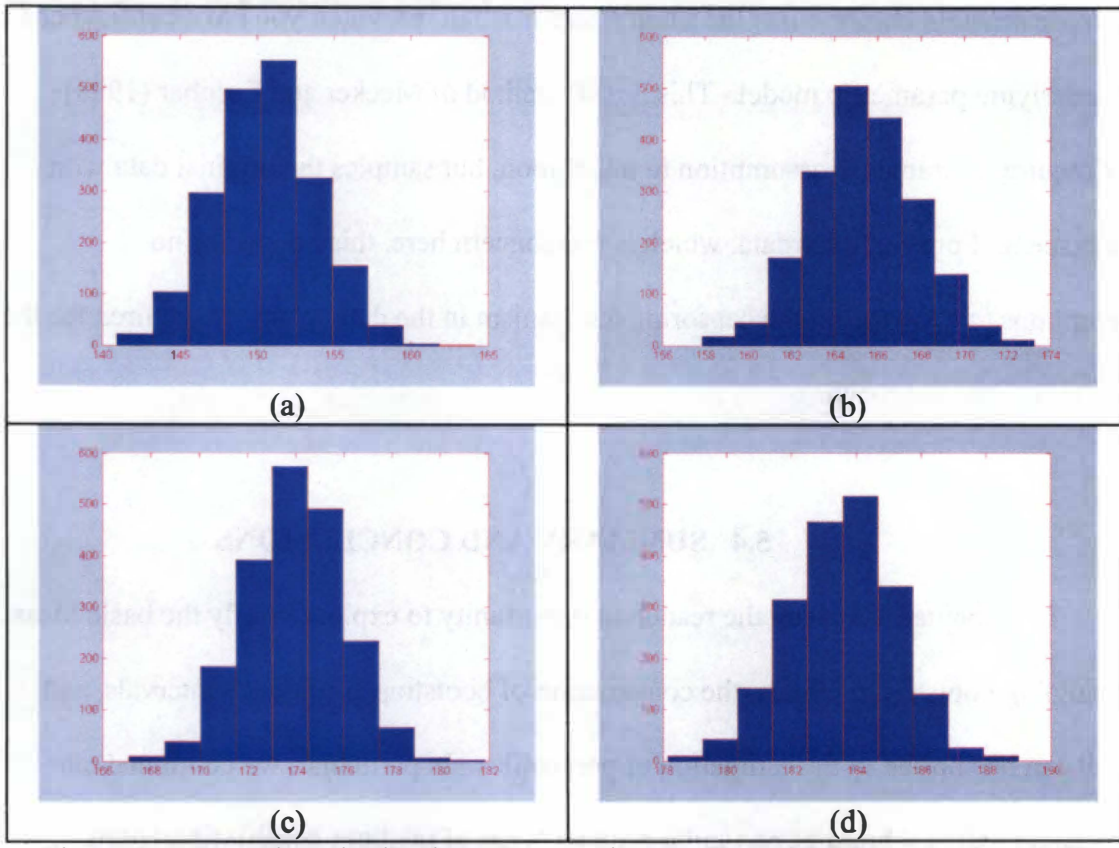


Figure 5.6. Sampling distribution of percentiles for Type 5 MDF under the NBSP method. (a) 1st, (b) 10th, (c) 25th, and (d) 50th.

be a more sensible choice when the sample size is small, provided you have confidence in the underlying parametric model. This NBSP method of Meeker and Escobar (1998) does require a parametric assumption to build upon, but samples the original data with replacement. For reliability data, which is our concern here, this allows for no assumptions to be made on the censoring mechanism in the data, which is required for the fully parametric approach.

5.4 SUMMARY AND CONCLUSIONS

This chapter has given the reader an opportunity to explore briefly the basic ideas surrounding bootstrap methods, the construction of bootstrap confidence intervals, and how it can be applied to the estimation of percentiles. In particular, we continued our study of the internal bond of particular product types of medium density fiberboard.

The different methods described for obtaining bootstrap samples include the fully parametric, fully nonparametric, and a mix between the parametric and nonparametric methods. Along with the different sampling methods, four different types of bootstrap confidence intervals were discussed. These include the standard interval, bootstrap-t interval, percentile, and bias-corrected percentile interval. For Type 1 and Type 5 MDF, bootstrap confidence intervals for each of the described sampling methods were constructed for the 1^{st}, 10^{th}, 25^{th}, and 50^{th} percentiles. The asymptotic normal intervals were shown to aid in the comparison.

For a sufficiently large sample size, as is the case for Type 1 MDF, the bootstrap sampling distributions appear continuous (i.e. does not have any holes) and follow roughly a normal distribution. In this case, it is relatively a matter of preference as to

which of the bootstrap interval types is used. They all provide very similar and accurate results as was easily observed. This is even the case using the fully nonparametric approach, although some care should be taken when examining the 1st percentile. Also, this is useful since no distributional assumptions are required and the worry of misspecification of the model is alleviated.

Furthermore, with a large sample size, the bootstrap sampling distribution appears continuous, allowing for reliable results. Note the reader can understand it being recommended that when the sample size is large, nonparametric sampling is an appropriate safer choice and can be used more confidently. A large sample size helps to make up for information that is lost when not assuming a parametric distribution. Again, we repeat that any of the described interval types would be useful in this environment. They were all approximately the same, which is reassuring for the practitioner.

It was shown that when the sample size is sufficiently large, the methods for constructing bootstrap confidence intervals were comparable to the asymptotic intervals as would be expected. As the percentiles increased from 1 to 50, the confidence intervals became narrower, given the larger quantities of observed failure data. I.e., the standard errors grew smaller. This is especially seen with the fully nonparametric case. The interval for the 1st percentile is much wider than the intervals for the 10th, 25th, and 50th percentiles. However, this result follows naturally from the sampling method and the lack of observed failure data in the extreme lower tail. Although this occurs, the nonparametric bootstrap can provide accurate results when the sample size is large and is recommended when the parametric assumptions are suspect.

Conversely, when the sample size is much smaller, as is the case for Type 5 MDF,

and when sampling is done using the fully nonparametric method, the bootstrap sampling distributions can be anything but continuous and may or may not follow a normal distribution. Recall to always check the plots. Furthermore, the nonparametric bootstrap does not yield intervals that are similar. Naturally, this adds complications and requires other considerations than those recommended for the large sample case.

If no distributional assumptions can be made, it is recommended that the practitioner make use of the bootstrap-t intervals or perhaps as a first choice the bias-corrected percentile intervals with great humility. Doing this can still produce roughly accurate results for the median or quartiles using the nonparametric method when the sample size is small. These intervals help to alleviate some of the frustration that can be caused by having a sampling distribution that does not follow, at least roughly, a normal distribution. Furthermore, they are both second-order accurate intervals. However, we would recommend not using confidence intervals for the lower percentiles and instead resort to another approach. Thus, we should place little faith in the confidence intervals for the 1^{st} or 10^{th} percentiles of Type 5 MDF shown in Table 5.6. Recall the warnings of Polansky (1999) to not even estimate the lower percentiles when the plots appear discrete or "snaggle-toothed." Also, note Polansky (2000) and his helpful insights into using kernel smoothing to better estimate lower percentiles in smaller samples.

Ideally, the best answer to estimating lower percentiles realistically is to have a larger sample. Note that the 1^{st} percentile is not robust in any sample less than 100 because by just changing the minimum, we can make the 1^{st} percentile virtually any number less than the old minimum up to the second order statistic. We want to stress the non-robustness of the 1^{st} percentile in samples like Type 5 with n=74. It is recommend

that at least a sample of size 200 or more be used. Further study might be done in the future to investigate this in more detail. Next, three alternatives are suggested to get around this difficulty if cost is prohibitive.

First, as an alternative, we need to study the outliers, which can be classified as outliers due to measurement error or due to statistical variation. One might do bootstrapping in a way that takes into account the few or many outliers in any particular data set. We leave that for a future study.

Another approach to estimating lower percentiles with a small sample would be to use the multiple regression equation in Young and Guess (2002) for estimating IB for a much larger sample. Then, use that larger sample to get more realistic estimates on the lower percentiles. This would save money and time but would need to be continuously validated as an appropriate model by actual destructive sampling. Alternatively, engineering judgment and experiences could be incorporated into a helpful Bayesian approach to get more realistic estimates on the lower percentiles when the data is small. Chapter 14 of Meeker and Escobar (1998) provides a thorough treatment of Bayesian methods for reliability data.

Sampling using the NBSP method described by Meeker and Escobar (1998) may be a more sensible choice when the sample size is small, provided you have confidence in the underlying parametric model. Recall the information criteria discussed in chapter 4 combined with Q-Q plots help with such needed parametric assessments with less data. By constructing intervals in this manner, the bootstrap sampling distributions appear continuous and roughly follow a normal distribution. In this case, the confidence interval construction methods produced similar intervals. Although requiring at least

approximate parametric assumptions, this method was useful in constructing intervals for the extreme lower percentiles.

This NBSP method of Meeker and Escobar (1998) does require a parametric assumption to build upon, but samples the original data with replacement. For reliability data, which is our concern here, this allows for no assumptions to be made on the censoring mechanism in the data, which is required for the fully parametric approach. For our data here, we had no censoring.

As a brief aside, let us recall the dangers when a user may misspecify a model mentioned previously in the thesis. The distributions of IB appear continuous and following normal distributions approximately, but not perfectly. If it was perfectly normal, one can do exact confidence intervals. See, for example, Lawless (1982). Our approach with the nonparametric bootstrap protects the user from assuming a perfect normal distribution and still applies then, whereas the exact procedure would not be completely exact. Also, the reader will note that the bootstrap-t would approximate very well those exact confidence intervals in the perfectly normal world. This is a nice feature and extra validation in practice. When, however, exact procedures are not available we can still do bootstrapping.

The bootstrap sampling distributions from the important diagnostic graphs for the NBSP method appear normal. They should always be checked to prevent the misuse of bootstrap inappropriately. There is never getting something for nothing. The different bootstrap intervals produce relatively similar results and the choice of interval for use can be based on preference because of the normality of the sampling distributions. It is, then, recommended that when a small sample size cannot be avoided and when one has

confidence in parametric assumptions, that one should make use first of the NBSP method from Meeker and Escobar (1998).

The fully parametric bootstrap is useful for verifying classical results using the familiar textbook formulas. Otherwise, there is no significant advantage for using the parametric bootstrap over the commonly known classical formulas, except for double-checking. Recall previous related comments.

Overall, sample size is the key player in our game of bootstrapping. A large sample size allows for more promising results when no distributional assumptions are made. Smaller sample sizes give way to needed limitations. Chernick (1999) tells us that "the main concern in small samples is that with only a few values to select from, the bootstrap sample will under represent the true variability as observations are frequently repeated and the bootstrap samples themselves repeat." This does not mean that the bootstrap should not be used with small sample sizes. Rather, much greater care should be taken when analyzing the results and their accuracy. It has been recommended that in the case of constructing confidence intervals, that more than 1000 bootstrap samples should be generated. This number can be and should be increased even more when the sample size is small.

It is standard practice to create bootstrap samples the same size as the original data being sampled from, as we have done in our MDF examples. The practitioner, however, may find it useful to create bootstrap samples as large as the data with the most observations. For illustration, since Type 1 MDF had n=396 observations, we resampled from this data to create bootstrap samples of size n=396. However, for Type 5 MDF with n=74 observations, we would recommend a future practitioner resample from this data

and create bootstrap samples also of size n=396. By creating bootstrap samples in this manner, one is able to control the overall sampling variation and focus instead on other sources of variation that are of greater interest to the practitioner. This was done for the fully nonparametric bootstrap and a percentile interval was constructed for the 1st percentile. The results yielded [146.3 , 148.6] which, as expected, had a smaller length than that previously shown. Compare with Table 5.6 above. Even though what we have done is very standard currently, we recommend this alternative highly for practitioners. We thank Seaver (2004) for these helpful insights.

It is the hopes of this author that the reader will take away a general knowledge of the bootstrap and find it to be a useful and helpful tool for analyzing data (reliability data, in particular). The common and practical use of the computer and ease of implementing the bootstrap algorithms make it a good candidate for conducting statistical inference. Possible future work with respect to bootstrapping and MDF includes observing differences in percentiles over time and shift. Efron (2003) remarks, "These days statisticians are being asked to analyze much more complicated problems... . I believe, or maybe just hope, that a powerful combination of Bayesian and frequentist methodology will emerge to deal with this deluge of data and that computer-intensive methods like the bootstrap will facilitate the combination."

Chapter 6

Summary and Concluding Remarks

It has been the purpose of this thesis to introduce and illustrate useful methods for analyzing reliability data. The discussions in the previous chapters have been concise for the purpose of creating a setting conducive for the practitioner and to keep the thesis from being too long. However, if interested, the reader is encouraged to refer to the appropriate cited sources, among others that these authors cite, to obtain extra details. By presenting the subject in this manner, it is hopeful that the practitioner will be able to easily understand and implement these methods without having to sift through excessive theoretical discussions. Applications of these methods in the forest products industry were used throughout.

The forest products industry was an appropriate choice for demonstration since it has seen tremendous growth in recent years and impacts the economies of many countries. Recall the state of Tennessee has an annual impact of all forest products on the order of $22 billion per year, compared to Maine of around $9 billion per year.

With government regulations being enforced that limit the amount of available raw materials; it has become more important to focus on environmental issues and producing higher quality products. Statistical reliability and quality control methods, to name a few, have been employed more to monitor the quality of forest products. With significant research and applied efforts devoted to this area, data is plentiful. This certainly allows for many helpful examples and case studies to illustrate specific statistical methods. In particular, this thesis focused on applying reliability methods,

information criteria, and bootstrapping to better understand the strength to failure of the internal bond (IB) of medium density fiberboard (MDF). Recall that MDF is a high quality engineered timber product. IB, measured in pounds per square inch (psi), is one of the key metrics of quality that is obtained during destructive testing. Basically, IB is one measurement for the strength of MDF.

Chapter 2 reviewed current literature with a large focus on MDF, IB research, and its improvement. Recall as an example that Wang, Chen et al. (1999) investigated "a compression shear device for easy and fast measurement of the bonded shear strength of wood-based materials to replace the conventional method used to evaluate internal bond strength (IB)." They found that measuring strength of MDF or particleboard by the suggested compression shear strength and by the conventional approach of internal bond strength were significantly correlated. This provides an alternative approach to measuring strengths of materials. Mentioning again the above reference helps to illustrate the importance of understanding the strength of the IB. Other helpful resources regarding IB can be found in Chapter 2.

Additionally, Chapter 2 delves into the statistical literature pertaining to reliability, information criteria, and bootstrapping. The cited references are helpful in obtaining more thorough discussions of the topics covered in this thesis. Reliability data analysis, along with MDF, is the recurrent theme of this thesis. As previously mentioned, reliability data refers to survival or failure time data and the analysis of this data is an important topic for industry and government. Many great books are devoted to this topic, including the classic Meeker and Escobar (1998).

In Chapter 3, exploratory data analysis techniques were utilized to examine the IB

of MDF and to draw useful initial conclusions. It, also, helped motivate the need for Chapter 4 using information criteria. Histograms and scatter plots were shown first in Chapter 3 to gain preliminary insight on the distributions of the MDF product types with respect to IB. In particular, histograms for Types 1 and 2 revealed a symmetric distribution, which certainly can provide us with useful information regarding variability in the data plus insight on the underlying parametric distribution. Recall additional comments in Chapter 3. The scatter plot of each type intuitively tells us that Type 2 is a stronger product than Type 1. A t-test confirmed this with $p<0.0001$. Thus, Type 2 would more likely be used in shelving, for example, than Type 1.

The use and helpfulness of probability plots to characterize the underlying distribution of the IB for different product types were also described. The points on the plot will form an approximate straight line if the data set "conforms" to a particular distribution. Thus, these plots, in many instances, can correctly identify the parametric distribution. Meeker and Escobar (1998) illustrate instances, however, where these plots may not correctly identify the parametric distribution. This misspecification is largely based on sample size and subjective visualization. Quite obviously, then, these plots do have subjectivity to them and one must be aware of this. Probability plots are useful tools, but the results produced should not always be taken to be concrete and definitive.

When the parametric model assumption is weak or absent, nonparametric plots known as the Kaplan-Meier estimators, survival plots, or reliability plots provide another useful way to explore reliability data. These plots can provide useful insight and show surprising results from the data. For the different MDF product types, the survival plots clearly indicated that Type 2 is stronger than Type 1 MDF.

What is interesting to note here is that Type 2 has a higher density of 48 lbs/ft^3 than Type 1 with a density of 46 lbs/ft^3. When, product types of the same density but different thickness were compared, little to no differences was evident from the plots. Therefore, density appears to be a key driver in IB variability. This information was determined based on the plots rather than a particular statistical test. It is remarkable the amount of information that can be extracted from the data based on graphical procedures! Often, though, we do wish to know the parametric distribution with more statistical assurance. More exact inferences can follow from having this parametric model, when valid.

Probability plots were discussed in Chapter 3 and above in an exploratory data analysis context. They were revealed as a helpful yet subjective tool that can provide the practitioner with an appropriate starting point for determining the parametric distribution of a particular data set. Rather than dispensing with these plots, Chapter 4 presented tools that make probability plots more objective and allow the choice of parametric distribution to be much clearer. This is accomplished through the use of information criteria that assign a numeric score to each plot. The distribution characterized by the probability plot with the lowest score is considered to be the best among competing models. This is a great help to practitioners.

The particular information criteria discussed in Chapter 4 include Akaike's Information Criteria (AIC) and Bozdogan's Information Complexity Criterion (ICOMP). These criteria have a lack of fit term that accounts for the likelihood of the proposed model as well as a term that accounts for the number of parameters/complexity in the model. A first order approximate model was fit through the points on the probability plot

in order to score them using AIC and ICOMP. The background is provided for these two criteria as well as for how the probability plot scores were obtained. However, more references are provided to point the reader to more extensive information.

An example using MDF product Types 1, 2, 3, and 5 was shown to illustrate the usefulness of AIC and ICOMP. In particular, probability plots for the normal, lognormal, and Weibull distributions were constructed and shown along with their corresponding AIC and ICOMP scores. For example, Type 1 MDF probability plots had ICOMP scores of 1439.5, 1498.1, and 1687.5 for the normal, lognormal, and Weibull distributions, respectively. It is quite easy to see that the lowest score of 1439.5 corresponds to the normal distribution. Therefore, we determined that among the competing models, the normal distribution provided the best fit for Type 1 MDF. Recall that this was suspected based on the exploratory graphics developed in Chapter 3. However, information criteria make this assumption more plausible. Furthermore, Type 2 was determined to follow the lognormal distribution while Types 3 and 5 follow a normal distribution.

Aside from knowing the parametric distribution, percentiles are often of key importance in reliability studies. Reliability engineers are frequently interested in the time at which 10% (or even 1%) of a particular product will fail. Specifically, percentiles can help in understanding product warranties and their costs. The mean or standard deviation of failure time data is not usually of concern and would not provide as much information as the percentiles in this context. Recall, again, Meeker and Escobar (1998).

One of the difficulties in estimating the percentiles (in particular, lower percentiles) is that the data may not be plentiful in the lower tail of the distribution. Therefore, if the parametric distribution is available and is considered strong, then

estimating the percentiles parametrically can provide accurate results.

If the parametric distribution is not available, asymptotic approximations can be utilized. In particular, the sample percentiles have been shown to be asymptotically normal. However, unless the sample size is sufficiently large, these asymptotic intervals will not necessarily yield accurate results. It then becomes important to search for another method for estimating these desired characteristics. Bootstrapping methods were presented in Chapter 5 as a useful method for constructing confidence intervals for percentiles. Histograms provide a diagnostic to warn of the misuse of this method in some cases. As always, the practitioner/user must check procedures for appropriateness of use.

Chapter 5 presents several methods for resampling the data along with different methods for constructing bootstrap confidence intervals. Bootstrap sampling methods described included nonparametric, parametric, and mixed (i.e. nonparametric sampling for parametric inference or NBSP). The confidence interval methods included the standard, bootstrap-t, percentile, and bias-corrected percentile intervals. Sufficient background information was provided in order for the practitioner to implement the methods presented.

An application of these methods to the MDF data was shown also in Chapter 5. Bootstrap confidence intervals for the 1^{st}, 10^{th}, 25^{th}, and 50^{th} percentiles were constructed for Types 1 and 5 MDF. These two product types were chosen as they would aid in illustration and provide a useful contrast. Type 1 is a rather large sample size of n=396 IB observations while Type 5 is much smaller with n=74 IB observations.

It was shown that when the sample size is sufficiently large, the methods for

constructing bootstrap confidence intervals all produced similar results and were comparable to the asymptotic intervals as would be expected. As the percentiles increased from 1 to 50, the confidence intervals became narrower, given the larger quantities of observed failure data. I.e., the standard errors grew smaller. This is especially seen with the fully nonparametric case. The interval for the 1^{st} percentile is much wider than the intervals for the 10^{th}, 25^{th}, and 50^{th} percentiles. However, this result follows naturally from the sampling method and the lack of observed failure data in the extreme lower tail. Although this occurs, the nonparametric bootstrap can provide accurate results when the sample size is large and is recommended when the parametric assumptions are weak.

When the sample size is small, the nonparametric bootstrap does not yield intervals that are similar. In this case, the bootstrap sampling distributions can be discrete and standard methods are not usually recommended. To partly remedy this, it is loosely recommended that the bootstrap-t or the bias-corrected percentile methods be utilized. Doing this can still produce roughly accurate results for the median or quartiles using the nonparametric method when the sample size is small. These intervals help to alleviate some of the frustration that can be caused by having a sampling distribution that does not follow, at least roughly, a normal distribution. Furthermore, they are both second-order accurate intervals. However, we would recommend not using confidence intervals for the lower percentiles and instead resort to another approach such as imputation and/or Bayesian methods.

It is thus recommended, but not always, that when the sample size is small, that the practitioner makes use of the NBSP method described thoroughly by Meeker and

Escobar (1998). The plots can give warning when not to use this approach. As described, this method does require parametric assumptions, but samples the data as done in the fully nonparametric case. By constructing intervals in this manner, the bootstrap sampling distributions appear continuous and roughly follow a normal distribution. In this case, the confidence interval construction methods produced similar intervals. Although requiring at least approximate parametric assumptions, this method was useful in constructing intervals for the extreme lower percentiles.

As already mentioned several times before, to be able to say that improvements have been made, we must be able to measure reliability expressed in percentiles that allow for statistical variation. We need to make comparisons of these reliability measures between products and within products before and after process improvement interventions. Knowing when to trust confidence intervals and when not to trust them are crucial for managers and users of MDF to make successful decisions.

Through all the explanations, illustrations, and summaries, it is hopeful that the reader will come away with some insight on practically analyzing reliability data and will be able to implement the statistical methods presented into their own research. At this point, it is common to wonder what might come next. With regards to the IB of MDF, future work on studying other sources of variation present is a possibility.

In general, other work is helpful regarding information criteria and bootstrapping. In Chapter 4, the model fit to the probability plots was a simple first order approximation that assumed normal errors with constant variance. It is possible and likely that another model would provide a better fit. Certainly, it is also feasible to assume that the error variation is not constant and rather, instead, varies with the data. We leave that work to

another time.

With respect to bootstrapping, the methods presented in Chapter 5 are not exhaustive. Other methods for constructing intervals certainly exist and these may be explored in greater detail. Furthermore, more percentiles may be estimated for the MDF data to compare them over time, by shift, month, etc. Bootstrapping is receiving much more attention in recent years than it ever has. It is likely that bootstrap methods will be further developed, refined, and built upon to better suit the practitioner's needs. See comments in Chapter 5 about Type 5 MDF for alternative approaches for estimating lower percentiles as possible future work.

In any case, more research is always possible and likely to appear in future works. Other possibilities than those listed above are certain to be explored. Reliability data analysis has rich theoretical foundations as can easily be seen in excellent works such as Barlow and Proschan (1975) and Meeker and Escobar (1998). However, the theory leads to many applications that are of interest to engineers, researchers, and other practitioners in industry, government, academia, etc. It is for these applications that this thesis is written.

While focusing on the forest products industry and application to MDF, the methods discussed and illustrated flow easily into a plethora of other worlds. We close, appropriately, with words from Meeker and Escobar (1998) who assert that "reliability is being viewed as the product feature that has the potential to provide an important competitive edge. A current industry concern is in developing better processes to move rapidly from product conceptualization to a cost-effective highly reliable product. A reputation for unreliability can doom a product, if not the manufacturing company."

Bibliography

Akaike, H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principle.* Second international symposium on information theory, Budapest, Academiai Kiado.

Barlow, R. E. and Proschan, F. (1965). *Mathematical Theory of Reliability*. Wiley, New York, NY.

Barlow, R. E. and Proschan, F. (1974). *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt Rinehart and Winston, New York, NY.

Barlow, R. E. and Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing: Probability Models*. To Begin With, Silver Spring, MD.

Beran, R. (2003). The Impact of the Bootstrap on Statistical Algorithms and Theory. *Statistical Science*, **18**(2), 175-184.

Boos, D. D. (2003). Introduction to the Bootstrap World. *Statistical Science*, **18**(2), 168-174.

Bozdogan, H. (1987). Model Selection and Akaike's Information Criteria (AIC): The General Theory and Its Analytical Extensions. *Psychometrika*, **52**, 345-370.

Bozdogan, H. (1988). ICOMP: A New Model Selection Criterion. *Classification and Related Methods of Data Analysis*. In Hans H. Bock (Ed.), Amsterdam, North-Holland, Springer, 599-608.

Bozdogan, H. (1990). On the Information-Based Measure of Covariance Complexity and Its Application to the Evaluation of Multivariate Linear-Models. *Communications in Statistics-Theory and Methods*, **19**(1), 221-278.

Bozdogan, H. (1996). *A New Informational Complexity Criterion for Model Selection: The General Theory and Its Applications*. Invited Paper presented in the Session on Information Theoretic Models & Inference of the Institute for Operations Research & the Management Sciences (INFORMS), Washington, D.C.

Bozdogan, H. (2000). Akaike's Information Criterion and Recent Developments in Information Complexity. *Journal of Mathematical Psychology*, **44**(1), 62-91.

Bozdogan, H. (2001). *Professor Bozdogan's Class Notes of Statistics 563-564: Mathematical Statistics at the University of Tennessee, Knoxville.* Copyright Professor H. Bozdogan.

Bozdogan, H. and Bearse, P. (2003). Information Complexity Criteria for Detecting Influential Observations in Dynamic Multivariate Linear Models Using the Genetic Algorithm. *Journal of Statistical Planning and Inference*, **114**(1-2), 31-44.

Bozdogan, H. and Haughton, D. M. A. (1998). Informational Complexity Criteria for Regression Models. *Computational Statistics & Data Analysis*, **28**(1), 51-76.

Bozdogan, H. and Sclove, S. L. (1984). Multi-Sample Cluster-Analysis Using Akaike's Information Criterion. *Annals of the Institute of Statistical Mathematics*, **36**(1), 163-180.

Butterfield, B., Chapman, K., Christie, L., and Dickson, A. (1992). Ultrastructural Characteristics of Failure Surfaces in Medium Density Fiberboard. *Forest Products Journal*, **42**(6), 55-60.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury/Thomson Learning, Pacific Grove, CA.

Chernick, M. R. (1999). *Bootstrap Methods: A Practitioner's Guide*. Wiley, New York, NY.

Chow, P., Bao, Z., Youngquist, J. A., Rowell, R. M., Muehl, J. H., and Krzysik, A. M. (1996). Properties of Hardboards Made from Acetylated Aspen and Southern Pine. *Wood and Fiber Science*, **28**(2), 252-258.

Chow, P. and Zhao, L. (1992). Medium Density Fiberboard Made from Phenolic Resin and Wood Residues of Mixed Species. *Forest Products Journal*, **42**(10), 65-67.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge; New York, NY.

Deming, W. E. (1986). *Out of the Crisis*. Massachusetts Institute of Technology's Center for Advanced Engineering Design, Cambridge, MA.

Deming, W. E. (1993). *The New Economics*. Massachusetts Institute of Technology's Center for Advanced Engineering Design, Cambridge, MA.

Diaconis, P. and Efron, B. (1983). Computer-Intensive Methods in Statistics. *Scientific American*, **248**(5), 116-130.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap Confidence Intervals. *Statistical Science*, **11**(3), 189-212.

Efron, B. (1981). Nonparametric Standard Errors and Confidence Intervals. *Canadian Journal of Statistics*, **9**, 139-172.

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, **82**(397), 171-185.

Efron, B. (2003). Second Thoughts on the Bootstrap. *Statistical Science*, **18**(2), 135-140.

Efron, B. and Gong, G. (1983). A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *American Statistician*, **37**(1), 36-48.

Efron, B. and Tibshirani, R. (1991). Statistical-Data Analysis in the Computer-Age. *Science*, **253**(5018), 390-395.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York, NY.

English, B., Jensen, K. and Menard, J. (2004). Tennessee's forest and forest products industry and associated economic impacts for 2000. The University of Tennessee Agricultural Experiment Station Research Series. 61p.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proc. Camb. Phil. Soc.*, **24**, 180-190.

Ghosh, M., Parr, W. C., Singh, K., and Babu, G. J. (1984). A Note on Bootstrapping the Sample Median. *Annals of Statistics*, **12**(3), 1130-1135.

Giudici, P. (2003). *Applied Data Mining: Statistical Methods for Business and Industry.* John Wiley and Sons, New York, NY.

Gomez-Bueso, J., Westin, M., Torgilsson, R., Olesen, P. O., and Simonson, R. (2000). Composites Made from Acetylated Lignocellulosic Fibers of Different Origin - Part I. Properties of Dry-Formed Fiberboards. *Holz Als Roh-Und Werkstoff*, **58**(1-2), 9-14.

Guess, F. M., Edwards, D. J., Pickrell, T. M., and Young, T. M. (2003). Exploring Graphically and Statistically the Reliability of Medium Density Fiberboard. *International Journal of Reliability and Applications*, **4**(4), 97-110.

Guess, F. M., Walker, E., and Gallant, D. (1992). Burn-in to Improve Which Measure of Reliability? *Microelectronics and Reliability*, **32**, 759-762.

Hahn, G. J. and Meeker, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners*. Wiley, New York, NY.

Hall, P. (2003). A Short Prehistory of the Bootstrap. *Statistical Science*, **18**(2), 158-167.

Han, G. P., Umemura, K., Zhang, M., Honda, T., and Kawai, S. (2001). Development of High-Performance Uf-Bonded Reed and Wheat Straw Medium-Density Fiberboard. *Journal of Wood Science*, **47**(5), 350-355.

Hashim, R., Murphy, R. J., Dickinson, D. J., and Dinwoodie, J. M. (1994). The Mechanical-Properties of Boards Treated with Vapor Boron. *Forest Products Journal*, **44**(10), 73-79.

Hollander, M. and Wolfe, D. A. (1973). *Nonparametric Statistical Methods*. Wiley, New York, NY.

Hsu, W. E. (1993). *Cost-Effective Pf-Bonded Fiberboard*. Proceedings of the twenty-seventh Washington State University international particleboard/composite mater.

Kim, H., Guess, F. M., and Young, T. M. (2004). Using Data Mining Tools of Decision Trees in Reliability Applications. Submitted to a Professional Journal. For a Copy of This Paper, Send an Email to Timothy Young at Tmyoungl@Utk.Edu.

Labosky, P., Yobp, R. D., Janowiak, J. J., and Blankenhorn, P. R. (1993). Effect of Steam Pressure Refining and Resin Levels on the Properties of Uf-Bonded Red Maple MDF. *Forest Products Journal*, **43**(11-12), 82-88.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley-Interscience, Hoboken, N.J.

Lunneborg, C. E. (2000). *Data Analysis by Resampling: Concepts and Applications*. Duxbury, Australia; Pacific Grove, CA.

Maloney, T. M. (1993). *Modern Particleboard and Dry-Process Fiberboard Manufacturing*. Miller Freeman Inc., San Francisco, CA.

Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. Wiley, New York, NY.

Martinez, W. L. and Martinez, A. R. (2002). *Computational Statistics Handbook with MATLAB*. Chapman and Hall/CRC, Boca Raton, LA.

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley, New York, NY.

Meeker, W. Q. and Escobar, L. A. (2004). Reliability: The Other Dimension of Quality. *Quality Technology & Quantitative Management*, **1**(1), 1-25.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001). *Introduction to Linear Regression Analysis*. Wiley, New York, NY.

Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*. John Wiley and Sons, New York, NY.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996). *Applied Linear Regression Models*. Irwin, Chicago, IL.

O'Connor, P. D. T. (1985). *Practical Reliability Engineering, 2nd Ed.* Wiley, Chichester, Great Britain.

Ogawa, T. and Ohkoshi, M. (1997). Properties of Medium Density Fiberboards Produced from Thermoplasticized Wood Fibers by Allylation without Adhesives. *Mokuzai Gakkaishi*, **43**(1), 61-67.

Park, B. D., Riedl, B., Hsu, E. W., and Shields, J. (1998). Effects of Weight Average Molecular Mass of Phenol-Formaldehyde Adhesives on Medium Density Fiberboard Performance. *Holz Als Roh-Und Werkstoff*, **56**(3), 155-161.

Park, B. D., Riedl, B., Hsu, E. W., and Shields, J. (2001). Application of Cure-Accelerated Phenol-Formaldehyde (Pf) Adhesives for Three-Layer Medium Density Fiberboard (MDF) Manufacture. *Wood Science and Technology*, **35**(4), 311-323.

Parr, W. C. (1983). A Note on the Jackknife, the Bootstrap and the Delta Method Estimators of Bias and Variance. *Biometrika*, **70**(3), 719-722.

Parr, W. C. (1985a). The Bootstrap - Some Large Sample Theory and Connections with Robustness. *Statistics & Probability Letters*, **3**(2), 97-100.

Parr, W. C. (1985b). Jackknifing Differentiable Statistical Functionals. *Journal of the Royal Statistical Society Series B-Methodological*, **47**(1), 56-66.

Polansky, A. M. (1999). Upper Bounds on the True Coverage of Bootstrap Percentile Type Confidence Intervals. *The American Statistician*, **53**(4), 362-369.

Polansky A. M. (2000). Stabilizing Bootstrap-t Confidence Intervals for Small Samples. *The Canadian Journal of Statistics.* **28**(3), 501-526.

Rials, T. G., Kelley, S. S., and So, C. L. (2002). Use of Advanced Spectroscopic Techniques for Predicting the Mechanical Properties of Wood Composites. *Wood and Fiber Science*, **34**(3), 398-407.

Rowell, R. M., Youngquist, J. A., Rowell, J. S., and Hyatt, J. A. (1991). Dimensional Stability of Aspen Fiberboard Made from Acetylated Fiber. *Wood and Fiber Science*, **23**(4), 558-566.

Scott, D. W. (2003). The Case for Statistical Graphics. *AmStat News*, **315**, 20-22.

Seaver, W. (2004). Personal Communication. wseaver@utk.edu.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York, NY.

Suchsland, O. and Woodson, G. E. (1986). *Fiberboard Manufacturing Practices in the United States*. U.S. Department of Agriculture Forest Service's Agriculture Handbook No. 640, Washington, D.C.

Tsunoda, K., Watanabe, H., Fukuda, K., and Hagio, K. (2002). Effects of Zinc Borate on the Properties of Medium Density Fiberboard. *Forest Products Journal*, **52**(11-12), 62-65.

Urbanik, T. J. (1998). Strength Criterion for Corrugated Fiberboard under Long-Term Stress. *Tappi Journal*, **81**(3), 33-37.

Urmanov, A. M., Gribok, A. V., Bozdogan, H., Hines, J. W., and Uhrig, R. E. (2002). Information Complexity-Based Regularization Parameter Selection for Solution of Ill Conditioned Inverse Problems. *Inverse Problems*, **18**(2), L1-L9.

van Houts, J. H., Bhattacharyya, D., and Jayaraman, K. (2000). Determination of Residual Stresses in Medium Density Fibreboard. *Holzforschung*, **54**, 176-182.

van Houts, J. H., Bhattacharyya, D., and Jayaraman, K. (2001a). Reduction of Residual Stresses in Medium Density Fibreboard, Part 1. Taguchi Analysis. *Holzforschung*, **55**, 67-72.

van Houts, J. H., Bhattacharyya, D., and Jayaraman, K. (2001b). Reduction of Residual Stresses in Medium Density Fibreboard, Part 2. Effects on Thickness Swell and Other Properties. *Holzforschung*, **55**, 73-81.

van Houts, J. H., Winistorfer, P. M., and Wang, S. Q. (2003). Improving Dimensional Stability by Acetylation of Discrete Layers within Flakeboard. *Forest Products Journal*, **53**(1), 82-88.

Walker, E. and Guess, F. M. (2003). Comparing Reliabilities of the Strength of Two Container Designs: A Case Study. *Journal of Data Science*, **1**, 185-197.

Wang, S. and Winistorfer, P. M. (2000a). Fundamentals of Vertical Density Profiles Formation in Wood Composites. Part 4. Methodology of Vertical Density Formation under Dynamic Condition. *Wood and Fiber Science*, **32**(2), 220-238.

Wang, S., Winistorfer, P. M., and Young, T. M. (2004). Fundamentals of Vertical Density Profile Formation in Wood Composites. Part Iii. MDF Density Formation During Hot-Pressing. *Wood and Fiber Science*, **36**(1), 17-25.

Wang, S., Winistorfer, P. M., Young, T. M., and Helton, C. (2001). Step-Closing Pressing of Medium Density Fiberboard - Part 2. Influences on Panel Performance and Layer Characteristics. *Holz Als Roh-Und Werkstoff*, **59**(5), 311-318.

Wang, S. Y., Chen, T. Y., and Fann, J. D. (1999). Comparison of Internal Bond Strength and Compression Shear Strength of Wood-Based Materials. *Journal of Wood Science*, **45**(5), 396-401.

Weibull, W. (1939). A Statistical Theory of the Strengths of Materials. *Ing. Vetenskaps Akad. Handll*, **151**(1), 1-45.

Weibull, W. (1951). A Statistical Distribution Function of Wide Applicability. *Journal of Applied Mechanics*, **18**(1), 293-297.

Widsten, P., Laine, J. E., Tuominen, S., and Qvintus-Leino, P. (2003). Effect of High Defibration Temperature on the Properties of Medium-Density Fiberboard (MDF) Made from Laccase-Treated Hardwood Fibers. *Journal of Adhesion Science and Technology*, **17**(1), 67-78.

Williams, T. N. (2001). A Modified Six Sigma Approach to Improving the Quality of Hardwood Flooring. *Department of Forestry, Wildlife, and Fisheries*, University of Tennessee, Knoxville, 190.

Xu, W. and Winistorfer, P. M. (1995). Layer Thickness Swell and Layer Internal Bond of Medium Density Fiberboard and Oriented Strandboard. *Forest Products Journal*, **45**(10), 67-71.

Xu, W., Winistorfer, P. M., and Moschler, W. W. (1996). A Procedure to Determine Water Absorption Distribution in Wood Composite Panels. *Wood and Fiber Science*, **28**(3), 286-294.

Young, T. M. and Guess, F. M. (2002). Mining Information in Automated Relational Databases for Improving Reliability in Forest Products Manufacturing. *International Journal of Reliability and Applications*, **3**(4), 155-164.

Young, T. M. and Winistorfer, P. M. (1999). Statistical Process Control and the Forest Products Industry. *Forest Products Journal*, **49**(3), 10-17.

Young, T. M. and Winistorfer, P. M. (2001). The Effects of Autocorrelation on Real-Time Statistical Process Control with Solutions for Forest Products Manufacturers. *Forest Products Journal*, **51**(11-12), 70-77.

Yusuf, S., Imamura, Y., Takahashi, M., and Minato, K. (1995). Property Enhancement of Albizzia Waferboard by formaldehyde Treatment. *Mokuzai Gakkaishi*, **41**(2), 223-228.

Ziegel, E. (2003). Top Five Books for Statisticians. *AmStat News*, **315**, 25.

Vita

David J. Edwards is a graduate research assistant in statistics for the Tennessee Forest Products Center at the University of Tennessee. He received a B.S. in Mathematics with a minor in Statistics from Virginia Polytechnic Institute and State University and is currently pursuing an M.S. in Statistics from the University of Tennessee with plans to graduate in May 2004. David received two Hatcher Scholarships in Mathematics while attending Virginia Tech. He is a member of Golden Key International Honor Society, Phi Beta Kappa National Honor Society, Pi Mu Epsilon National Mathematics Honor Fraternity, and Alpha Phi Omega National Service Fraternity. David was a speaker at the 57th Annual Forest Products Society conference in Seattle, WA in June 2003. He presented further work at the Massachusetts Institute of Technology's International Conference on Information Quality in November 2003. He served as a Statistics Instructor for the Ronald McNair Post-Baccalaureate Achievement Program during summer 2003. After obtaining a Master's degree, David will pursue a Ph.D. in Statistics from Virginia Tech and intends to remain in an academic setting in order to teach and conduct research.