Spring 5-31-2014

Data analysis and simulation for warranties and golf handicaps

Sonia Bandha

New Jersey Institute of Technology

ABSTRACT

DATA ANALYSIS AND SIMULATION FOR WARRANTIES AND GOLF HANDICAPS

by
Sonia Bandha

In this dissertation, we discuss the application of data analysis and numerical simulation in order to gain insight into problems related to warranty cost management and the effectiveness of the golf handicap system. Despite the commonalities of the approaches, we will discuss these problems in turn.

For many moderately high value items with a substantial sales volume (such as automobiles), a warranty is used as an important element of marketing products as a better warranty typically signals higher product quality to customers. Much recent research on modeling and optimization of servicing costs for Non-Renewing Free Replacement Warranties (NR-FRW) assumes that the consumers' usage profile is known. Such an assumption is unrealistic for many consumer durables. In such cases, it would be pragmatic to assume that the usage rate should be modeled by a probability distribution. This research seeks to model and minimize the expected costs of servicing strategies for NR-FRW; this is accomplished using a numerical technique known as simulated annealing while considering a variety of usage rate distributions. The relationship between the usage rate distribution and product life-length is modeled using the Accelerated Failure Time (AFT) formulation. We use a copula based approach to capture the adverse impact of increasing product usage rate on its time-to-failure. We obtain a unique copula based on the marginal distributions of both the usage rate and the product life-length, which we call the AFT Copula. The underlying dependency of our copula is evaluated using non-parametric measures of association. The Mean Time to First Failure (MTTF) indicates which usage rate distributions most likely correspond to highly reliable products. We

found that certain warranty servicing strategies were more cost efficient than other commonly used approaches. We use data analysis techniques on a traction motor data set to study the practicality of our approach. The results obtained from this data are in qualitative agreement with our previous results.

The ability of a golfer is measured by a player's handicap which is an estimate of his/her potential based on previously played games. The handicap system is administered by the United States Golf Association (USGA); it is designed to enable players of differing abilities to compete against each other on an equitable basis. Most previous studies in golf have focused on analyzing golf scores. The goal of this research is to study the effectiveness of the current handicapping system. We use the AT&T Golf Tournament League data set for our study; this data set contains scores and handicaps of golfers from almost 100 different tournaments. In this study, we use data analysis methods including filtering to remove outliers and goodness of fit tests to determine the most appropriate distribution for the golf scores. Because each handicap requires a separate fit, we develop a technique of minimizing the average ranks of the candidate distributions in order to obtain the single best distribution for all handicaps. For this data set, the generalized extreme value distribution is the most appropriate. In order to investigate the effectiveness of the handicap system, we conduct simulations of competitions between golfers with varying handicaps based on the empirical and fitted data for golf scores. These simulations indicate that a player with a lower handicap has an advantage over a player with a higher handicap.

DATA ANALYSIS AND SIMULATION FOR WARRANTIES AND GOLF HANDICAPS

by
Sonia Bandha

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences
Department of Mathematics and Computer Science, Rutgers-Newark

May 2014

APPROVAL PAGE

DATA ANALYSIS AND SIMULATION FOR WARRANTIES AND GOLF HANDICAPS

Sonia Bandha

Dr. David J. Horntrop, Dissertation Co-Advisor — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Thomas Spencer III, Dissertation Co-Advisor — Date
Professor of School of Management, Walden University

Dr. Fadi Deek, Committee Member — Date
Provost and Senior Executive Vice President, NJIT

Dr. Wenge Guo, Committee Member — Date
Assistant Professor of Mathematical Sciences, NJIT

Dr. Aridaman K. Jain, Committee Member — Date
Professor of School of Management, Walden University

BIOGRAPHICAL SKETCH

Author: Sonia Bandha

Degree: Doctor of Philosophy

Date: May 2014

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
  New Jersey Institute of Technology, Newark, NJ, 2014
- Master of Science in Applied Mathematics,
  Ohio University, Athens, Ohio, 2008
- Master of Science in Mathematics,
  S. C. D. Government College, Punjab, India, 2005
- Bachelor of Science with Mathematics Honors,
  Government College for Women, Punjab, India, 2003

Major: Applied Statistics

Presentations and Publications:

S. Bandha, D. J. Horntrop, "Copula-based modeling and computational solutions to some warranty cost management problems", *The Dana Knox Student Research Showcase*, NJIT, April, 2014.

S. Bandha, D. J. Horntrop, "Computational solutions to some warranty cost management problems", *Statistics Seminar, Department of Mathematical Sciences*, NJIT, November, 2013.

S. Bandha, D. J. Horntrop, T. Spencer III, "Simulation of golf handicap scores via distribution of moving average order statistics", *The IXth Annual Graduate Student Research Day*, NJIT, October, 2013.

S. Bandha, D. J. Horntrop, T. Spencer III, "Simulating golf handicaps using empirical and fitted data", *New England Symposium of Sports in Statistics*, Harvard University, Cambridge, MA, September, 2013.

S. Bandha, D. J. Horntrop, T. Spencer III, "Golf handicap scores modeled via distribution of averages of moving order statistics", *Frontiers in Applied and Computational Mathematics*, NJIT, June, 2013.

S. Bandha, D. J. Horntrop, "Copula-based modeling and computational solutions to some warranty cost management problems", *IIE Annual Conference and Expo*, San Juan, Puerto Rico, May, 2013.

S. Bandha, D. J. Horntrop, "Application of simulations for warranty cost management problems", *Frontiers in Applied and Computational Mathematics*, NJIT, May, 2012.

I would like to dedicate my work to my parents
Mrs. Kamlesh Bandha and Mr. Mohan Lal Bandha,
my brothers and sisters-in-law
Manoj and Siddhi Bandha,
Sunil and Vani Bandha,
and
my nephew and niece
Girik & Rayna

ACKNOWLEDGMENTS

First of all, I want to express my deep gratitude to my Professor Manish C. Bhattacharjee (deceased) who initially supported me to begin my doctoral dissertation work. I wish to specially thank my advisor, Professor David J. Horntrop for his guidance throughout this research by giving me the confidence that I could be good at this field of statistics, and for really challenging me every step of the way to leave no stone unturned, and to approach any problem from multiple angles. From beginning to end, he has been a steadfast source of information, ideas, support, and energy. I am indebted to him for his patience and guidance during discussions and detail examination of this manuscript. His profound knowledge and experience have strengthened this dissertation in many ways.

I am very much grateful to Professor Thomas Spencer for getting me interested in the research related to golf. I highly appreciate his support, encouragement, and guidance throughout this project. He spent his valuable time in discussing various possible solutions related to the problems in the project. His insightful discussions and comments greatly shaped the project.

I wish to thank my dissertation committee members, Professor Aridaman K. Jain, Professor Fadi Deek and Professor Wenge Guo for their valuable suggestions, involvement and willingness to be present in my committee.

I wish to express my sincere thanks to Professor B. P. Iskandar (Professor, Department of Industrial Engineering, Faculty of Industrial Technology, Bandung Institute of Technology, Indonesia) for his valuable time to provide insightful thoughts about the warranty related project. I would like to extend my thanks to Professor Bharatendra K. Rai (Chair, Department of Decision and Information Sciences Charlton College of Business, University of Massachusetts, Dartmouth) and Steven

Swiss (President of AT&T Golf Tournament League 2004-2005) for sharing with me the data that helped me greatly in my research.

I want to express my gratitude to Professor Jonathan Luke and Professor John K. Bechtold who have been very helpful as the chairmen of the department during my doctoral program. I would also like to thank Eileen Michie and Susan Sutton for being extraordinarily understanding and supportive in every possible way to make my stay pleasant at NJIT. I wish to extend my appreciation to the faculty, staff and fellow graduate students of the Department of Mathematical Sciences as a whole, whom I have known during my stay at NJIT.

I want to specially thank brother Manoj Bandha and my sister-in-law Siddhi Bandha for always motivating and supporting me. Their selfless love and sacrifices have been the greatest encouragement throughout this doctoral program. I am highly appreciative of them allowing me to stay with them through out my doctoral program. I am blessed by having them in my life.

I owe my deepest gratitude to my parents, Kamlesh Bandha and Mohan Lal Bandha whose love, blessings, and innumerable sacrifices have given me the strength to believe in myself to accomplish my dream. I would also thank my brother Sunil Bandha and sister-in-law Vani Bandha for their love and support from miles away.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF TABLES
(Continued)

Table **Page**

LIST OF TABLES
(Continued)

LIST OF TABLES
(Continued)

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

Data analysis includes many techniques that help describe facts, detect patterns, develop explanations, and test hypotheses. Simulation involves the use of the computer to shed insight into problems for which analytic solutions are unavailable. Both of these modes of inquiry are applicable in many areas. In this dissertation, these approaches are used to study warranties and golf handicaps; despite the great differences in these applications, many of the approaches used are common. The first five chapters are related to warranties and last five are related to golf handicaps.

In Chapter 2, we briefly introduce the concept of warranties, present an overview of the classification of warranties, discuss the types of repairs that a manufacturer may use to remedy failures, and give a literature review about previous research on warranties. Chapter 3 deals with the development of the repair strategy and the formulation of the two-dimensional cost model. This chapter also highlights how the warranty cost is sensitive to the usage rate with the help of copulas and the Mean Time to First Failure (MTTF). Chapter 4 summarizes the simulation techniques. The results are obtained for different servicing scenarios with an overall objective of minimizing the warranty cost. Chapter 5 involves the use of data analysis to study the traction motor data set to verify the practicality of our cost model. Chapter 6 provides the summary of the results and describes future work in optimizing warranty costs.

In Chapter 7, we begin the portion of the dissertation on the golf handicap system with a discussion of the golf handicap system along with the basic terminologies used to calculate the handicap of a golfer. This chapter also includes a discussion on application of order statistics to calculate the handicap index. Chapter 8 uses the AT&T Golf Tournament League data set to analyze the golf scores. Chapter 9 describes a validation of the simulation approach based upon the data to observe the behavior of the handicap indices. In Chapter 10, we conduct simulations of competitions between golfers of differing handicaps based on the empirical and the fitted data. Chapter 11 includes the summary of the simulations and describes future work in analyzing golf handicaps.

CHAPTER 2

AN OVERVIEW OF WARRANTIES

In the automotive industry, modern manufacturing techniques and fierce competition in the market have posed serious challenges to the manufacturer in terms of differentiating its product from the competition. As a result, a warranty is used as a tool to attract consumers in order to increase sales and revenue. Offering a warranty results in additional costs to the manufacturer from the servicing of the claims made by the consumers. In order to understand the concept of warranty, this chapter highlights the importance of warranty from the viewpoint of both the manufacturer and consumer. Along with this, different kinds of warranties and a brief literature overview of strategies for servicing warranties will be discussed.

2.1 Introduction to Warranties

A warranty is a contractual obligation that serves as a manufacturer's assurance to a buyer of adequate product performance by requiring the manufacturer to rectify the failures that may occur while the warranty coverage is in effect [1]. An item is said to have failed if the item is not functioning satisfactorily during the given period of time of the contract. In broad terms, the purpose of a warranty is to establish the liability of the manufacturer that the product will perform the intended function for a specified period of time. Consumers benefit by receiving assurance that any kind of failure will be rectified [2]. Manufactures gain an effective promotional tool to sell their product; they can differentiate their product by offering longer warranties versus competitor selling a nearly identical product [2]. The warranty policy can provide information regarding the product quality and reliability.

2.1.1 Role of a Warranty

Warranties serve as an integral part of nearly all commercial and government transactions involving purchase of products [1]. Rapidly developing new technologies, fierce competition and exposure to a global marketplace are the characteristics of modern manufacturing and sales, together with savvy and demanding consumers. A consumer (customer/buyer) typically views a product in terms of its attributes, which represents 'a complex cluster of value satisfactions' [2], which are not always easy to quantify. From a producer's (manufacturer/seller) viewpoint, offering a warranty has cost implications that are additional to the costs of manufacturing, promotion, marketing and other related costs that are accepted as standard business practices. Since the costs of a base warranty are typically factored into the sale price as part of the initial transaction; the manufacturer, for obvious reasons, has an incentive to keep the costs of servicing warranties down.

The costs associated with offering a warranty are influenced by several factors such as design reliability, the actual terms of the warranty, the usage intensity of the item by the consumer, operating environment of the item and the logistics of servicing the item; of these factors, only the design reliability of the product is exclusively controlled by the manufacturer [3]. The reliability of a product depends on decisions made during design and production phases, operating environment, usage mode and intensity, and so forth [2, 3]. Statistically, it is defined as the probability that the product is in function for a specified time period when operating under stated conditions. Due to variations in the manufacturing, inherent and design reliability can differ. For example, defective bearings in a washing machine can result in early failure. Since the occurrence of failures is uncertain, it is critical to build up a model to solve the problem of predicting failures over time. The kind of mathematical formulation to be used depends on the system characterization and type of approach used. Therefore, modeling product failures requires probability distribution functions to model time to first failure and counting processes to model subsequent processes.

2.1.2 Standard Nomenclature Used in a Warranty

When the performance of a product fails below the specified level, it is then considered to have failed. The actions taken to rectify failures include either a replacement or various types of repairs. Some of the commonly used terminologies in the warranty research are discussed below.

Replacement/Perfect Repair: Under this strategy [2], the failed product is replaced by an identical product from manufacturer's stock of new products. This type of rectification action has the advantage of restoring the product to its initial stage. The corresponding rate of degradation of the product resulting from a replacement is the same as for a new product.

Minimal Repair: Under this strategy [2], the condition of a product is restored to its corresponding condition just prior to failure. Such repairs are often called "bad - as - old", while replacements are called "good - as - new" repairs.

Imperfect Repair: Under this strategy [2], a failed equipment is restored to a condition intermediate between that achieved by a minimal repair and a replacement.

2.1.3 Classification of Warranties

We will discuss list of some standard classifications that are used in the warranty literature.

One-Dimensional (1-D) and Two-Dimensional Warranties (2-D)

One-dimensional warranties [2] are characterized by an interval defined in terms of a single variable which is most typically time or age. Two-dimensional warranties [2] are categorized by time or age and usage.

Renewing and Non-Renewing Warranties

A warranty is renewing [2] if the warranty begins again when a failed item is replaced or otherwise repaired. A warranty is non-renewing [2] if repairs or replacement have no impact on the warranty period. A warranty cycle for a renewing warranty starts from the time of sale and is a random variable whose value depends on the warranty period, the total number of failures under the warranty, and the actual failure times. The majority of warranties are non-renewable in which the warranty cycle, which is the same as the warranty period, is not random, but predetermined since the warranty obligation will be terminated as soon as warranty period of time passes after the sale.

Free-Replacement Warranties (FRW)

Under free-replacement warranty, the manufacturer provides all repairs or replacements at no charge. These sorts of warranties are usually combined with renewing and non-renewing warranties in 1-D or 2-D. For instance, under a Non-Renewing Free-Replacement Warranties (NR-FRW) [2] the manufacturer agrees to repair or provide replacements for failed items free of charge up to the end of the original warranty period. In comparison, under Renewing Free-Replacements warranties, the manufacturer agrees to repair or provide a replacement free of charge along with a new warranty under original terms and conditions.

Pro-Rated Warranties (PRW)

Pro-rated warranties [4, 5] refer to warranty policies in which the cost of servicing of failed items under warranty is shared between the manufacturer and the consumer. Such warranty policies charge a fixed percentage of warranty costs to the consumers and the remaining percentage is incurred by the warranty provider. Pro-rated warranties are therefore relatively more 'manufacturer friendly'.

Group/Fleet Warranties

These refer to warranties that cover a group of similar items as a whole, which are typically large commercial/governmental transactions (e.g., purchase of a fleet of vehicles or aircraft) [4, 5]. A buyer's preference for these warranties is based on the convenience of having the entire fleet covered as a single unit. For the same reason, a manufacturer may expect fewer claims since long lasting units in the group can offset the cost of early failures. Group warranties can be thought of as extensions of non-renewing FRW and PRW warranties covering multiple items with umbrella coverage of all items in the group.

Extended Warranties

These are service contracts entered into between a buyer and a seller, with coverage beginning when the base warranty expires [4, 5]. Such a contract is effective until a later time than the original expiration time of the warranty, can be of any specified duration, and is purchased separately from the item. Here are many variations for the terms offered in practice by an extended warranty; extended warranties can be identical to the base warranty or may contain other provisions (e.g. exclusions of labor costs, cost sharing for parts etc.). Nowadays extended warranties are offered on a wide range of products, including automobiles, electronics, appliances, and many other items.

Reliability Improvement Warranties (RIW)

These extend a base warranty to include guarantees on product reliability as an additional component of the corresponding contract [4, 5]. Such warranties are meaningful for complex repairable systems intended for repeated or relatively long period of use, or for systems which are mission critical where safety is paramount (e.g., space shuttles). A RIW contract often includes a guaranteed and demonstrated "mean time to failure" requirement with the contractors' fees based on their ability to deliver the corresponding reliability.

2.2 Literature Review

The goal of our research is to model warranty costs from the manufacture's point of view. The costs associated with the warranty servicing are unpredictable since claims and the cost to rectify the failed items are uncertain. Different types of warranty policies are in common use for effective management of product warranty. A variety of mathematical models have been developed and discussed in the literature to study various aspects of warranties.

2.2.1 Literature Review of Warranty Servicing Cost

The first definitive reference handbook covering a variety of warranty policies and mathematical models of various related engineering and management issues, was published by Blischke and Murthy [5]. Blischke [6] provided a survey of cost models which were proposed for different warranty situations. Blischke and Murthy presented a three part paper review as a taxonomy for warranty policies ([7], [8] and [9]). The first paper [7] proposed a taxonomy for warranty to assist the manager responsible for product warranty in choosing appropriate alternatives for evaluation before a final choice is made. The second review paper [8] presented an alternate forms of system characterization for the study of warranties from three different perspectives, those of the consumer, the manufacturer and the public policy decision-maker. The third review paper [9] carried out a comprehensive review of these models by classifying them into three categories based on consumer, manufacturer and public policy decision maker perspectives. Murthy and Djamaludin [1] provided the next comprehensive literature review of warranties and highlighted issues of interest to manufacturers in the context of managing new products from an overall business perspective. Historically the works and reviews referenced above are useful in a practical sense; however, much of these have been recently supplemented by the work of Blischke, Karim and Murthy [2] on warranty data collection and analysis. The warranty data collections book [2] covers additional topics such as product reliability assessment methods as well as warranty claims data collection and their statistical analysis and is more encyclopedic in scope relative to our research. Accordingly, the rest of the literature review will focus on research literature on controlling expected warranty servicing costs from the manufacturer's point of view.

Biedenweg [10] proved that for non-renewing free replacement 1-D warranties with rectification options as minimal repair or replacement, the optimal strategy replaces a failed product with a new item for all failures occurring up to a certain clock-time measured from the time of initial purchase, and rectifies all subsequent failures by appropriate repair until the warranty expires. This idea of partitioning the warranty period into two intervals corresponding to replacement period and a repair period respectively was then fruitfully exploited and generalized by adding a third interval to the partition (Nguyen and Murthy [11]) in which failures occurring in the middle interval are rectified using a stock of used items. The three-interval partition idea was also used by the same authors (Nguyen and Murthy [12]) to extend the original two-interval model of Biedenweg by adding a third interval (middle interval) where failed units are either replaced or minimally repaired and a new warranty is given after every such repair (a variation of renewing warranty). In a subsequent paper, Jack and Van der Schouten [13], conjectured the structure of the optimal (i.e., expected total cost minimizing) warranty servicing strategy under the set up first proposed by Nguyen and Murthy (*ibid.*). Their conjecture was indeed proved several years later by Jiang, Jardine and Lugitigheid [14].

Jack and Murthy [15] further pursued the cost implications of a strategy where only minimal repairs are carried out for failures in the initial and terminal intervals, with at most one replacement in the middle interval. It may be noted here that while the optimal 1-D strategy as shown by Jiang et al.(*ibid.*) requires continuous monitoring and is not easy to implement in practice; Jack and Murthy's idea of pursuing the three interval strategy without a requirement of continuous monitoring of the working unit is a pragmatic option. Its objective is to optimize the choice of how and where to locate the boundaries of the middle interval. Although the resulting cost minimizing policy is, strictly speaking, theoretically sub-optimal among all possible warranty strategies; it may be nearly optimal under suitable configurations of repair costs, warranty length, and life length parameters - a possibility that has not been sufficiently explored. In any case, the corresponding three interval warranty servicing strategy is well understood by the producers due to its simplicity and is also easily explained to buyers; features that go a long way to earn acceptability by both producers and consumers. Indeed, as recent literature suggests; by virtue of its pragmatism and simplicity, the modeling and analysis of cost implications of three-interval based strategies have become a benchmark for measuring the cost-effectiveness of such strategies with different possible repair options.

An article by Iskander et al. [17] illustrated a new warranty servicing strategy for items sold with two-dimensional warranty where the failed item is replaced by a new one when it fails for the first time in a specified region of the warranty and all other failures are repaired minimally. Jack et al. [18] provided a repair-replace strategy for the manufacturer of a product sold with a two-dimensional warranty. The strategy is based on a specified region of the warranty defined in terms of age and usage with the first failure in the region rectified by replacement and all other failures being minimally repaired. Yun et al. [19] investigated a corresponding model for 1-D warranties which are sensitive to the usage rate via a strategy that allows for various 'degrees of repair' including minimal repair and replacement as the two extremes for rectifying the first failure in the middle interval. A distinguishing feature of their approach is to allow the cost of an 'imperfect repair' to depend on the degree of repair undertaken.

The majority of research literature on modeling and optimization of warranty servicing costs for non-renewing free warranties (NR - FRW) assume a constant and known usage rate of the product by a customer. Such an assumption is unrealistic for many moderately high-value products with substantial sales volumes (such as new automobiles) where a customer's use profile of the product may at best be statistically known in terms of a usage rate distribution. Relatively little attention appears to have been paid to investigate the cost of warranties from this aspect, with few exceptions to the best of our knowledge. Jack et al. [18] mentions such a modeling possibility as a passing remark without pursuing it any further. Chukova and Johnston [20] have considered this modeling approach more explicitly, with the usage rate constant for each customer. Chukova and Varnosafaderani [21] modeled warranty costs with a probabilistically specified usage rate.

2.2.2 Warranty Cost Analysis

The first step in analyzing the cost of a warranty to a manufacturer is to model product failures and associated costs either at component or product level [1]. In 1-D warranties, failures are modeled via a probability distribution which depends on the nature of rectification action. For failures rectified with replacements, the formulation follows a renewal process with replacement times being negligible. If all failures are minimally repaired with repair times negligible then the formulation is a non-stationary Poisson process with specified intensity function. In comparison to the 1-D case, 2-D warranties have received less attention in terms of modeling failures. In this case, failures are points in a plane with age plotted against usage. One approach in 2-D uses a two-dimensional distribution function to model failures; see Iskandar [22] and Murthy et al. [23]. Another approach involves modeling usage as a function of time so that failures are effectively modeled by one-dimensional point process formulation [1]. Iskandar [22] used such an approach by suggesting a linear model relating age and time.

In this dissertation, we have considered an alternative formulation for modeling and analysis of warranty servicing costs: a consumer's usage rate is assumed to be constant, but unknown to the producer/seller of the product except via its marginal distribution. For any specific consumer, the assumption of a constant usage rate is not unrealistic at least up to a first approximation. Although the manufacturer may not know and possibly can not ascertain a perspective consumer's usage rate, sufficient past data on warranty claims can enable the manufacturers to construct statistically valid models of the product's usage rate and times to failure.

2.3 Concluding Remarks

Reducing warranty costs is an issue of great interest to the manufacturers. In this chapter, we have discussed the overview of warranties along with strategies adopted previously to address the claims to rectify the failures. Next, we will discuss the strategy adopted in this dissertation and evaluate the warranty servicing cost model as a result of rectification of failures.

CHAPTER 3

WARRANTY MODEL FORMULATION

When a repairable item fails under warranty, the manufacturer has the option of repairing it or replacing it with a new one. This chapter focusses on an approach to model failures for the case of products sold with 2-D non-renewable free replacement warranty (NR-FRW), given that usage rate of the customer is unknown to the manufacturer. The failure pattern is influenced by the design of the product, the manufacturing process, the intensity of usage by the consumer, as well as several other uncontrollable factors. Since the failure patterns can not be determined precisely, we will construct an appropriate probability model. Using this new approach to model failures and claims under warranty, the overall expected warranty cost model is evaluated.

3.1 Modeling Failures

For 2-D warranties, the effects of both age and usage play an important role in the costs associated with the product degradation and failure. The usage can be the output of the device (e.g., number of pages printed by photocopier), distance traveled (e.g., miles covered by an automobile), or the amount of time the product has been used (e.g., the number of hours vacuum cleaner has been used). We confine our research to a two-dimensional warranty which expires when the item reaches an age W or reaches a usage level, U whichever comes first. For example, the warranty offered by an automobile manufacturer for a new automobile is typically 3 years and 36,000 miles. The two-dimensional (2-D) warranty offered at the time of sale is defined by the rectangular warranty region [2] given by

$$\Omega = [0, W) \times [0, U) \tag{3.1}$$

Failures in two-dimensional warranties are random points scattered over the defined warranty region. Hence, the cost analysis of these failures involves intricate calculations. For a customer with a constant usage rate y, the effective warranty period, is

$$W^e = min(W, U/y) \tag{3.2}$$

Figure 3.1 depicts two common cases for warranties expiration. A consumer with a low usage rate has warranty expiring at the end of the warranties specified time period W. On the other hand, a consumer with a relative high usage rate has warranty expiration at $U/Y < W$ [18].

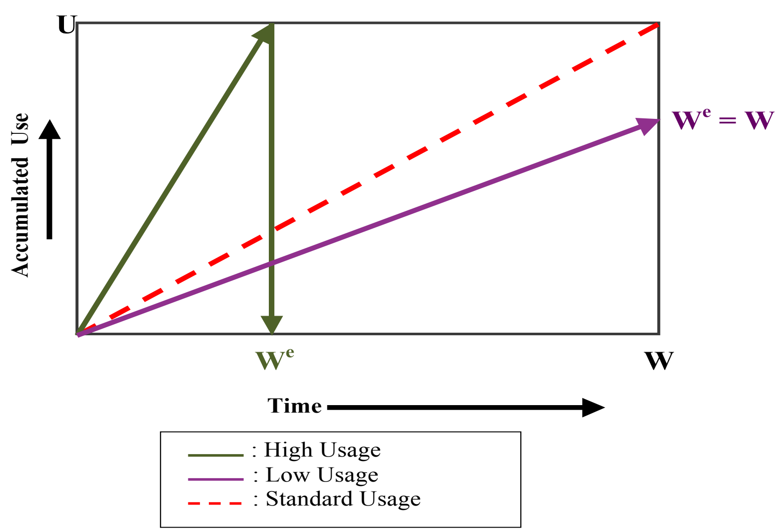


Figure 3.1 Effective warranty region for a given usage rate y.

3.1.1 Statistical Functions to Model Failures

We now introduce some of the statistical functions needed to model failures [2]. Let T be a continuous random variable denoting the time to failure of an item with corresponding distribution function $F(t)$, which is the probability that the item fails before time t. $f(t) = F'(t)$ is the failure probability density function. The survival (or reliability) function $\bar{F}(t)$ ($\bar{F}(t) = 1 - F(t)$) is used to evaluate the probability that the item survives at least a given period of time t. The hazard function $h(t) = f(t)/\bar{F}(t)$ characterizes the effect of age on item failure more explicitly than $F(t)$ or $f(t)$. At last, $H(t) = \int_0^t \mathrm{h(t')}\,\mathrm{d}t'$ is the cumulative hazard function, also known as cumulative failure rate function.

The approach to modeling failures assumes that the usage rate, denoted by Y, varies from customer to customer, is unknown, but is constant for a given customer. Therefore, Y is a random variable that can be modeled using a probability density function. Let $g(y)$ denote the density function of usage rate. For a given usage rate $Y = y$, the total usage of an item with an age x is given by u. The total usage is then expressed as

$$u = yx, \qquad 0 \leq u < \infty \tag{3.3}$$

The conditional hazard function $h(x|y)$ is the function of the time to first failure for given the usage rate. The failures over time are modeled via a counting process [2, 18] which is characterized via a conditional intensity function $\lambda_y(x)$ which is a non-decreasing function of x and y. For minimal repairs, $\lambda_y(x) = h(x|y)$ [2, 18]. If failed items are replaced by new ones, then this counting process is a renewal process associated with conditional distribution $F(x|y)$.

3.1.2 Modeling Product Failures

The usage intensity of a product varies across the consumer population. During the design stage of the product, decisions are made about component reliabilities in order to ensure that the product has the desired reliability at some nominal usage rate [18]. When the actual usage rate differs from this nominal value, the reliabilities of the components are affected which in turn affects the overall product reliability. With an increase in usage rate, the rate of product degradation increases which accelerates the time to failure. Consequently, an increase in usage rate leads to decrease in product reliability.

The Accelerated Failure Time (AFT) model (Lawless et al. [24]) can be used to model the effect of usage rate on reliability. Let y_0 be the nominal usage rate associated with time to first failure T_0. As the usage rate changes to y, the time to first failure becomes T_y. The following expression links the relationship between product life length and usage rate:

$$\frac{T_y}{T_0} = \left(\frac{y_0}{y}\right)^{\gamma} \tag{3.4}$$

where γ is an acceleration parameter such that $\gamma \geq 1$.

The associated probability functions are now obtained. For initial time to failure T_0 and scale parameter α_0, the associated conditional cumulative distribution function is given by $F_0(x|\alpha_0)$. Effectively, for time to failure T_y, the scale parameter $\alpha(y)$ and associated $F(x|\alpha(y))$ are expressed as:

$$\alpha(y) = \left(\frac{y_0}{y}\right)^{\gamma} \alpha_0 \tag{3.5}$$

and

$$F(x|\alpha(y)) = F_0\left(\left(\frac{y_0}{y}\right)^{\gamma} x \,\middle|\, \alpha_0\right). \tag{3.6}$$

For the warranty cost model formulation (Jack et al. [18]) we will choose the values: $y_0 = 1$, $\alpha_0 = 1$ and $\gamma = 2$. Using the AFT formulation, (3.4) depicts the existence of an inverse relationship between time to failure and usage rate. In order to numerically compute the association between the two, we will use a Copula approach [25] that links the relationship between usage rate and effective life length through a probabilistic approach and non-parametric measures of association. An advantage of using this approach is that it allows the dependence relationship to be studied by separating the marginal effects.

3.2 Copula Formulation

From the manufacturer's viewpoint, for the case of an unknown usage rate of customers, the warranty servicing cost must be averaged over the usage rate profile of the target group of customers. It is here that copulas play a significant role in capturing the impact of usage rate on the product life length for modeling warranty costs. Below, we briefly summarize those salient facts and basic results for bivariate copulas that will be required.

3.2.1 Introduction to Copulas

Copulas provide a convenient way to express joint distributions of two or more random variables. The joint distribution can be separated into two contributions: the marginal distributions of each variable by itself, and the copula that combines these into a joint distribution. As per Nelsen [25], copulas are usually defined on the unit square I^2, where $I = [0, 1]$.

Definition A two-dimensional function C from I^2 to I is called a copula [25] if it satisfies following properties:

1. For every $u, v \,\epsilon\, I$,

$$C(u, 0) = C(0, v) = 0;\ C(u, 1) = u \text{ and } C(1, v) = v$$

2. For every u_1, u_2, v_1, $v_2 \,\epsilon\, I$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_{2,}v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

The next theorem plays the main role in the theory of copulas that describes the relationship between the joint distribution function and the corresponding copula.

Theorem 3.2.1. (Sklar's Theorem) *Let S be a joint distribution function with margins F and G. Then, there exists a copula C such that for all x, y,*

$$S(x, y) = C(F(x), G(y)) \tag{3.7}$$

If F and G are continuous, then C is unique. Otherwise, the copula C is uniquely determined on $Range(F) \times Range(G)$. Conversely, if C is a copula and F and G are distribution functions, then the function S defined above, is a joint distribution function with margins F and G.

In virtue of Sklar's theorem [25], (3.7) can be replicated for the product life length X and usage rate Y with marginal distributions $F_X(x)$ and $G_Y(y)$. The joint distribution S of x and y, is expressed as

$$S(x, y) = C(F_X(x), G_Y(y)) \tag{3.8}$$

3.2.2 Copulas and Association

Dependence properties and measures of association are interrelated. The latter numerically quantify the extent of dependence between the components of a random vector. Since the copula corresponding to a joint distribution describes its dependence structure, it might be appropriate to use measures of dependence which are copula-based, so called measures of concordance. Among the most widely known scale-invariant measures of association for copulas are the population versions of Kendall's tau (τ) and Spearman's rho (ϱ) that measure concordance [25, 26].

"Kendall's tau" (τ_{XY}) [25, 26] is the difference between the probabilities of concordance and discordance for two independent and identically distributed pairs (X_1, Y_1) and (X_2, Y_2) each with a common joint distribution S, i.e.,

$$\tau_{XY} = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}$$

These probabilities can be evaluated by integrating over the distribution of (X_2, Y_2). If C is the copula uniquely determined by (X, Y), then one can show (see Nelsen [25]) that Kendall's tau for (X, Y) is given as:

$$\tau_C = 4\left[\int_0^1 \int_0^1 C(u,v)\, dC(u,v)\right] - 1 \tag{3.9}$$

"Spearman's rho" (ϱ_{XY}) [25, 26] coefficient is associated with a pair (X, Y) defined as

$$\varrho_{XY} = P\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_3) < 0\},$$

where (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) are three independent random vectors with a common joint distribution function S. Hence ϱ_{XY} is the difference between the probabilities of concordance and discordance of the random vectors (X_1, Y_1), (X_2, Y_3). Note that, X_1 and Y_1 have a joint distribution S with an associated copula C whereas X_2 and Y_3 are independent and identically distributed uniform $(0, 1)$. In terms of a copula C, Spearman's rho(ϱ_C) is :

$$\varrho_C = 12\left[\int_0^1 \int_0^1 (C(u,v) - uv)\, du\, dv\right] \tag{3.10}$$

3.2.3 AFT Copula Family

Using the AFT-model formulation in combination with copula, we will now obtain a unique copula, which we will refer to as the "family of AFT Copulas". Such copulas are of interest since they allow flexibility in the underlying dependency structure, which can be particular summarized by Kendall's τ and Spearman's ϱ. The effect of usage rate on the product life length can be explained with the help of the copula. The measures of association can be evaluated once the AFT copula is obtained. As an illustration, we will derive the AFT copula assuming a uniform distribution of the usage rate and Weibull AFT lifetimes conditional on rate of use y. Similar AFT copulas can be obtained for different choices of distributions for usage rate such as gamma, normal, and Weibull distributions. Let x follow the Weibull failure time distribution conditional on usage rate y with β as the shape parameter. The corresponding distribution functions are:

$$\overline{F}(x|y) = \exp(-(x(y^{\gamma}))^{\beta}),$$

$$f(x|y) = -\frac{d\overline{F}}{dx}(x|y) = \beta y^{\gamma\beta} x^{\beta-1} \exp(-(x(y^{\gamma}))^{\beta}).$$

Hence, the joint probability density (of time to failure X and usage rate Y) is:

$$\begin{aligned} s(x,y) &= g(y) f(x|y) \\ &= \beta y^{\gamma\beta} x^{\beta-1} \exp(-(x(y^{\gamma}))^{\beta})\, g(y) \qquad x > 0,\ y > 0 \end{aligned} \tag{3.11}$$

The joint cumulative distribution function of (X, Y) is then evaluated as:

$$\begin{aligned} S(x,y) &= \int_0^x \int_0^y s(t,r)\, dr\, dt \\ &= \beta \int_0^x t^{\beta-1} \left\{ \int_0^y \exp(-(t(r^{\gamma}))^{\beta}) r^{\gamma\beta} g(r)\, dr \right\} dt, \end{aligned} \tag{3.12}$$

with marginal cumulative distribution function of time to failure X, given by

$$F(x) = P(X \leq x, Y < \infty) = S(x, \infty).$$

Hence, the copula that uniquely corresponds to $S(x, y)$ using Sklar's theorem is

$$C(u,v) = S(F^{-1}(u), G^{-1}(v)) \tag{3.13}$$

In particular, for uniform distribution on (a, b) for the usage rate, we get

$$C(u,v) = S(F^{-1}(u), a + (b-a)v),$$

where $x = F^{-1}(u)$, $0 < u < 1$, is the unique solution of $u = S(x, b)$, such that

$$\begin{aligned}
u &= \frac{\beta}{b-a}\int_0^x t^{\beta-1}\int_a^b \exp(-(t(r^\gamma))^\beta) r^{\gamma\beta}\,dr\,dt \\
&= \frac{1}{b-a}\int_a^b\int_0^x t^{\beta-1}\exp(-(t(r^\gamma))^\beta)\,dt\,r^{\gamma\beta}\,dr \\
&= \frac{1}{b-a}\int_a^b\int_0^{(x(r^\gamma))^\beta} \exp(-z)\,dz\,dr \qquad (\text{setting } (x(r^\gamma))^\beta = z) \\
&= \frac{1}{b-a}\int_a^b (1-\exp(-(x(r^\gamma))^\beta))\,dr \qquad (3.14)
\end{aligned}$$

The final form of the AFT copula is

$$C(u,v) = \frac{\beta}{b-a}\int_0^x \int_a^{a+(b-a)v} \exp(-(t(r^\gamma))^\beta) t^{\beta-1} r^{\gamma\beta}\,dr\,dt \qquad (3.15)$$

where $x = F^{-1}(u)$. Clearly, both $F^{-1}(u)$ and $C(u, v)$ can be evaluated using numerical methods, which will be discussed in later.

3.3 Warranty Servicing Strategy

The statistical functions and formula shown in Sections 3.1 and 3.2 are now used to form the cost model. Modeling of expected costs in servicing warranties and later minimizing these costs has been the main focus of research from manufacturer's perspective. While the problem of optimizing expected warranty costs can be formulated as a finite horizon dynamic programming problem in continuous time; typically the path to finding an optimal strategy exploits the specific features of the problem at hand. In the warranty context, reasonably realistic models of product failures and repair options typically make exact analytical answers unobtainable and thus dictate the need for computational solutions in a discretized search space. In this process, we must deal with the so called 'curse of dimensionality' in dynamic programming that refers to increasing computational complexity which grows exponentially with the dimension of the state variable, the number of available (repair) options, and the relatively coarseness of the grid.

3.3.1 2-D Warranty Servicing Strategy

When a repairable product fails under warranty, the manufacturer has the option of either repairing the failed item or replacing it by a new one. The first option costs less than the second but the repaired item has a greater probability of failing during the remainder of the warranty period. Therefore, it is important for the manufacturer to choose an appropriate servicing strategy in order to minimize the expected cost of servicing the warranty for each product sold.

The repair-replace strategy used here is similar to that used by Jack et al. [18] which was initially introduced by Jack and Murthy [15]. According to this strategy, the total effective duration of warranty period is split into three disjoint intervals $[0, K)$, $[K, L)$, $[L, W^e)$ starting from the time of initial purchase, where $0 < K < L < W^e$, with W^e denoting the clock-time indicating the expiry of warranty.

The intuitive justification of the three-interval based strategy can be graphically shown in Figure 3.2. The initial time interval $[0, K)$ can be thought of as the time period when a product is relatively young and relatively less prone to fail. Hence choosing the costliest option of a new replacement for failures in this interval when the product is “young” may not be justifiable. Similarly, for a product with degrading profile of performance, a functioning product during the terminal interval $[L, W^e)$ is “old” and has a high likelihood of failure; however, the end of the warranty period is also rather soon, there is a lack of justification for fixes that are costlier than minimal repairs. If the only failure remedies are replacements or minimal repairs, the case for undertaking minimal repairs only during $[0, K)$ and $[L, W^e)$ is quite clear. However, the time interval $[K, L)$ in the middle can be interpreted as time period when the product is neither too young nor too old. In the previous studies, a replacement

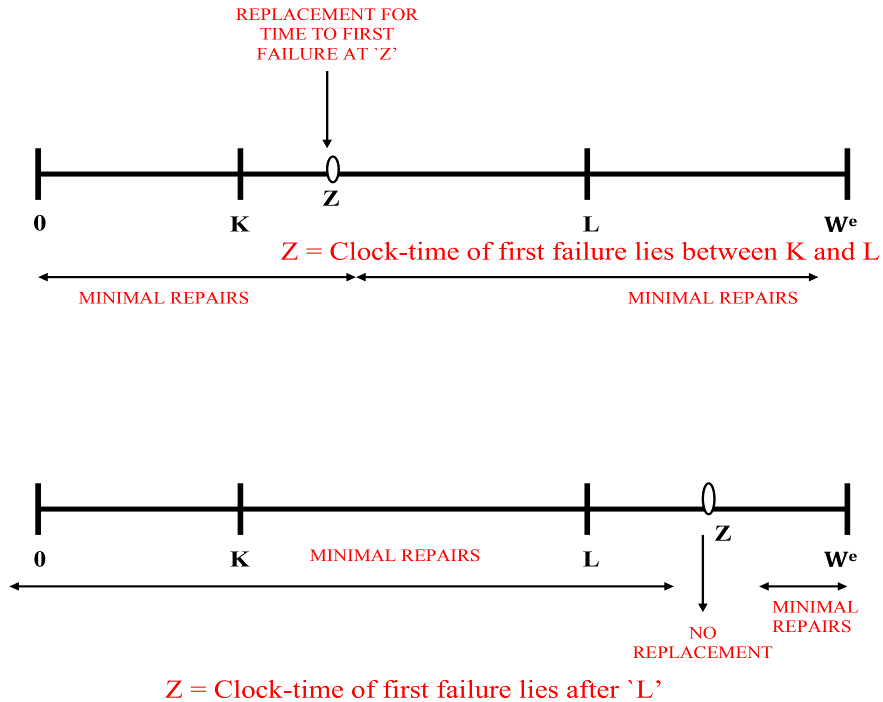


Figure 3.2 Graphical representation of warranty servicing strategy by splitting the warranty into three disjoint intervals.

in $[K, L)$ has been restricted to the first failure there in, if any. This choice seems reasonable because (i) there is no obvious criterion to decide how many replacements in $[K, L)$ should we undertake and, (ii) the fact that number of replacements beyond a critical threshold will make the warranty too expensive for the manufacturer.

3.3.2 Warranty Servicing Cost

For given usage rate y, the warranty of a product according to (3.2) will expire after time W^e as graphically demonstrated in Figure 3.1. The expected warranty servicing cost under the repair-replace strategy can be obtained by calculating the conditional servicing cost in each of the three intervals and then combining the results [18].

Since the first and subsequent failures during the interval $[0, K)$ are rectified via minimal repairs, so the expected repair cost for this period is

$$c_m \int_0^K h(x|y)\,dx = c_m H(K|y) \tag{3.16}$$

where c_m is the cost of a minimal repair.

The expected warranty servicing cost for the middle interval $[K, L)$ depends on the occurrence of the first failure after time K. The first failure in the region is rectified by replacement and all the subsequent failures being minimally repaired. If X denotes the time at which the first failure occurs after time K while still lying in the interval $[K, L)$, then the expected warranty servicing cost conditional on $X = x$ is given by

$$c_r + c_m \int_x^{W^e} h\left((t-x)\,|\,y\right)\,dt = c_r + c_m H\left((W^e - x)\,|\,y\right) \tag{3.17}$$

since the first failure results in the replacement by a new item. Here, c_r is the cost of the replacement; it is assumed that $c_m < c_r$.

For all failures after the time L, yet before W^e, then the expected warranty cost for rectifying these failures in the remainder of the warranty period is

$$c_m \int_L^{W^e} h(x|y)\,dx = c_m[H\left(W^e\,|\,y\right) - H\left(L\,|\,y\right)] \tag{3.18}$$

Finally, the total expected warranty cost is given by $I(K, L)$ which is the sum of the conditional costs from the equations (3.16)-(3.18), each integrated over the density function of usage rate $g(y)$:

$$\begin{aligned} I(K,L) &= \int_0^\infty \{c_m H(K|y)\} g(y)\,dy \\ &+ \int_0^\infty \left\{ \frac{c_m}{\overline{F}(K|y)} \left(\int_K^L [\rho + H((W^e - x)|y)] f(x|y)\,dx \right) \right\} g(y)\,dy \\ &+ \int_0^\infty \left\{ c_m [H(W^e|y) - H(L|y)] \frac{\overline{F}(L|y)}{\overline{F}(K|y)} \right\} g(y)\,dy \end{aligned} \tag{3.19}$$

The cost obtained in (3.19) is sensitive to the type of usage rate distribution used and the association between product life-length and usage rate. Based on Mean Time To First Failure (MTTF) [27], the most reliable usage rate distribution is selected. Since the consumer's usage rate is known via a probability distribution, non-parametric measures of association such as Kendall's tau and Spearman's rho are beneficial to depict the effect of usage rate on product life-length.

MTTF [27] is a reliability term based on methods and procedures for lifecycle predictions for a product. MTTF is sometimes misunderstood to be the life of the product instead of the expected values of the times to failure. Based on this value, it is easy to judge the reliability of each of the distribution in addition to the resulting lowered optimal warranty costs.

Reliability of an item can be defined as the probability that an item will perform a defined function without failure under stated conditions for a stated period of time. For an item at a given time x, reliability is described as

$$R(x) = P(X > x) \tag{3.20}$$

In general, MTTF is calculated as

$$MTTF = \int_0^\infty R(x)\, dx \tag{3.21}$$

Mean time to first failure in the context of warranties is redefined as

$$MTTF = \int_{y_{min}}^{y_{max}} \int_0^\infty R(x \mid y) g(y)\, dx\, dy \tag{3.22}$$

In (3.22), $R(x|y)$ is the reliability function at time x for given usage rate y and $g(y)$ is the probability density function of the usage rate. Using this expression, MTTF values for several common usage rate distributions are computed and compared to select the most reliable distribution. These distributions were all selected to have same mean and variance.

Table 3.1 Mean Time to First Failure for the Usage Rate Distributions

Usage Rate Distribution	Mean Time to First Failure
Uniform Distribution	1.7725
Normal Distribution	0.3356
Gamma Distribution	0.1705
Weibull Distribution	3.1953

From Table 3.1, it is observed that gamma distribution has the lowest value of MTTF and Weibull distribution has the highest value of MTTF. This implies that if the usage rate follows the gamma distribution, the warranty costs to service claims will be highest in comparison to other distributions. In order to track and improve the reliability of their products, manufacturing organizations must utilize an accurate and concise method to specify and measure the reliability. The distribution of the usage rate will clearly play an important role in warranty servicing costs.

We are unaware of the prior use of copulas to observe the effect of usage rate on time to failure. Hence, the choice to use the copula here is novel. In the Section 3.2 of Copula Formulation, we introduced non-parametric measures of association in context of copulas. Using equations (3.7), (3.11) and (3.12) we can obtain the copula for each of the usage rate distributions considered above. To verify our assumption about the inverse relationship between usage rate and product life-length as expressed in (3.4), we will compute the measures of association for each selected usage rate distributions. Since Kendall's tau and Spearman's rho provide alternative measures of the same association, we have included only the results for Kendall's tau in Table 3.2. It is observed that for each distribution, there is a higher negative value of Kendall's tau indicating negative correlation between usage rate and product life-length. This analysis is similar to that shown in the AFT formulation (3.4).

Table 3.2 Measure of Association Using Copula

Usage Rate Distribution	Kendall's Tau
Uniform Distribution	-0.9938
Normal Distribution	-0.8636
Gamma Distribution	-0.7626
Weibull Distribution	-0.8044

3.4 Concluding Remarks

In this chapter, we have discussed the servicing strategy adopted to obtain the expression for expected warranty cost (3.19). Also, we have observed and evaluated the factors affecting the cost function. Based on mean time to first failure, the usage rate distribution which is most reliable can be determined. In addition, the correlation between usage rate and product life-length is evaluated and a negative value is an indicator of inverse relationship between the two. It is clear that in general the optimal choice of K and L are not amenable to an analytic solution, but must be obtained numerically, which is the topic of the next chapter.

CHAPTER 4

SIMULATION OF OPTIMAL WARRANTY COST

Optimizing warranty policies from the perspective of manufacturers is an area of considerable managerial interest. Under a 2D warranty, policy we observed that the expected warranty cost incurred in following the strategy adopted to rectify failures depends on the parameters of the system and of the warranty itself. In this chapter, we will discuss the numerical methods used to obtain optimal values at which the cost can be minimized under the repair-replace strategy.

4.1 Numerical Procedures

Optimization of the warranty cost is performed using the numerical technique known as simulated annealing [28]. Simulated annealing is a method for solving unconstrained and bound - constrained optimization problems. This method, which is based Markov Chain Monte Carlo techniques, mimics the physical process of heating a solid material and then slowly cooling it by lowering its temperature to decrease defects, thus minimizing the system energy which corresponds to ground state of solid [28]. At each iteration of the simulated annealing algorithm, a new point as a candidate solution is randomly generated. The distance of the new point from the current point is based on a probability distribution with a scale proportional to the temperature. The algorithm accepts all new points that lower the objective, but also, with a certain probability, points that raise the objective. By accepting points that raise the objective, the algorithm avoids being trapped in local minima in early iterations and is able to explore globally for better solutions. Following this technique, the cost function $I(K, L)$ will be minimized at the optimal values K^* and L^* with respect to K and L, respectively.

Simulated annealing requires that after the integral expressions of cost in (3.19) be evaluated. Since it is impossible to find an antiderivative for each of the integrals as evaluated in equations (3.16) - (3.18), we have used a numerical integration technique known as Romberg Integration to obtain an approximate solution [29]. The Romberg numerical method is based on a Newton-Cotes formula that estimates the integral of a function over a given interval by dividing the interval into 2^k subregions estimating the integral within each subregion, then summing the estimates [29]. Additionally to evaluate the AFT copula function (3.15) and corresponding Kendall's tau (3.9) and Spearman's rho (3.10), a quadrature numerical method based on recursive adaptive Simpson rule is applied to approximate those integrals [31].

The analysis presented by Jack et al. [18] introduced the repair-replace strategy and summarized the optimal warranty cost under the assumption of known usage rate. Since such an assumption is unrealistic, we have extended the analysis with the assumption that the usage profile of customers is known via probability distribution. Our simulations proceed as follow: We begin by simulating a usage rate. The optimization routine uses the numerical integration method to evaluate the cost in (3.19) at each step as it finds the minimum cost. This process is repeated for various usage rates and the results are averaged to obtain the mean cost.

4.2 Discussion of Results

In order to obtain the results, the warranty and reliability parameters are selected suitable to the cost model and then the above numerical techniques are applied. Below is the choice of parameters considered here.

- Warranty Policy: $W = 2$ (years) and $U = 2$ ($\times 10^4$ km), so that $\frac{U}{W} = 1$
- Design Reliability: $y_0 = 1(\times 10^4$ km per year), $\alpha_0 = 1$ (year) and $\beta = 2$
- AFT Model: $\gamma = 2$

For the time to failure distribution, we used the Weibull distribution, as it is commonly used in the literature. The distributions used for the usage rate are uniform, normal, gamma and Weibull all having the same mean and variance [30]. Since the usage rate must be confined to a finite interval, we have used truncated versions of normal [31], gamma [32] and Weibull [33] distributions. The moments of truncated normal, truncated gamma and truncated Weibull distributions are computed using the formulae based in the article by J. W. Jawitz [30]. The specific usage rate distributions along with their parameters are Uniform(0.1,5), Normal(3.0656, 1.4587), Gamma(5.2477, 0.6559) and Weibull(1.1076, 2.3773).

We will divide our analysis into two parts: (a) Effect of usage rate, and (b) Effect of cost ratio on the optimal warranty cost. The cost ratio is defined as the ratio of cost of replacement to that of repairs $\rho = \frac{c_r}{c_m}$ with $c_m < c_r$. We will consider the choice of these cost parameters as $c_m = 1$ and $c_r = 2$, so that $\rho = 2$. For a given usage rate interval, the results in each case contain the values of K^*, L^*, $I(K^*, L^*)$ and W^e where (*) means the optimal value of the parameter and W^e is the effective warranty period.

For the results summarized in each of the tables, the optimal cost is obtained using the warranty servicing strategy involving only minimal repairs and replacement. The characterization of the optimal servicing strategy states that an upper bound to this cost can be obtained by considering the situation when all the failures are rectified only via minimal repairs. If $J^m(y)$ is the expected cost for minimal repairs then

$$J^m(y) = \begin{cases} 4y^4, & \text{if } y \le 1, \\ 4y^2, & \text{if } y > 1. \end{cases}$$

The optimal cost $I(K^*, L^*)$ is bounded above by the corresponding expected value of minimal repair costs $J^m(y)$ as

$$I(K^*, L^*) \le \int_0^\infty J^m(y)\, dG(y). \tag{4.1}$$

4.2.1 Effect of Usage Rate

In this section, we begin with the discussion of the results when the manufacturer uses the same warranty servicing strategy regardless of the usage rate of the customer.

Table 4.1 Expected Cost of Overall Usage Interval

Distribution	K*	L*	I(K*,L*)	W^e	Upper Bound
Uniform	0.1901	0.3182	11.032	[0.4,2]	33.905
Normal	0.1885	0.3263	11.711	[0.4,2]	38.3240
Gamma	0.1883	0.3290	12.0589	[0.4,2]	39.8204
Weibull	0.1874	0.3084	5.2968	[0.4,2]	19.4096

Table 4.1 shows that how the optimal warranty servicing cost is affected by the nature of distribution of the usage rate. In each case, the overall cost is less than the upper bound in (4.1) as expected. Additionally, it is clearly visible that the optimal warranty servicing cost is lowest for the Weibull distribution. This also agrees with our intuition since the MTTF in Table 3.1 was the greatest for the Weibull distribution.

It seems quite natural for a manufacturer to use different warranty servicing strategies for customers with differing product usage rates for a number of reasons. A single interval places the replacement interval at very different position in the life-length interval of a product for heavy user versus a light user. Also, heavy users of a product have a shorter effective warranty period. However, its is not practical for a manufacturer to have a unique warranty servicing strategy for every customer. Thus, it seems natural for a manufacturer to categorize customers according to usage rate [34]. For example, in a three category system, customer could be thought of as low, medium, or heavy users.

The optimal costs are obtained here for all the usage rate distributions for both a three and six category system with categories determined by a regular partition of the range of possible usage rates. In each case, the results for categorized treatment are compared with those obtained from the overall interval. In order to compare the overall costs varying categorizations, we have computed the overall weighted expected cost for the case of categorized treatment.

Case : Usage Rate: Uniform Distribution

Table 4.2 Optimal Warranty Servicing Cost for Overall Interval and Categorized Treatment

Optimal Cost for Overall Interval				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 5$	0.1901	0.3182	11.032	[0.4,2]
Expected Cost for Three Equally Probable Sub-Intervals				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 1.7333$	0.4461	0.8829	0.6410	[1.1539,2]
$1.7333 < y < 3.3667$	0.2709	0.5292	3.3866	[0.5940,1.1539]
$3.3667 < y < 5$	0.1928	0.3403	9.1413	[0.4,0.5940]
Weighted Optimal Cost of the Sub-Intervals = 4.3905				
Optimal Cost for Six Equally Probable Sub-Intervals				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 0.9166$	2.08E-04	2.95E-04	0.1056	2
$0.9166 < y < 1.7333$	0.4466	0.9317	0.5947	[1.1538,2]
$1.7333 < y < 2.5500$	0.3420	0.7353	1.3481	[0.7843,1.1538]
$2.5500 < y < 3.3667$	0.2774	0.5577	2.5176	[0.5941,0.7843]
$3.3667 < y < 4.1834$	0.2294	0.4649	4.1329	[0.4781,0.5941]
$4.1834 < y < 5$	0.1946	0.3370	6.1917	[0.4,0.4781
Weighted Optimal Cost of the Sub-Intervals = 2.4823				

Case : Usage Rate: Normal Distribution

Table 4.3 Optimal Warranty Servicing Cost for Overall Interval and Categorized Treatment

Optimal Cost for Overall Interval				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 5$	0.1885	0.3263	11.711	[0.4,2]
Expected Cost for Three Sub-Intervals				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 1.7333$	0.4520	0.9174	0.5029	[1.1539,2]
$1.7333 < y < 3.3667$	0.2714	0.5248	4.8576	[0.5940,1.1539]
$3.3667 < y < 5$	0.192	0.389	9.0935	[0.4,0.5940]
Weighted Optimal Cost of the Sub-Intervals = 5.6309				

Optimal Cost for Six Sub-Intervals				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 0.9166$	0.0066	0.0066	0.0462	2
$0.9166 < y < 1.7333$	0.4504	0.9505	0.4824	[1.1538,2]
$1.7333 < y < 2.5500$	0.3452	0.7243	1.7128	[0.7843,1.1538]
$2.5500 < y < 3.3667$	0.2776	0.5606	3.7578	[0.5941,0.7843]
$3.3667 < y < 4.1834$	0.2292	0.3898	5.3941	[0.4781,0.5941]
$4.1834 < y < 5$	0.1944	0.3362	5.2426	[0.4,0.4781]
Weighted Optimal Cost of the Sub-Intervals = 3.3050				

Case : Usage Rate: Gamma Distribution

Table 4.4 Optimal Warranty Servicing Cost for Overall Interval and Categorized Treatment

Optimal Cost for Overall Interval				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 5$	0.1883	0.3290	12.0589	[0.4,2]

Optimal Cost for Three Sub-Intervals				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 1.7333$	0.4544	0.9549	0.4707	[1.1539,2]
$1.7333 < y < 3.3667$	0.271	0.53	5.3464	[0.5940,1.1539]
$3.3667 < y < 5$	0.1921	0.3806	9.1855	[0.4,0.5940]
Weighted Optimal Cost of the Sub-Intervals = 6.1699				

Optimal Cost for Six Sub-Intervals				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 0.9166$	0.0158	0.0158	0.0169	2
$0.9166 < y < 1.7333$	0.4516	0.9668	0.4632	[1.1538,2]
$1.7333 < y < 2.5500$	0.3453	0.7262	1.9901	[0.7843,1.1538]
$2.5500 < y < 3.3667$	0.2772	0.5898	4.0691	[0.5941,0.7843]
$3.3667 < y < 4.1834$	0.2292	0.4339	5.3798	[0.4781,0.5941]
$4.1834 < y < 5$	0.1945	0.3143	5.3444	[0.4,0.4781]
Weighted Optimal Cost of the Sub-Intervals = 3.6096				

Case : Usage Rate: Weibull Distribution

Table 4.5 Optimal Warranty Servicing Cost for Overall Interval and Categorized Treatment

Optimal Cost for Overall Interval				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 5$	0.1874	0.3084	5.2968	[0.4,2]

Optimal Cost for Three Sub-Intervals				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 1.7333$	0.4499	0.8642	0.9177	[1.1539,2]
$1.7333 < y < 3.3667$	0.2674	0.5256	2.7503	[0.5940,1.1539]
$3.3667 < y < 5$	0.1921	0.3652	3.6686	[0.4,0.5940]
Weighted Optimal Cost of the Sub-Intervals = 1.8796				

Optimal Cost for Six Sub-Intervals				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 0.9166$	0.0073	0.0073	0.1846	2
$0.9166 < y < 1.7333$	0.4454	0.9191	0.8373	[1.1538,2]
$1.7333 < y < 2.5500$	0.342	0.7265	1.3943	[0.7843,1.1538]
$2.5500 < y < 3.3667$	0.2767	0.5818	1.8482	[0.5941,0.7843]
$3.3667 < y < 4.1834$	0.2292	0.4668	2.1092	[0.4781,0.5941]
$4.1834 < y < 5$	0.1945	0.3779	2.1625	[0.4,0.4781]
Weighted Optimal Cost of the Sub-Intervals = 1.0552				

Tables 4.2-4.5 display the optimal costs for overall and stratified intervals. It is noted that the optimal cost for the single usage rate interval is around double the overall cost for the three category system and triple the overall cost for the six category system. Thus, an increase in the sub-intervals (categories of customers) lower the overall cost for the manufacturer. Similar to our observation in Table 4.1, it is also noted that among the usage rate distributions the overall cost for the Weibull distribution is the lowest and for the gamma distribution is highest. This is in agreement with our results for MTTF in Section 3.3. These results also demonstrate how sensitive the value of optimal warranty cost is to the type of usage rate distribution.

4.2.2 Effect of Cost Ratio

In this section, we study the impact of an increase in cost ratio on the optimized cost. The optimization results in the previous section are evaluated using a cost ratio of replacement to repairs of 2, i.e., $\rho = 2$. Next, for the sake of simplicity we, will consider the possible choices of cost ratio as $\rho = 7, 15$. The results are given below for the same usage rate distributions for overall usage interval and the stratified sub-intervals of the usage rate interval.

Case : Usage Rate: Uniform Distribution

Table 4.6 Optimal Warranty Servicing Cost for Observing the Effect of Cost Ratio

(a) Optimal Warranty Servicing Cost for Overall Usage Interval

ρ	K*	L*	I(K*,L*)
2	0.190	0.319	11.032
7	0.193	0.231	13.870
15	0.194	0.210	17.334

(b) Optimal Warranty Servicing Cost for Categorized Treatment

Cost Ratio $\rho = 2$				
y	K*	L*	I(K*,L*)	W^e
$0.1 < y < 1.7333$	0.4461	0.8829	0.6410	[1.1539,2]
$1.7333 < y < 3.3667$	0.2709	0.5292	3.3866	[0.5940,1.1539]
$3.3667 < y < 5$	0.1928	0.3403	9.1413	[0.4,0.5940]
Cost Ratio $\rho = 7$				
y	K*	L*	I(K*,L*)	W^e
$0.1 < y < 1.7333$	1.33E-04	1.90E-04	0.8502	[1.1539,2]
$1.7333 < y < 3.3667$	0.2782	0.3466	4.8709	[0.5940,1.1539]
$3.3667 < y < 5$	0.1928	0.3527	10.8079	[0.4,0.5940]
Cost Ratio $\rho = 15$				
y	K*	L*	I(K*,L*)	W^e
$0.1 < y < 1.7333$	3.51E-05	7.32E-05	0.8502	[1.1539,2]
$1.7333 < y < 3.3667$	3.22E-05	5.09E-05	6.0049	[0.5940,1.1539]
$3.3667 < y < 5$	0.1931	0.2393	13.4692	[0.4,0.5940]

Case : Usage Rate: Normal Distribution

Table 4.7 Optimal Warranty Servicing Cost for Observing the Effect of Cost Ratio

(a) Optimal Warranty Servicing Cost for Overall Usage Interval

ρ	K*	L*	I(K*,L*)
2	0.1885	0.3263	11.711
7	0.1915	0.2315	15.2059
15	0.1967	0.2101	19.1773

(b) Optimal Warranty Servicing Cost for Categorized Treatment

Cost Ratio $\rho = 2$				
y	K*	L*	I(K*,L*)	W^e
$0.1 < y < 1.7333$	0.4520	0.9174	0.5029	[1.1539,2]
$1.7333 < y < 3.3667$	0.2714	0.5248	4.8576	[0.5940,1.1539]
$3.3667 < y < 5$	0.192	0.389	9.0935	[0.4,0.5940]

Cost Ratio $\rho = 7$				
y	K*	L*	I(K*,L*)	W^e
$0.1 < y < 1.7333$	5.53E-05	8.17E-05	0.6962	[1.1539,2]
$1.7333 < y < 3.3667$	0.2772	0.3518	6.9252	[0.5940,1.1539]
$3.3667 < y < 5$	0.1920	0.3990	10.9311	[0.4,0.5940]

Cost Ratio $\rho = 15$				
y	K*	L*	I(K*,L*)	W^e
$0.1 < y < 1.7333$	2.52E-05	4.51E-05	0.6962	[1.1539,2]
$1.7333 < y < 3.3667$	0.0045	0.0045	8.6808	[0.5940,1.1539]
$3.3667 < y < 5$	0.1925	0.2359	13.8624	[0.4,0.5940]

Case : Usage Rate: Gamma Distribution

Table 4.8 Optimal Warranty Servicing Cost for Observing the Effect of Cost Ratio

(a) Optimal Warranty Servicing Cost for Overall Usage Interval

ρ	K^*	L^*	$I(K^*,L^*)$
2	0.1881	0.3286	12.0589
7	0.1916	0.2303	15.7316
15	0.1945	0.2073	19.7789

(b) Optimal Warranty Servicing Cost for Categorized Treatment

Cost Ratio $\rho = 2$				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 1.7333$	0.4520	0.9174	0.5029	[1.1539,2]
$1.7333 < y < 3.3667$	0.2714	0.5248	4.8576	[0.5940,1.1539]
$3.3667 < y < 5$	0.192	0.389	9.0935	[0.4,0.5940]
Cost Ratio $\rho = 7$				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 1.7333$	2.48E-05	4.69E-05	0.6755	[1.1539,2]
$1.7333 < y < 3.3667$	0.2778	0.3511	7.6609	[0.5940,1.1539]
$3.3667 < y < 5$	0.1921	0.3414	11.0342	[0.4,0.5940]
Cost Ratio $\rho = 15$				
y	K^*	L^*	$I(K^*,L^*)$	W^e
$0.1 < y < 1.7333$	3.66E-05	5.66E-05	0.6755	[1.1539,2]
$1.7333 < y < 3.3667$	1.08E-05	2.27E-05	9.5293	[0.5940,1.1539]
$3.3667 < y < 5$	0.1924	0.2357	13.9828	[0.4,0.5940]

Case : Usage Rate: Weibull Distribution

Table 4.9 Optimal Warranty Servicing Cost for Observing the Effect of Cost Ratio

(a) Optimal Warranty Servicing Cost for Overall Usage Interval

ρ	K*	L*	I(K*,L*)
2	0.1874	0.3084	5.2968
7	0.1923	0.2226	6.9770
15	0.2041	0.2144	8.6152

(b) Optimal Warranty Servicing Cost for Categorized Treatment

Cost Ratio $\rho = 2$				
y	K*	L*	I(K*,L*)	W^e
$0.1 < y < 1.7333$	0.4499	0.8642	0.9177	[1.1539,2]
$1.7333 < y < 3.3667$	0.2674	0.5256	2.7503	[0.5940,1.1539]
$3.3667 < y < 5$	0.1921	0.3652	3.6686	[0.4,0.5940]
Cost Ratio $\rho = 7$				
y	K*	L*	I(K*,L*)	W^e
$0.1 < y < 1.7333$	0.13E-04	1.32E-04	1.1924	[1.1539,2]
$1.7333 < y < 3.3667$	0.2795	0.3391	4.0468	[0.5940,1.1539]
$3.3667 < y < 5$	0.1921	0.3601	4.4025	[0.4,0.5940]
Cost Ratio $\rho = 15$				
y	K*	L*	I(K*,L*)	W^e
$0.1 < y < 1.7333$	1.64E-05	5.30E-05	1.1924	[1.1539,2]
$1.7333 < y < 3.3667$	1.96E-05	4.67E-05	4.7894	[0.5940,1.1539]
$3.3667 < y < 5$	0.1932	0.2371	5.5731	[0.4,0.5940]

Tables 4.6-4.9 demonstrate the results observing the effect of cost ratio for overall and stratified usage rate interval. The impact of cost ratio has similar trend for both the uncategorized and categorized treatment. As the cost ratio increases from 2 to 15, i.e., increasing the cost of replacement, it becomes expensive to service the failures as would be expected. Numerically, the replacement interval shrinks to zero as cost ratio increases indicating that the manufacturer will rectify failures only though minimal repairs. For example, from Table 4.6(a) for ρ=2, the interval length for replacement is 0.129 whereas for ρ=15 the interval length for replacement becomes 0.016.

4.3 Concluding Remarks

In this chapter, we have observed how the categorized treatment in comparison to uncategorized treatment results in lower warranty servicing costs for a manufacturer. Increasing the cost ratio leads to higher warranty servicing costs and eventually shrinks the replacement interval to zero. The overall objective of our research is achieved by investigating the best choice of usage rate distribution that minimizes the optimal warranty cost. In the next chapter, to illustrate practicality of our model we will calculate the optimal warranty cost using a data set.

CHAPTER 5

OPTIMAL WARRANTY COSTS BASED UPON DATA ANALYSIS

Using the repair-replace model for addressing the failures of repairable products, we set up a cost model in Chapter 3. We then studied various warranty servicing strategies for many different usage rate distributions in Chapter 4. We now apply our model to a product for which a data set is available.

5.1 Description of Data

The data under consideration here is taken from the maintenance records of a locomotive traction motor [35]. The data set includes the time since inception of service and miles accumulated by different traction motors when they were returned to the maintenance depot upon failing. There are a total of 147 failures recorded in this data set, of which only the first 40 arrive at the depot [35]. It is assumed that a free replacement warranty has been provided by the manufacturer to cover first 500 days and 20,000 miles after the motor is sold.

5.2 Preliminary Data Analysis

Prior to application of warranty model on the given data, it is important to inspect the data by verifying the source of the data as well as the data itself includes the variables and units of the measurements [36]. The objectives of preliminary data analysis is to summarize the information content of the data in an easily understood format with the help of descriptive statistics, represent the data graphically, and prepare the data for detailed statistical analysis.

Descriptive statistics [37] describe the basic features of the data considered in a study, and also provide simple summaries about the sample and the descriptive measures together with graphical analysis to form the basis of quantitative analysis of the data. Some measures that are commonly used to describe a data set are measures of central tendency and measures of variability or dispersion. Measures of central tendency include the mean, median, and mode, while measures of variability include the standard deviation (or variance), the minimum and maximum values of the variables, skewness, and kurtosis. Skewness is a measure of symmetry, or more precisely, lack of symmetry. As a general rule, for data skewed to the right, the mean will be greater than the median and vice versa for the data skewed to the left. Kurtosis is a measure of whether the data are peaked or flat relative to the normal distribution. The reference standard is a normal distribution, which has kurtosis of 3. Data sets with high kurtosis (leptokurtic) tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails and with low kurtosis (platykurtic) tend to have a flat top near the mean rather than a sharp peak. Another measure of variability is indicated via the quartiles Q_1 (25%), Q_2 (50%) and Q_3 (75%) that divide the data set into four equal groups, each group comprising a quarter of the data. Table 5.1 presents the descriptives for both the time to failure and usage rate. The data for time to failure indicates positive skewness and high kurtosis which shows that the notion of normal distribution can not be applied. The usage rate data has negative skewness and high kurtosis which also indicates that it is not normally distributed.

Table 5.1 Descriptive Statistics of Time to Failure and Usage Rate for the Original Data

Time to Failure							
Sample Size	Mean	Std. Dev	Skewness	Kurtosis	Q_1	Q_2	Q_3
40	0.6477	0.7255	1.9916	6.9180	0.1219	0.4520	0.8082
Usage Rate							
Sample Size	Mean	Std. Dev	Skewness	Kurtosis	Q_1	Q_2	Q_3
40	2.7258	0.7468	-1.2354	5.0881	2.4361	2.8287	3.1704

Graphs allow one to observe patterns and distributions of data that are difficult to discern by looking at the raw data [37]. Histograms are popularly used as graphical representation of frequency distribution for quantitative data. The data is grouped according to some specified grouping algorithm and then frequencies are plotted in the form of vertical bar chart. Figures 5.1 and 5.2 show the histogram plots for time to failure and usage rate. The histogram for time to failure indicates that majority of the observations are clustered to the right indicator of positive skewness with extreme observations clearly visible. It is noted that the histogram of usage rate is roughly symmetric with slightly visible negative skewness. These plots are in agreement with the results in Table 5.1.

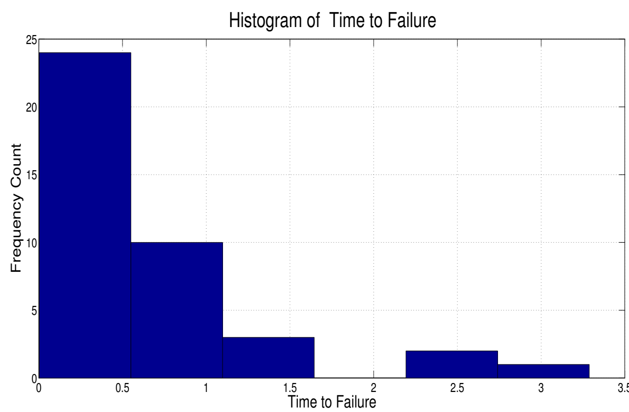


Figure 5.1 Histogram of time to failure for original data.

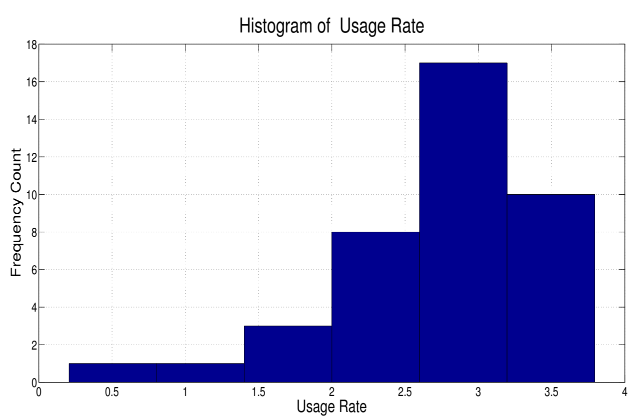


Figure 5.2 Histogram of usage rate for original data.

5.2.1 Data Filtering

Data filtering is helpful in getting in identifying extreme observations that can significantly affect the outcome of the tests and statistical procedures. Visual techniques using box plots are frequently used to identify the outliers, particularly in extreme cases. Box plots or box and whiskers plots were introduced by John Tukey [37]. The spacings between the different parts of the box plot help indicate the degree of dispersion (spread) and skewness in the data, and identify outliers. This is a plot in which the middle 50% of the data, i.e., all observations between the quartiles Q_1 and Q_3 is represented by a rectangle; the horizontal line within the box indicates the location of median Q_2. The remaining data are indicated by vertical lines outside the box or by points beyond the lines. The length of each of these vertical lines ("whiskers") is calculated as 1.5 times the interquartile range (IQR) where IQR = Q_3 - Q_1. Figures 5.3 and 5.4 show box plots of the time to failure and usage rate for the data. In both case, we observe that there are data points lying outside the "whiskers". Using Tukey's [37] procedure, data values are identified as outliers if they are outside the interval $[Q_1 - 1.5\,IQR, Q_3 + 1.5\,IQR]$ and as extreme outliers if they lie outside the interval $[Q_1 - 3\,IQR, Q_3 + 3\,IQR]$. Hence, we will consider these points as outliers.

When outliers have been detected, it is recommended to not include this data in statistical analysis. The goal is not only to bring the data set into a consistent state but also to ensure an accurate and complete representation of the real-world constructs to which the data refer, in this instance, to the traction motor data. Hence, we will remove the outliers shown in Figures 5.3 and 5.4 from the traction motor data prior to obtaining the best fitting distribution.

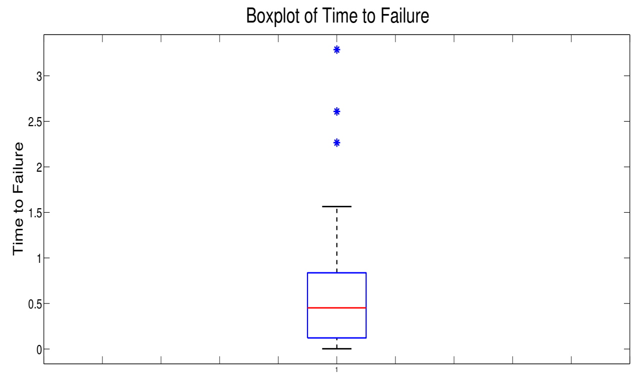


Figure 5.3 Boxplot of time to failure for original data.

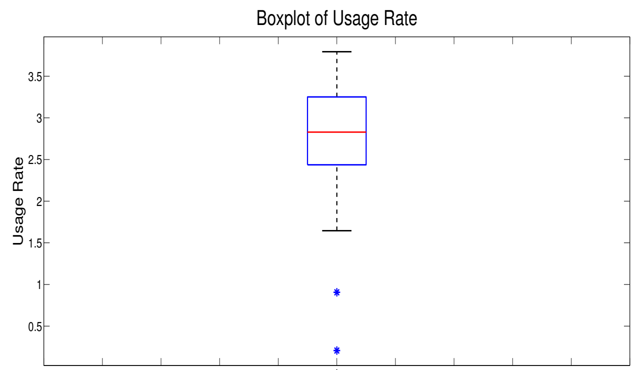


Figure 5.4 Boxplot of usage rate for original data.

5.2.2 Data Fitting

Fitting an inappropriate distribution to the given data would certainly lead to failure in making future predictions. Distribution fitting is aimed at finding the distribution that more closely corresponds to the observed frequency of the data than others, depending on the characteristics of the phenomenon and of the distribution. Pearson [38] gave a four-parameter system of probability density functions, and fitted the parameters by what is called the "method of moments". In most cases, it is helpful to fit two or more distributions, compare the results, and select the most appropriate model. A slightly more sophisticated test is to compute the moments of the actual data distribution and to examine them for fit to the chosen distribution. As a preliminary step to fitting distribution, we will look briefly at the methods used for estimating the parameters of the candidate distributions. The most commonly use methods are method of moments (MOM), maximum likelihood estimates (MLE), least square estimates (LSE) and method of L-moments.

The *method of moments (MOM)*, introduced by Karl Pearson [38], is the oldest method of deriving point estimators. The method of moments is based on matching the sample moments with the corresponding population (distribution) moments and is based on the assumption that sample moments should provide good estimates of the corresponding population moments. Because the population moments are often functions of the population parameters, corresponding population and sample moments are equated to solve for these parameters in terms of the moments. (`http://www.itl.nist.gov/div898/handbook/eda/section3/eda3651.htm`)

Maximum likelihood estimation (MLE) begins with writing a mathematical expression known as the Likelihood Function of the sample data. The likelihood of a set of data is the probability of obtaining that particular set of data given the chosen probability distribution model. This expression contains the unknown model parameters. The values of these parameters that maximize the sample likelihood are known as the Maximum Likelihood Estimates or MLE's. (`http://www.itl.nist.gov/div898/handbook/apr/section4/apr412.htm`)

In *least squares (LS)* estimation, the unknown values of the regression coefficients in the regression function are estimated by evaluating the numerical values for the parameters that minimize the sum of the squared deviations between the observed responses and the functional portion of the model. For the linear model, the least square estimates of the parameters are computed by equating the partial derivatives of the sum of the squared deviations, with respect to the parameters of interest, equal to zero and further solving for those estimated parameters. (`http://www.itl.nist.gov/div898/handbook/pmd/section4/pmd431.htm`)

The *L-moments* [39] are L-statistics (linear combinations of order statistics, “L” for “linear”) that are used for characterizing the shape of a probability distribution and estimating the distribution parameters, particularly for data where sample sizes are commonly small. Unlike product moments, the sampling properties for L-moments statistics are nearly unbiased, even in small samples, and are near normally distributed. These properties make them well suited for characterizing data that commonly exhibit moderate to high skewness.

Once the best parameters in the possible distributions have been determined, the best fit distribution must be determined. The typical tests for goodness of fit [40] compare the actual distribution function of the data with the cumulative distribution of the distribution that is being used to characterize the data, to either accept the hypothesis that the chosen distribution fits the data or to reject it. The goodness of fit (GOF) tests [41, 42] measure the compatibility of a random sample with a theoretical probability distribution function. In other words, these tests show how well the distribution fits to the data. The general procedure consists of defining a test statistic which measures the distance between the hypothesis and the data, and then calculating the probability of obtaining data which have a still larger value of this test statistic than the value observed, assuming the hypothesis is true. This probability is called the confidence level. Small probabilities (say, less than one percent) indicate a poor fit. Especially high probabilities (close to one) correspond to a fit which is too good to happen very often. In this section, we will discuss briefly the Chi-Square, Anderson Darling and Kolmogorov-Smirnov tests [43, 44].

Chi-Square Goodness of Fit Test

The Chi-Square goodness of fit test [43, 44], also referred to as the Chi-Square test for a single sample is used to determine the goodness of fit of the experimental data with a theoretical distribution. In this test, observed values are compared with theoretical or expected values. Observed values are those that the researcher obtains empirically through direct observation; theoretical or expected values are developed on the basis of some hypothesis. This test is applied to the observed data after it has been divided into bins on categories. This is actually not a restriction since it is a straightforward to bin data and is usual that one wishes to construct a histogram or frequency table is before conducting the Chi-Square test.

Every hypothesis test requires the statistician to state a null hypothesis and an alternative hypothesis in such a way that they are mutually exclusive. For a Chi-Square goodness of fit test, the hypotheses take the following form.

H_0: The data follow the specified distribution.

H_A: The data do not follow the specified distribution.

The hypothesis can be tested using the Chi-Squared statistic defined as:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed frequency count for the ith level of a categorical variable, and E_i is the expected frequency count for the ith level of the categorical variable. The expected frequency counts at each level of the categorical variable are equal to the sample size times the hypothesized proportion from the null hypothesis, i.e., $E_i = np_i$. For the Chi-Square approximation to be valid, the expected frequency should be at least 5. This test is not valid for small samples, and if some of the counts are less than five, it is required to combine some bins in the tails. Since the test statistic is a Chi-Square, the associated probability (p-value) is computed using $k - 1$ degrees of freedom where k is the number of levels of the categorical variable. On comparing the p-value to the given significance level, the null hypothesis is rejected when the p-value is less than the significance level.

Kolmogorov-Smirnov Test (K-S test)

This Kolmogorov-Smirnov test [43, 44] is used to decide if a sample comes from a hypothesized continuous distribution. It is based on the empirical cumulative distribution function (ECDF). The distribution of the K-S test statistic does not depend on the underlying cumulative distribution function being tested. Another advantage is that it is an exact test where the Chi-Square goodness of fit test depends on an adequate sample size for the approximations to be valid.

For a Kolmogorov-Smirnov goodness of fit test, the hypotheses take the following form:

H_0: The data follow the specified distribution.

H_A: The data do not follow the specified distribution.

The hypothesis can be tested using the test statistic defined as:

$$D_n = \sup_x |F_n(x) - F(x)|$$

where $F(x)$ equals some hypothesized distribution function and $F_n(x)$ is the empirical distribution function denoting the fraction of sample values less than or equal to x for a given random sample of size n.

At any point x, a large difference between the empirical distribution $F_n(x)$ and the hypothesized distribution $F(x)$ would suggest that the empirical distribution does not equal the hypothesized distribution. Therefore, we reject the null hypothesis if D_n is too large.

Anderson-Darling Test (A-D)

The Anderson-Darling test [43, 44] is a statistical test to determine whether a given sample of data is drawn from a specified theoretical probability distribution. The Kolmogorov-Smirnov test is distribution free in the sense that the critical values (threshold values to which the value of the test statistic in a sample is compared to determine whether or not the null hypothesis is rejected) do not depend on the specific distribution being tested. On the other hand, the Anderson-Darling test makes use of the specific distribution in calculating critical values.

For an Anderson-Darling goodness of fit test, the hypotheses take the following form:

H_0: The data follow the specified distribution.

H_A: The data do not follow the specified distribution.

The hypothesis can be tested using the test statistic defined as:

$$A^2 = -n - \frac{1}{2}\sum_{i=1}^{n}(2i-1)[\ln(F(X_i)) + \ln(F(X_{n-i+1})]$$

where A is the test statistic to assess if the given data $\{X_1 < \cdots < X_n\}$ from a random sample of size n comes from a cumulative distribution function F. The test statistic can then be compared against the critical values of the theoretical distribution, which are given by Scholz et al. [45] for many popular distributions. The null hypothesis is rejected if the test statistic is greater than the critical value.

5.2.3 Model Selection

The parameter estimation and goodness of fit tests are useful in model building. An approach to model selection when dealing with complex models is to prepare a list of candidate models, use methods of parameter estimation and compare the results. The selected distributions are compared based on the goodness of fit tests. Out of the list of well known distributions, we have selected Weibull, normal, gamma and lognormal distributions as most of the research in reliability has been performed using these distributions.

Once the fitting procedure is completed, the distributions are ranked as per the goodness of fit tests. Since the goodness of fit test statistics indicate the distance between the data and the fitted distributions, the distribution with the lowest test statistic is assigned rank of 1 with the next best model assigned rank of 2, etc., to compare the fitted models and select the valid model. The disadvantage of Chi-Square test is its sensitivity to the grouping of the data. Though, Anderson−Darling and Kolmogorov-Smrinov tests do not depend on the grouping of the data, the former test has higher power than the latter and detects discrepancies in the tails (NIST/SEMATECH e-Handbook of Statistical Methods, `http://www.itl.nist.gov/div898/handbook/`). While all these tests were applied, the Anderson-Darling test is the best test for small samples [45]. Tables 5.3 and 5.4 display the ranking of the distributions for time to failure and usage rate, respectively. According to the fitting results using the A-D test, the Weibull distribution is the most appropriate model while for usage rate we have to choose between Weibull (3P) and normal distributions, since the test statistic values for Weibull (3P) and normal are extremely close in values, 0.08 and 0.09, respectively. We select the normal distribution because of difficulties in calculating truncated moments for the Weibull (3P) distributions.

Table 5.2 Ranks of Distributions for Time to Failure Using the Goodness of Fit Tests

Distribution	K-S	A-D	Chi-Sq.
Gamma	1	2	5
Gamma (3P)	7	6	N/A [a]
Lognormal	5	5	2
Lognormal (3P)	4	4	1
Normal	2	3	3
Weibull	3	1	4
Weibull (3P)	6	7	N/A[a]

[a]The test says not a suitable fit

Table 5.3 Ranks of Distributions for Usage Rate Using the Goodness of Fit Tests

Distribution	K-S	A-D	Chi-Sq.
Gamma	5	5	2
Gamma (3P)	4	4	5
Lognormal	6	6	6
Lognormal (3P)	2	3	4
Normal	3	2	3
Weibull	7	7	7
Weibull (3P)	1	1	1

5.3 Warranty Data Analysis

Next, we will make use of the fitting information in the warranty model and obtain the optimal warranty costs for the given scenario [46]. As discussed in the previous section, the Weibull distribution is used to model time to failure for the traction motor. The estimated parameters of the distribution are scale =1.21 and shape=0.5. Additionally, the normal distribution is used to model usage rate with estimated parameters are standard deviation=0.72 and mean=2.77. We have considered two-dimensional ($\Omega = [0, W) \times [0, U)$) non-renewing free replacement warranty where W=2 (years) and U=2 ($\times 10^4$ km). The time to first failure and subsequent failures are modeled via the AFT approach as discussed in Chapter 3.

Table 5.4 shows the results for the optimal cost over the complete usage interval and for stratified treatment of usage interval into three and six sub-intervals. It is noted that the total optimal cost is almost two times the weighted cost for three category system and almost 2.5 times the weighted cost for the six category system. Therefore, the grouping of consumer population into a large number of groups each

Table 5.4 Expected Cost for Overall and Stratified Usage Intervals

y	K^*	L^*	$I(K^*,L^*)$	W^e
0.01<y<4.5	0.1974	0.3314	12.0376	[0.44,2]
Usage Interval Divided into Three Sub-Intervals				
y	K^*	L^*	$I(K^*,L^*)$	W^e
0.01<y<1.5067	0.5467	0.9011	0.2488	[1.3274,2]
1.5067<y<3.0033	0.2945	0.5279	7.8367	[0.6659,1.3274]
3.0033<y<4.5	0.2038	0.399	7.0811	[0.4444,0.6659]
Weighted Optimal Cost for Sub-Intervals = 6.3565				
Usage Interval Divided into Six Sub-Intervals				
y	K^*	L^*	$I(K^*,L^*)$	W^e
0.01<y<0.7583	0.0046	0.0047	0.0041	2
0.7583<y<1.5067	0.5427	0.9217	0.2459	[1.3274,2]
1.5067<y<2.2550	0.3783	0.7434	2.197	[0.8869,1.3274]
2.5500<y<3.0033	0.2993	0.614	6.2561	[0.6659,0.8869]
3.0033<y<3.7517	0.2468	0.4921	6.1571	[0.5331,0.6659]
3.7517<y<4.5	0.2092	0.3953	2.156	[0.4444,0.5331]
Weighted Optimal Cost for Sub-Intervals = 4.8524				

with different service treatment is helpful to the manufacturers in making wise use of resources. These conclusions are similar to those in Chapter 4.

5.4 Concluding Remarks

We have discussed the warranty cost modeling on the data related to products sold with 2-D warranty. After conducting the preliminary statistical fitting tests, we selected the appropriate distribution for the time to failure and usage rate. The optimal warranty costs are obtained and trends are observed similar to that in Chapter 4. The warranty data allows the manufacturer to assess the factors potentially effecting the overall profits and the results of the optimal costs are useful in making such decisions.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK ON WARRANTIES

Based on a three interval servicing strategy consisting of minimal repairs in the first and third time interval and replacement in the second time interval, we successfully constructed a warranty cost model that was optimized using simulated annealing. We observed that the optimal cost is sensitive to the usage rate distribution. The mean time to first failure is valuable indicator of which usage distributions tend to result in lower warranty servicing costs. Instead of servicing all customer warranties in exactly the same manner, it is observed that grouping customers according to usage rate led to lower warranty servicing costs for the manufacturer. Increasing the ratio of cost of replacement to repair leads to higher warranty servicing costs for the manufacturer and a reduction in the size of the replacement time interval. Also, it was seen that using a copula based on AFT model is obtained to capture the adverse impact of usage rate on the product life-length.

Future work includes the investigation on other joint distribution models of life-length and product use to examine their suitability as underlying building blocks of a warranty cost model. Using those models we will derive and numerically compute the copulas along with their dependency measures, to judge the potential suitability of such models. We will potentially search for other parametric copula families with a negative Kendall's tau, and superimpose on them reasonable marginal distributions of product life-length and usage rate. Additionally, we will consider possible models for fair pricing of extended warranties that are alternative to those used in the literature. We will model and analyze expected warranty servicing costs for the servicing strategy for an imperfect repair. We recently received a data set from Dr. Bharatendra K. Rai containing the information regarding usage and time to failures for vehicles. We will complete preliminary data analysis and apply our model to compare the results to those obtained in Chapter 5.

CHAPTER 7

INTRODUCTION TO GOLF HANDICAPS

In the game of golf, a person's playing ability is represented by the "handicap" [47]. The handicap can also be used to track a golfer's progress over time. The handicap is designed to allow golfers of different abilities to compete against each other in an equitable fashion. In this chapter, a brief review of the historical development of handicapping systems will be given followed by a mathematical description of the handicapping system currently in use.

7.1 The USGA Handicap System

The United States Golf Association (USGA), a non profit organization founded in 1894, has developed a handicap system with the expressed purpose of "making the game of golf more enjoyable by enabling golfers of differing abilities to compete on an equitable basis" [48]. The system provides a handicap for each player and adjusts a player's rating up or down as the player's game changes over time. The purpose of the system has always been to attempt to level the playing field for golfers of differing abilities, so that those golfers can compete equally. In match play the weaker player receives handicap strokes on certain designated holes in conjunction with the handicapping scheme. Before proceeding to investigate about the handicap system, it is important to understand the history and terms along with their associated roles in the system.

7.1.1 History of the Handicap System

The term "handicap" evolved during the 17th and 18th centuries from "hand in cap", a game consisting of two players having the objective of establishing fairness between them. The handicapping system gained prominence in Great Britain and Ireland during the late 19th century but efforts to make handicapping more uniform did not cover all courses and golfers. The popularity of golf made it increasingly difficult for the mentors to monitor the handicaps of fellow players; many clubs adopted mathematical procedures to determine the calculation of the handicaps. The most popular method was to take the average of a players best three scores for the year; however, that did not receive widespread acceptance or support. Though various governing bodies attempted to make handicapping more uniform and widespread, their efforts did not cover all courses and golfers. In contrast to the way golf developed in Great Britain and Ireland, there was one central golf authority in the US. After years of study and experimentation, the USGA adopted the first nationwide handicap format at a meeting on October 11, 1911, at Baltusrol Golf Club in Springfield, N.J. (`http://www.nebgolf.org/handicapping_education.html`)

Over the past 100 years, many changes and refinements have been made to the system in an attempt to minimize the advantage of one player over another. During the first several decades of the USGA Handicap System, improvements related to courses and players were made by various regional golf associations. By using the three best scores of the season as the basis for determining handicaps, it was clear the handicap system was intended to measure a player's potential, not playing ability. Later, the USGA increased the number of scores from the three lowest scores to the 10 best rounds with a minimum of 50 scores needed to obtain a handicap. This change gave an advantage to the average players, who now had a better chance of playing to their handicaps. Unfortunately, that increase triggered confusion as regional golf associations could not agree on the number of rounds from which to take the 10 best scores for handicap purposes. In 1967, the USGA reduced the requirement to 10 of the last 20 scores, which remains operative today. As the game continued to develop, the USGA constantly examined the intricacies of the system within the foundations established years ago. Irrespective of all the changes to the handicap system, one aspect that has remained constant is that the handicap is the measure of a players potential rather than an actual scoring average. (http://www.usga.org/news/2011/October/History-Of-Handicapping--Part-I/)

The fairness of the handicap has been researched and analyzed by a number of authors. Scheid [49], for example, suggested that the winning proportion of a better golfer is between $60 - 85\%$ provided the handicap difference between the two players was more than three strokes. Then, Pollock [50] addressed that, under certain modeling assumptions, the better golfer has an advantage in both medal (total strokes) and match play. The findings by Pollock and Scheid were related to the old handicap system, which was based on treating the handicap as a function of course rating only. In the early 1990s, the slope system was introduced to further generalize the old handicap system by considering course difficulty. Tallis [51] investigated that in various team competitions the existing golf handicapping systems can be extremely unfair. Larkey [52] reviewed the handicap system and studied other statistical applications in golf. Bingham et al. [53] considered medal play between two golfers and used the new slope system for handicapping. James [54] too included the details about the newer handicap system and examined the issue of existence of bias in the handicap system. The research in this dissertation is related to that of James but follows a different approach which will be discussed later.

The historical background of the handicap system includes the usage of the key terms of the USGA handicap system such as “Course Handicap”, “Course Rating”, “Handicap Differential”, “Handicap Index”, “Slope Rating”, “Slope System”, “USGA Handicap System” and “USGA Course Rating System”, which are explained in detail in the next section (USGA Manual 2002-2005) [47].

7.1.2 Important Terms of the USGA Handicap System

The terms [47] discussed in this section are the backbone of the USGA handicap system.

Gross Score: The gross score is the number of actual strokes taken by a player plus any penalty strokes. A player's most recent scores are denoted by $X_{j-19}, X_{j-18}, \cdots, X_j$.

Scratch Golfer: A scratch golfer is a player whose average score over 18 holes is equal to par. For example, if a standard course has a par of 72, a scratch golfer would need to be capable of averaging a score of 72.

Bogey Golfer: A bogey golfer is a player who averages a bogey, or one shot over par, on every hole they play. Since a bogey is one stroke above par per hole, a bogey golfer would average a score of 90 on a typical par 72 course and have a handicap of 18.

Course Rating: Course rating indicates the evaluation of the playing difficulty of a course for scratch golfers. It is usually expressed in terms of strokes. A par 72 course that is easy might have a course rating of 68.9; one that is difficult might have a course rating of 74.5. That means that a scratch golfer should be expected to average 68.9 strokes at the easier course and 74.5 at the more difficult one.

Bogey Rating: A bogey rating is the evaluation of the playing difficulty of a course for bogey golfers under normal course and weather conditions. It is expressed in strokes and is based on yardage and other obstacles to the extent that they affect the scoring difficulty of the bogey golfer.

Slope Rating: Slope rating is a measurement of the relative difficulty of a course for players who are not scratch golfers compared to the course rating. Currently bogey golfers are used to determine the slope rating. It is computed from the difference between the bogey rating and the course rating. The lowest slope rating is 55 and the highest is 155. A golf course of standard playing difficulty has a USGA slope rating of 113.

Handicap Differential: A handicap differential, to one decimal place, represents a golfer's score in relation to the USGA course rating adjusted to take into account slope rating. For a given course rating and slope rating, the handicap differential is computed using the most recent gross scores $X_{j-19}, X_{j-18}, \cdots, X_j$ as shown in the expression below:

$$D_i = \frac{(X_i - \text{Course Rating}) \times 113}{\text{Slope Rating}}, \text{where } i = j-19, j-18, \cdots, j \qquad (7.1)$$

Handicap Index: A handicap index indicates the measurement of a player's potential ability on a course of standard playing difficulty. It is expressed as a number taken to one decimal place and is used for conversion to a course handicap. The handicap index formula is based on the 10 best handicap differential(s) in a player's scoring record of the 20 most recent scores. (A handicap index is not issued to a player who has fewer than five acceptable scores). The handicap index is denoted by I_j and computed in two steps:

- The average of the lowest ten differentials $D_{j-19^*}, D_{j-18^*}, \cdots, D_{j-10^*}$ is obtained.
- The result is multiplied by .96 and truncated to 1 decimal place to obtain I_j.

Course Handicap: A course handicap indicates the number of handicap strokes a player receives from a specific set of tees at the course being played to adjust his scoring ability to the level of scratch golfer. The USGA course handicap accounts for slope rating and may award extra strokes on a difficult course or take away strokes on an easier course. Every golf course that is part of the USGA handicap system has a chart showing course handicaps for players based on their handicap index and the slope rating of the tees played. For example, the chart may show that a 14.5 handicapper playing tees with a slope of 108 has a course handicap of 13, or playing tees with a slope of 138 has a course handicap of 16. For a given slope rating and computed value of handicap index, after calculating the handicap indices I_j, the course handicap is computed using:

$$H_j = \frac{I_j \times \text{Slope Rating}}{113} \tag{7.2}$$

The resulting number is rounded off to the nearest whole number. Once the course handicap is computed, the number of handicap strokes is applied during the round. For example, a course handicap of 4 means the player has this score is reduced by one stroke, on each of the four toughest handicap holes. As a special case, for slope rating of 113, the course handicap is equal to the handicap index. For a course with slope rating greater than 113, the course handicap is greater than the handicap index. Golfers usually covert their handicap indices to handicaps using appropriate USGA course handicap tables available at the course to be played.

7.2 Applications of Statistics for the Handicap Calculation

Now, the role of statistics in calculation of the course handicap will be discussed. As mentioned in "Statistics in Golf" [52], this is the most mature area in the statistical analysis of golf. Order statistics are used to obtain the handicap index with the help of an expression of a probability density function, discussed in detail below.

7.2.1 Introduction to Order Statistics

We begin by defining the order statistics. Order statistics and functions of these statistics play an important role in numerous practical applications [55, 56]. Order statistics are a set of values in a random sample which have been ordered from least to greatest, where the *order* is not being used to describe a sequence in which they are drawn. If random variables $\{X_i\}$, $i = 1, 2, \cdots, n$ and their associated observations $\{x_i\}$, $i = 1, 2, \cdots, n$ are arranged in ascending order of magnitude and written as

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$$

with ordered set of observations,

$$x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$$

then $X_{i:n}$ is said to be ith-order statistic in a sample of size n.

7.2.2 Application of Order Statistics in Handicap Calculation

The probability density function of the handicap index requires ten order statistics. For the sake of brevity, we will discuss the derivation of the probability density function of the average of the first two order statistics $X_{1:n}$ and $X_{2:n}$ from a random sample of size n.

Joint Distribution of Any Two Order Statistics

For any given two order statistics x_i and x_j with their probability density functions as $f(x_i)$, $f(x_j)$ and cumulative distribution functions as $F(x_i)$, $F(x_j)$, respectively the combined joint probability density function [55, 56] is :

$$f(x_{i,j:n}) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}\{F(x_{i:n})\}^{i-1}\{F(x_{j:n}) - F(x_{i:n})\}^{j-i-1} \times \{1 - F(x_{i:n})\}^{n-j} f(x_{i:n}) f(x_{j:n}) \qquad 0 < x_{i:n} < x_{j:n} < 1 \tag{7.3}$$

A simpler version of the above equation can be used to describe the joint distribution of first two ordered handicap differentials substituted for first two order statistics. Therefore, for i=1 and j=2, the joint distribution function of $X_{1:n}$ and $X_{2:n}$ can be expressed as:

$$f_{1,2:n}(x_1, x_2) = n(n-1)\{1 - F(x_2)\}^{n-2} f(x_1) f(x_2), \qquad -\infty < x_1 < x_2 < \infty \tag{7.4}$$

Single Order Statistic

The marginal probability density function [55, 56] of a random variable $X_{i:n}$ $(1 \leq i \leq n)$ corresponding to an ordered observation x_i is obtained by integrating the joint probability density of x_i and x_j in (1.3) to obtain:

$$f_{i:n}(x_{i:n}) = \frac{n!}{(i-1)!(n-i)!}\{F(x_{i:n})\}^{i-1}\{1-F(x_{i:n})\}^{n-i}f(x_{i:n}) \qquad (7.5)$$

$$-\infty < x_{i:n} < \infty$$

A simpler version of the above equation is the marginal distribution of a handicap differential substituted for a given single order statistic. Therefore, for i=1 and j=2, the marginal distribution function for single ordered statistics $X_{1:n}$ and $X_{2:n}$ is expressed as:

i =1 : $\quad f_{1:n}(x_{1:n}) = n\{1-F(x_{1:n})\}^{n-1}f(x_{1:n}),\ -\infty < x_{1:n} < \infty$

i =2 : $\quad f_{2:n}(x_{2:n}) = n(n-1)\{F(x_{2:n})\}\{1-F(x_{2:n})\}^{n-2}f(x_{2:n}),\ -\infty < x_{2:n} < \infty$

Distribution of an Average of First Two Order Statistics

Consider $X_{1:n}$ and $X_{2:n}$ from a random sample of size n. Therefore, the probability density function for $\bar{X}$, $\left(\bar{X} = \frac{X_{1:n}+X_{2:n}}{2}\right)$, can be obtained using convolutions:

$$f_{\bar{X_n}}(\bar{x}) = 2n(n-1)\int_{-\infty}^{\bar{x_n}}[1-F(2\bar{x_n}-x_1)]^{n-2}f(x_1)f(2\bar{x_n}-x_1)\,dx_1 \quad -\infty < \bar{x_n} < \infty \qquad (7.6)$$

Though the underlying distribution of the gross scores is unknown, we can obtain (7.6) for independent and identically distributed $X_i's$ random variables following a uniform $(0,1)$ distribution.

$$f_{\bar{X}_n}(\bar{x}) = \begin{cases} 2n[(1-\bar{x})^{n-1} - (1-2\bar{x})^{n-1}] & \text{if } 0 < \bar{x} < 1/2 \\ 2n(1-\bar{x})^{n-1} & \text{if } 1/2 < \bar{x} < 1 \end{cases} \tag{7.7}$$

So given the underlying distribution of the sample of golf scores, the density function of the handicap index can be obtained by a similar approach. However, these calculations become unwieldy for order statistics more than three in the sample. It is not tractable to obtain, much less use, an analogous expression to (7.7) for the first 10 order statistics as required for the handicap index calculation. A numerical approach is required to retrieve the average of first ten order statistics (i.e., handicap index) which is needed for estimating the handicap.

Now that we understand the handicap system and the algorithm for the handicap calculation, we need a source of golf scores for the numerical scheme. In the next chapter, we will analyze a data set that will give us the information we need for our numerical study.

CHAPTER 8

STATISTICAL ANALYSIS OF THE EMPIRICAL DATA

In order to study the effectiveness of the handicap system, we need golf scores for use in simulations. These can be obtained via exploratory analysis of a data set of gross scores of golfers with varying handicaps. To begin, it is necessary to understand the underlying distribution of the data sets of scores for each of the given handicaps. In this chapter, we will begin by introducing the source of the data set. To understand the data set, we will use descriptive statistics which exhibit the basic features of the data set both summarized graphically as well as numerically. The statistical analysis of the data set provides the detailed fitting information of the underlying family of distribution of the golf scores corresponding to each handicap.

8.1 Analysis of Data Set

An empirical study is conducted with the help of the AT&T Golf Tournament League data set. The data was obtained from Steven Swiss, President of AT&T Golf Tournament League 2004-2005. This data set contains scores up to 10 tournaments for each year plus a special end of year tournament for each of the 9 years 1994 to 2002. This data set contains over 10,800 scores from 84 tournaments for golfers with handicaps ranging from 3 to 46 for games played on 27 different golf courses.

8.1.1 Summary Statistics

In this section, we will look at a number of statistics that are intended to describe the sample and summarize its information. We will use the same set of descriptive statistics as used earlier for warranties in Chapter 5. As a part of preliminary analysis, the descriptives are computed for the given data set and shown in Table 8.1. For every

Table 8.1 Descriptive Statistics of the Scores from the Data Set

						Quartiles		
Handicap	Sample Size	Mean	Std. Dev.	Skewness	Kurtosis	Q_1	Q_2	Q_3
3	4	77.500	2.887	0.000	1.000	75	78	80
4	12	79.833	2.368	0.636	1.486	78	79	83
5	22	80.636	3.303	−0.142	2.281	78	81	82
6	36	81.444	4.890	0.217	2.146	78	81	86
7	78	81.667	4.302	0.772	3.880	78	82	84
8	76	82.342	4.933	0.334	2.614	79	82	86
9	106	84.793	5.675	0.727	3.331	81	84	88
10	130	84.569	4.893	0.879	4.706	81	84	87
11	186	86.742	5.872	0.791	5.774	83	86	90
12	280	87.357	5.513	0.635	3.497	84	87	91
13	318	88.711	6.048	1.382	8.136	84	88	93
14	368	89.147	5.070	0.140	2.423	85	89	93
15	372	91.296	6.472	0.870	4.061	87	90	94
16	486	91.296	5.897	0.448	3.124	87	91	95
17	564	92.036	5.650	0.670	3.445	88	91	95
18	596	93.661	5.973	0.532	3.398	89	93	97
19	554	94.267	5.798	0.548	3.632	90	94	97
20	590	95.861	6.406	−0.025	2.725	92	96	100
21	581	96.463	6.403	0.319	2.952	92	96	101
22	524	97.660	6.597	0.514	3.665	93	97	102
23	519	98.881	6.951	0.582	3.430	94	98	104
24	477	100.065	6.554	0.424	3.505	95	100	104
25	443	100.348	6.206	0.252	3.193	96	100	104
26	423	102.312	7.242	0.992	5.737	98	102	106
27	354	101.901	6.235	0.462	3.921	98	102	106
28	298	103.567	7.728	0.972	6.401	98	103	108
29	276	104.029	6.676	0.295	2.804	99	103	108
30	241	106.701	7.439	0.128	2.741	101	107	112
31	233	106.554	7.398	0.394	3.919	102	106	111
32	169	107.462	7.703	0.135	3.037	103	107	113
33	174	109.328	6.979	0.179	2.545	104	109	114
34	112	109.054	7.690	0.364	3.056	104	109	114
35	162	111.284	7.650	0.862	4.914	106	110	117

Table 8.1 (Continued) Descriptive Statistics of the Scores from the Data Set

						Quartiles		
Handicap	Sample Size	Mean	Std. Dev.	Skewness	Kurtosis	Q_1	Q_2	Q_3
36	129	113.457	9.120	0.635	4.554	107	113	120
37	148	115.216	9.764	0.611	3.482	109	113	119
38	117	114.111	8.178	0.414	3.089	108	114	120
39	115	117.696	9.331	0.333	2.350	110	116	127
40	134	120.373	10.399	0.893	3.530	112	119	126
41	94	120.266	9.007	0.134	2.668	115	120	126
42	36	122.694	7.270	0.116	2.173	118	122	127
43	35	126.400	7.770	0.756	2.816	121	124	132
44	18	126.222	7.134	0.456	1.968	120	125	131
45	5	131.200	14.738	1.086	2.714	120	130	130
46	5	130.000	4.472	1.500	3.250	128	128	128

handicap, the mean lies around the second quartile (median), confirming the location where majority of the scores are concentrated. For example, for handicap 15, the mean is given by 91.296 which is near the value of the median given by 91. The standard deviation indicates that higher the handicaps, the more spread out the data set is likely to be. The differing behavior for handicaps greater than 40 is caused by small sample sizes. It is generally assumed in the golf literature that the scores are normally distributed. This assumption does not apply to the given data because scores for varying handicaps are slightly positive skewed and have fluctuating values of kurtosis (either less than or greater than 3). In each case, the scores are skewed to the right, most values are concentrated to the left of the mean, with extreme values to the right. In addition, majority of the handicaps have kurtosis greater than 3 indicating that most of the scores are concentrated around the mean with thicker tails than for a Gaussian. On the other hand, a few handicaps have kurtosis smaller than 3 which shows that the scores are wide spread around the mean. Overall, the values of skewness and kurtosis indicate that the handicaps are asymmetrical and have peaked distributions as compared to normal distribution.

8.1.2 Graphical Summaries

Visualization techniques are the ways of creating and manipulating graphical representations of data. Graphs provide insight into data under investigation and illustrate important concepts. The section begins with discussion of bivariate plot, i.e., scatterplot. Then box plots and histograms will be shown.

A scatterplot gives a good visual picture of the relationship between the two variables, and aids the interpretation of the correlation coefficient or regression model. Figure 8.1 contains a scatterplot of gross scores versus handicaps. In this plot, an increasing trend in gross scores is observed as the handicap increases, as would be expected. The spread of the scores indicates increasing variability with increase in handicaps. It is also observed that for the smallest and largest handicap values, the data is sparse. Hence, our further analysis only consider the data for handicaps 10 through 40.

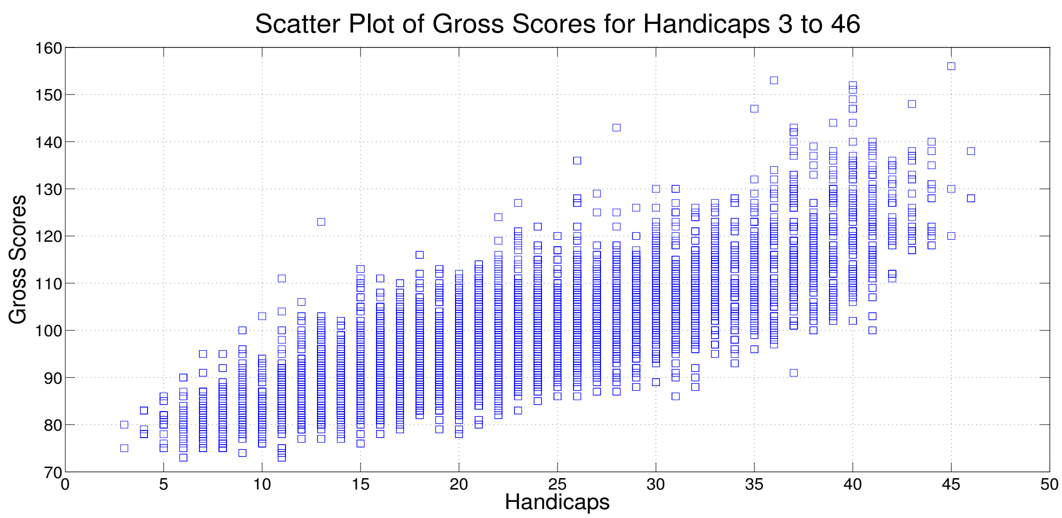


Figure 8.1 Scatter plot of gross scores from the original data set for handicaps 3 through 46.

Another way of presenting the graphical summary is via histograms, which combine data into groups or classes as a way to exhibit the details of a data set while at the same time illustrating the data's overall pattern. On a histogram, the x-axis represents the data values arranged into classes while the y-axis shows the number of occurrences in each class. Figure 8.2 contains the histograms only for every third handicap for brevity. It is observed that with increase in handicap, the mean in scores increases as indicated by the horizontal shift in the plots in the right. The increase in variability can be observed by the increasing width of the histogram. Moreover, we can also infer that bulk of the data is at the left and right tail is longer which confirms that the data is positively skewed. With reference to kurtosis values in Table 8.1, we observe that handicaps having high value of kurtosis tend to have distinct peak near the mean, decline rapidly and have heavy tails. Hence, Figure 8.2 observes similar trends as observed in Table 8.1.

Another way of presenting a graphical summary is a box plot which is especially useful for the detection of outliers. Figure 8.3 shows the box plots for handicaps from 10 to 40. Similar to the trend in histograms, as in Figure 8.2, the variability in scores is observed with the vertical shift in the box plots with increase in handicap. The presence of outliers in the plots indicates atypical behavior in the data.

8.1.3 Data Cleaning

Data cleaning (cleansing, scrubbing, etc.) is a process of identifying and repairing the existing incorrect or corrupt records in a data set. Although outliers are often considered as an error or noise, they may carry important information. The goal is not only to bring the data set into a consistent state but also to ensure an accurate and complete representation of the real-world constructs to which the data refer, in this instance, to golf scores. The box plots in the Figure 8.3 have been useful in visually locating the outliers which are plotted separately as points. Following Tukey

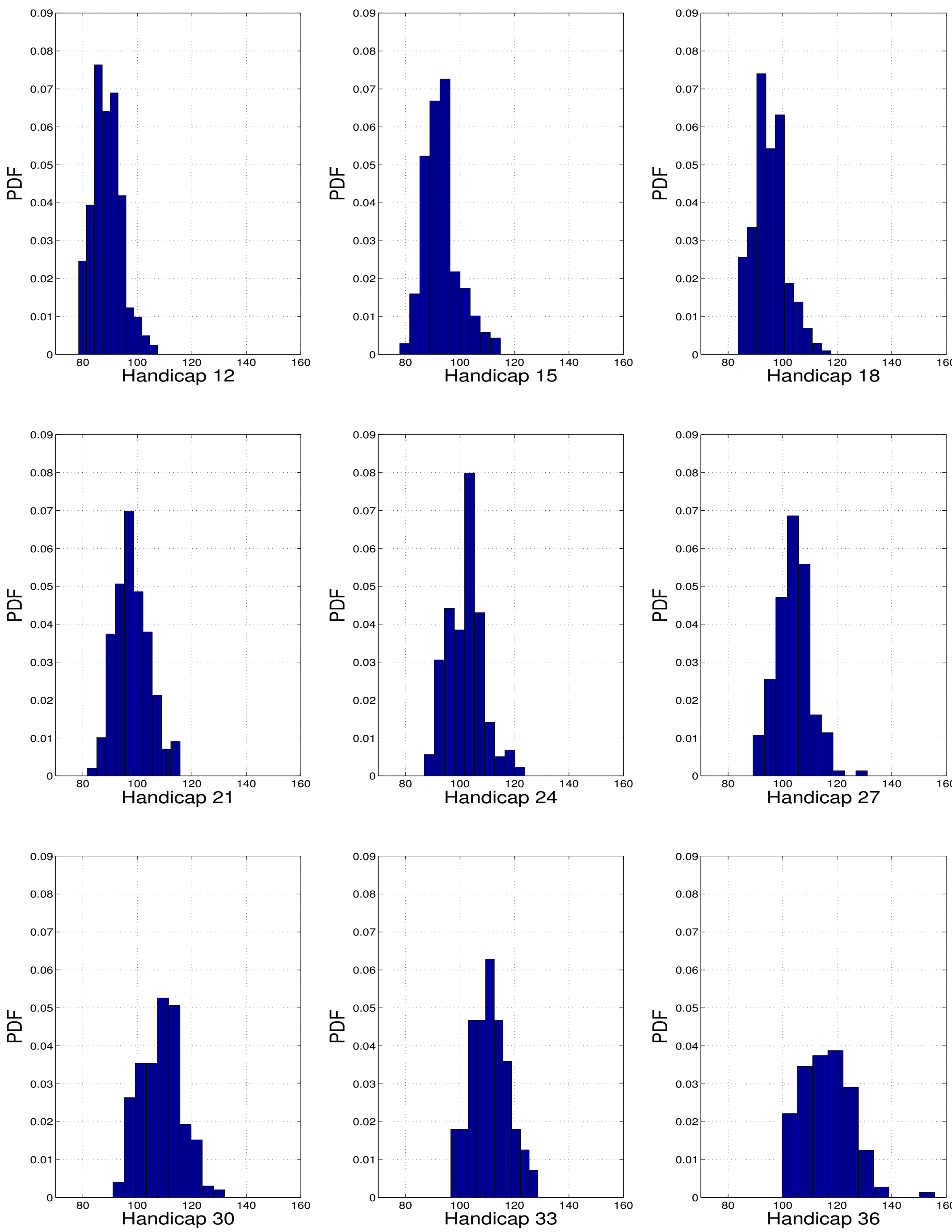


Figure 8.2 Histograms of gross scores from original data sets for selected handicaps.

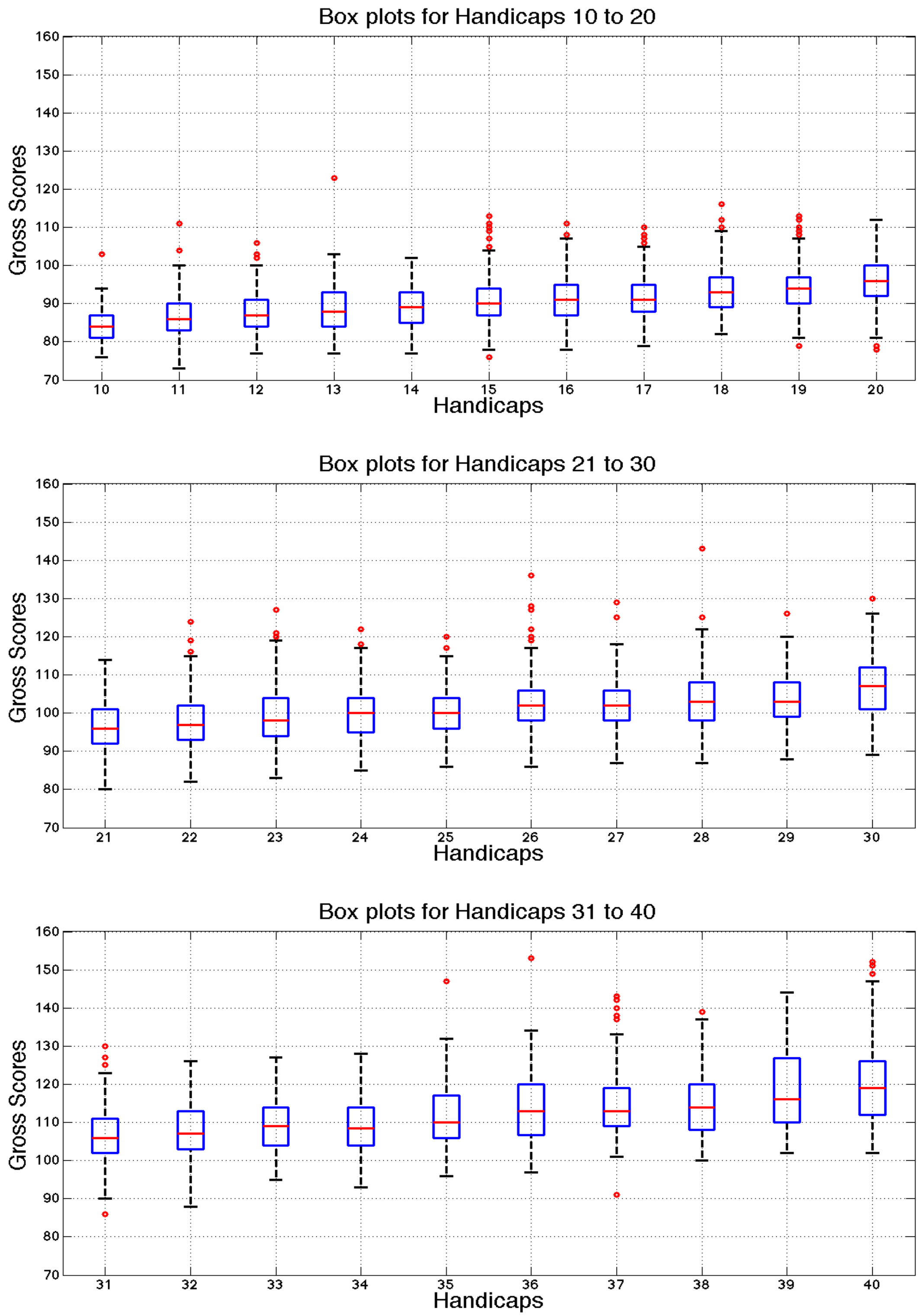


Figure 8.3 Box plots of handicaps from 10 to 40.

[37], the removal of these observations is done and the resulting data are used for making statistical inference. After the data cleaning is completed, the scatter plot of filtered data of gross scores can be compared with the original data set as shown in Figure 8.4. Clearly, the outliers that are prominent in the scatterplot of original data are no longer present in the plot for filtered data. Statistically, the variability in the scores is reduced in each case.

Table 8.2 gives the descriptive statistics for the data set after the filtering. In comparison with Table 8.1, the effect of removal of outliers is prominent in the change of the numerical values of skewness and kurtosis in Table 8.2. For each of the handicaps, the decrease in value of the skewness indicates that the data values are getting clustered towards the center. Hence, the variability in the data is effectively reduced. Filtering causes the value of kurtosis to become less than 3 for majority of the data sets. This means that the distribution of scores has flattened peaks and slender tails. The next step in data analysis is to use the information from a sample to draw inferences concerning the population from which the sample was drawn.

8.1.4 Selection of Model

Most of the literature in golf research assumes the normal distribution as the underlying distribution of golf scores because of the central limit theorem [57]. Since both the original and fitted data sets are skewed and have kurtosis values other than 3, we need appropriate statistical fitting tests to find the underlying distribution. Many parameter estimation and goodness of fit tests that are useful in model building have been discussed in detail in Chapter 5. An approach to model selection includes preparing a list of candidate distributions, using methods of parameter estimation to evaluate the parameters of the distributions, and performing the goodness of fit tests to obtain the parameter values for each distribution. The selected distributions are compared based on the results obtained from fitting procedure.

Table 8.2 Descriptive Statistics of the Scores for Each of the Handicaps After Filtering

						Quartiles		
Handicap	Sample Size	Mean	Std. Dev.	Skewness	Kurtosis	Q_1	Q_2	Q_3
10	128	84.281	4.345	0.267	2.590	81	84	87
11	182	86.286	5.024	−0.131	3.331	83	86	90
12	272	86.904	4.894	0.238	2.578	84	87	90
13	316	88.494	5.412	0.443	2.563	84	88	92
14	368	89.147	5.070	0.140	2.423	85	89	93
15	346	90.266	4.898	0.254	2.903	87	90	93
16	474	90.903	5.409	0.175	2.589	87	91	94
17	552	91.685	5.175	0.418	2.926	88	91	95
18	588	93.412	5.607	0.282	2.732	89	93	97
19	534	93.798	5.040	0.190	2.833	90	94	97
20	586	95.980	6.264	0.062	2.587	92	96	100
21	581	96.463	6.403	0.319	2.952	92	96	101
22	514	97.272	6.018	0.102	2.528	93	97	101
23	508	98.429	6.291	0.187	2.327	93	97	101
24	469	99.725	6.059	0.087	2.752	95	100	104
25	434	100.000	5.737	−0.069	2.634	96	100	104
26	405	101.553	5.780	0.017	2.627	98	102	105
27	352	101.759	5.955	0.170	2.880	98	102	106
28	294	103.163	6.858	0.117	2.479	98	103	107
29	275	103.949	6.555	0.212	2.576	99	103	108
30	240	106.604	7.300	0.040	2.533	101	107	112
31	222	105.946	6.121	−0.015	2.922	102	105	110
32	169	107.462	7.703	0.135	3.037	103	107	113
33	174	109.328	6.979	0.179	2.545	104	109	114
34	112	109.054	7.690	0.364	3.056	104	109	114
35	160	110.931	6.958	0.293	2.425	106	110	116
36	128	113.148	8.451	0.105	2.428	107	113	119
37	140	114.486	8.026	0.477	2.632	109	113	119
38	116	113.897	7.876	0.274	2.773	108	114	120
39	115	117.696	9.331	0.333	2.350	110	116	127
40	129	119.256	8.859	0.512	2.581	112	119	125

Preparing the data for distribution fitting is one of the most important steps taken, since the analysis of the results depend on whether the data is correctly collected and specified. The most commonly used format in probability data analysis is an unordered set of values obtained by observing some random process. The order of values in a data set is not important and does not affect the distribution fitting results. This is one of the fundamental differences between distribution fitting (and probability data analysis in general) and time series analysis where each data value is connected to some time point at which this value was observed.

The standard rule of thumb is that the number of observations in the data set should be more than 100 in order to perform the fitting procedure correctly. Very large samples (tens of thousands of data points) might cause some computational problems when fitting distributions to the data, and there might be a need to reduce the sample size by selecting a subset of that data. However, in our case the data sets are of reasonable size.

Once the goodness of fit tests are applied to the given data sets, the final results include the list of distributions along with their ranks and test statistic values. The test statistic values are calculated for various significance levels (α) and the acceptance of the null hypothesis is tested for each of those levels. Since the goodness of fit test statistics indicate the distance between the data and the fitted distributions, the distribution with the lowest test statistic is assigned rank of 1 with the next best model assigned rank of 2, etc., to compare the fitted models and select the valid model. The Anderson-Darling test was used here as the goodness-of-fit test due to its power in detecting discrepancies in the tails and the fact that it does not depend upon grouping of the data.

In our work, we considered 55 distributions in the fitting procedure. For each of these distributions, the associated parameters were estimated based on the data. Methods used here include the method of moments (MOM), maximum likelihood estimates (MLE), least square estimates (LSE) and method of L-moments. (Each of these methods were discussed in Chapter 5). The method of estimation of parameters of the distribution associated with the scores depends on the nature of the fitted distribution. Hence, different distributions require different methods of estimation. For example, exponential, two-parameter gamma, and logistic models that use method of moments. However, for many other distributions, the Method of Moments does not yield closed form expressions for parameter estimates; and in such cases the Maximum Likelihood Estimation (MLE) method is often used.

Next, we will discuss the results obtained from the fitting procedure. The fitted distributions are arranged in increasing order of test statistics ranging from smallest to largest. Among all the candidate distributions, the one with lowest test statistic is selected as the most appropriate fitted model. For the sake of brevity, we have only included some of the most well known distributions in Table 8.3. In order to make the selection simpler, we have calculated overall average rank across the fitted distributions for each handicap. It is observed that the generalized extreme value distribution (GEV distribution) has the minimum overall average rank for both the raw and filtered data sets of scores even among distributions not included in Table 8.3.

Table 8.3 Overall Average Rank for Popularly Used Distributions

Distribution	Raw Data	Filtered Data
Generalized Extreme Value	8.5	7.1
Lognormal (3P)	10.4	10.9
Gen. Gamma (4P)	11.2	11.7
Beta	11.9	11.9
Gamma (3P)	11.9	12.7
Gamma	19.6	16.5
Normal	25.4	20.9

Table 8.4 compares the ranks of the GEV and normal distributions for both the raw and filtered data sets as an example of the behavior of the filtering procedure.

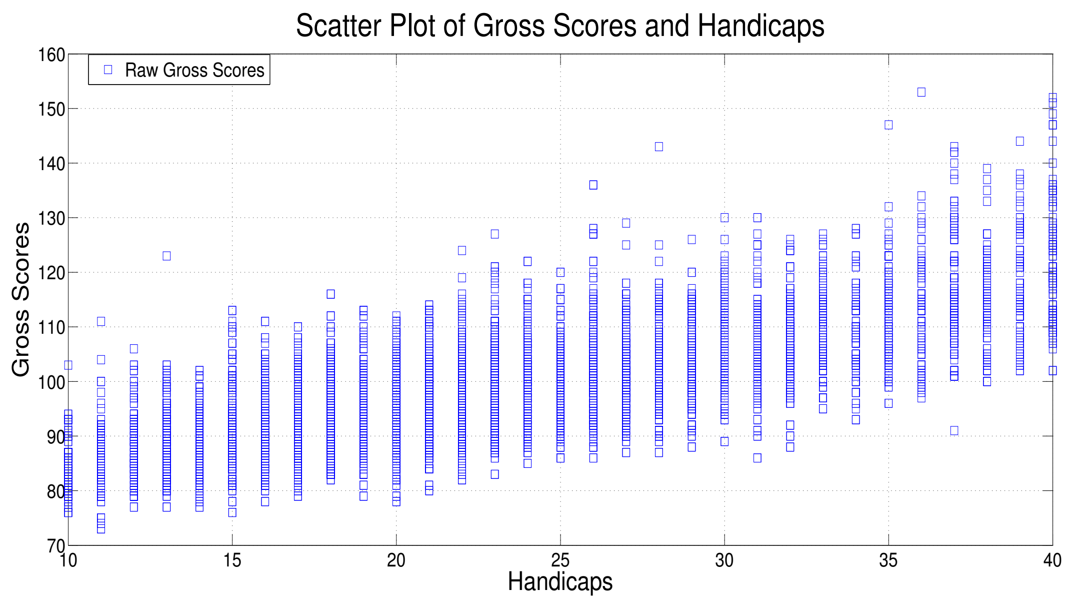
We observe that filtering lowered the overall average rank for GEV from 8.5 to 7.1. It is also noted that the the normal distribution has high overall average rank in comparison with GEV. However, there are a few handicaps for which the rank of GEV was worse than the normal distribution. But, since the overall rank is much lower for GEV in comparison to the normal distribution, GEV distribution was selected as the appropriate distribution.

In general, the Generalized Extreme Value distribution consists of three parameters namely, shape, scale and location [59]. The scale must be positive, the shape and location can take any real value. Table 8.5 gives a complete list of parameters of GEV distribution for each of the handicaps for both the raw and filtered data sets. On comparing the parameter values for the original and filtered data, we find that the location parameter is almost the same in value. However the values of the scale and shape parameters for fitted data are different from the original data. In addition, all the values for the shape parameter for the fitted version are more negative as compared to original data.

Figure 8.7 displays the scatter plots for original, filtered and fitted data in order to visually compare the effect of filtering and fitting on the data in comparison to the original data. The scatter plot of raw data clearly shows the variance of the scores is greater in comparison to the other plots. After filtering, the pattern of the spread of scores confirms lowering of the variance. In order to obtain a similar scatterplot for the fitted data, random samples of scores for each handicap are generated using the parameters from Table 8.5. We observe that the plot captures similar trend to that of the filtered version. Hence, these plots provide important graphical summary helpful in understanding the effect of procedures used in the data analysis.

Table 8.4 Ranking of the Generalized Extreme Value and Normal Distributions for Handicaps from Original and Filtered Data Sets

	GEV Distribution		Normal Distribution	
Handicap	Original	Filtered	Original	Filtered
10	2	3	36	28
11	16	27	29	7
12	2	1	32	25
13	8	4	30	33
14	7	7	26	26
15	4	14	38	28
16	4	3	30	21
17	5	2	35	29
18	3	1	27	26
19	13	18	32	22
20	1	2	7	10
21	9	9	27	27
22	5	4	27	25
23	6	6	28	23
24	21	8	12	4
25	23	4	8	9
26	10	7	30	1
27	20	16	26	21
28	8	3	28	25
29	1	1	29	28
30	5	5	3	8
31	15	17	31	5
32	23	23	29	29
33	2	2	24	24
34	14	14	30	30
35	5	4	30	30
36	13	3	9	6
37	4	2	32	29
38	9	4	6	5
39	5	5	27	27
40	1	2	30	36
Average	8.5	7.1	25.4	20.9



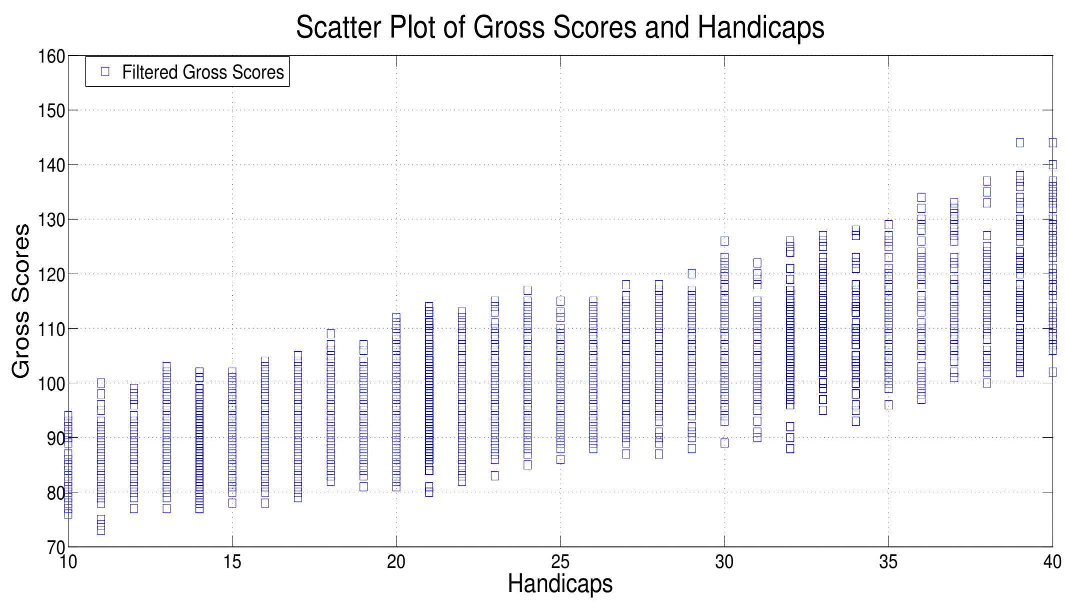


Figure 8.4 Comparing the scatter plots of gross scores for handicaps 10 to 40 from original and filtered data.

Table 8.5 Estimated Values of Parameters of the Fitted Generalized Extreme Value Distribution for Handicaps

	Original Data			Filtered Data		
Handicap	Shape	Scale	Location	Shape	Scale	Location
10	−0.103	4.201	82.536	−0.200	4.139	82.588
11	−0.171	5.168	84.517	−0.325	5.046	84.64
12	−0.110	4.851	85.039	−0.205	4.694	85.001
13	−0.042	4.858	86.099	−0.108	4.844	86.169
14	−0.230	4.960	87.222	−0.230	4.960	87.222
15	−0.022	5.152	88.434	−0.191	4.600	88.353
16	−0.156	5.403	88.908	−0.227	5.252	88.854
17	−0.073	4.792	89.592	−0.135	4.667	89.547
18	−0.146	5.410	91.229	−0.198	5.328	91.224
19	−0.137	5.171	91.908	−0.231	4.880	91.909
20	−0.287	6.246	93.613	−0.264	6.226	93.708
21	−0.181	5.995	93.929	−0.181	5.995	93.929
22	−0.158	6.022	95.010	−0.237	5.895	95.014
23	−0.127	6.232	95.987	−0.208	6.084	95.973
24	−0.212	6.191	97.586	−0.289	6.075	97.608
25	−0.242	6.008	98.067	−0.324	5.871	98.069
26	−0.116	6.170	99.391	−0.289	5.801	99.531
27	−0.192	5.796	99.495	−0.232	5.761	99.530
28	−0.143	6.809	100.490	−0.236	6.718	100.590
29	−0.167	6.215	101.340	−0.183	6.190	101.340
30	−0.268	7.389	104.020	−0.287	7.354	104.03
31	−0.171	6.700	103.670	−0.277	6.074	103.780
32	−0.218	7.357	104.540	−0.218	7.357	104.540
33	−0.228	6.809	106.680	−0.228	6.809	106.680
34	−0.173	7.136	105.990	−0.173	7.136	105.990
35	−0.085	6.540	108.020	−0.153	6.463	108.060
36	−0.194	8.494	109.950	−0.260	8.246	110.050
37	−0.056	8.156	110.940	−0.097	7.094	111.020
38	−0.200	7.757	110.940	−0.235	7.686	110.940
39	−0.160	8.748	113.860	−0.160	8.748	113.860
40	0.021	8.093	115.530	−0.072	7.686	115.330

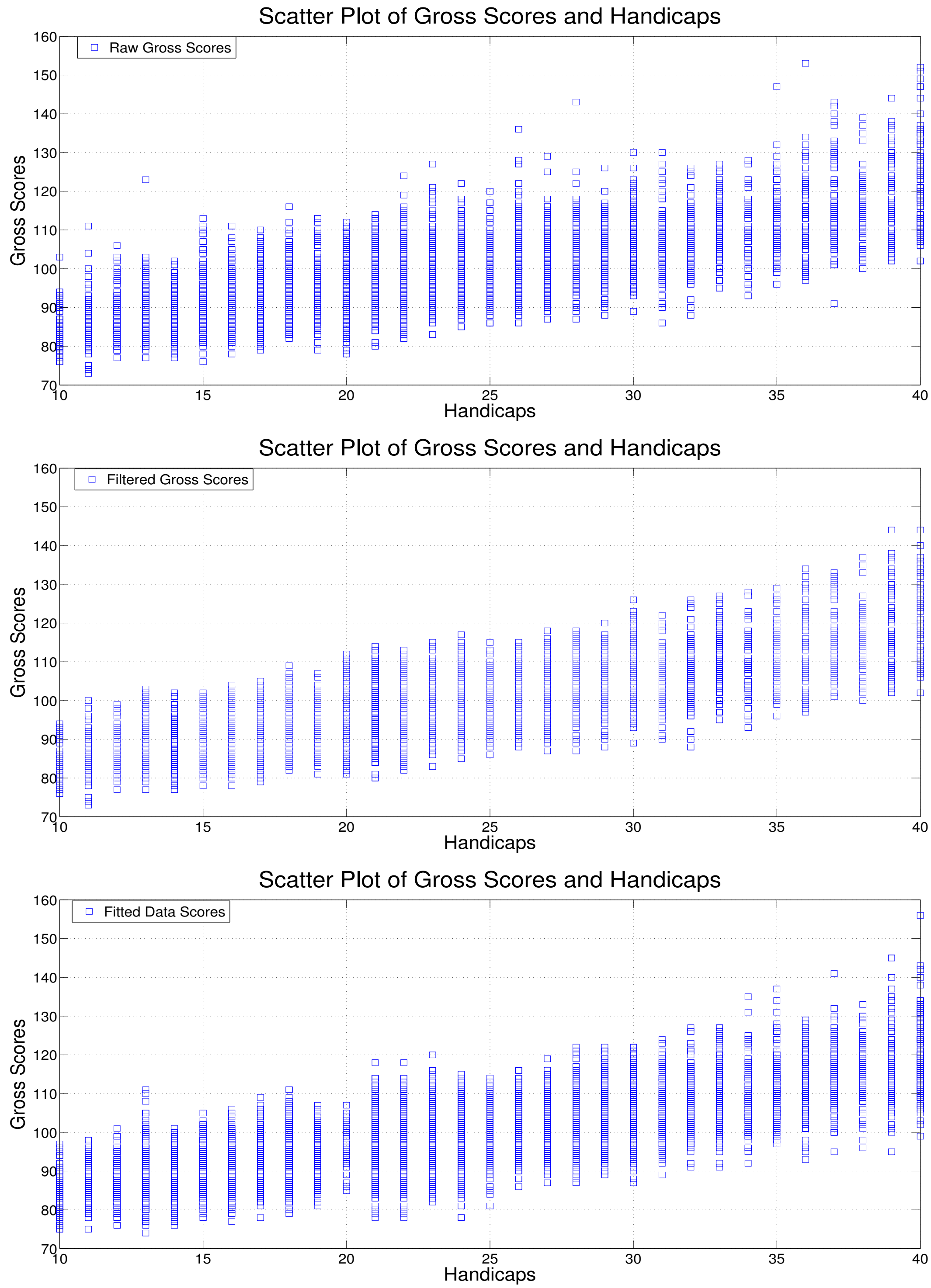


Figure 8.5 Comparing the scatter plots of gross scores for handicaps 10 to 40 from original, filtered and fitted data.

Data filtering and fitting have been used to introduce smoothness by reducing variability in the scores. But there are still fluctuations in the parameter values of GEV distribution. To obtain smooth results, regression is a popularly used strategy to construct a fit through scatter of points. We will consider linear regression in order to determine models based on the parameters of the fitted distribution.

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data [58]. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. Before attempting to fit a linear model to observed data, it is important to determine whether or not there is a relationship between the variables of interest. This does not necessarily imply that one variable causes the other but that there is some significant association between the two variables. A scatterplot is a helpful tool in determining the strength of the relationship between two variables. If there appears to be no association between the proposed explanatory and dependent variables (i.e., the scatterplot does not indicate any increasing or decreasing trends), then fitting a linear regression model to the data probably will not provide a useful model. A valuable numerical measure of association between two variables is the coefficient of determination (R-squared) which indicates the proportionate amount of variation in the response variable explained by the independent variables in the linear regression model. The closer to 1 the R-squared is, the more variability is explained by the linear regression model. Once the model is obtained, the validity is checked by looking at the probability value in the regression procedure. According to this procedure, the regression model fits well when the probability value is less than the significance level of $\alpha = 0.05$.

It is known that the fitted distribution for both the original and filtered data sets is GEV distribution. The distribution has parameters such as shape, scale and location. Each of the parameters is plotted as a function of handicap and a linear fit is attempted in order to explain the relationship between the parameter.

Figure 8.6 is the scatter plot of the original values of the shape parameter. Visually, there does not appear to be a strong trend in the data. Nonetheless, linear regression was applied to the data. The best linear model for the shape parameter is - 0.0006 $\times$ handicap - 0.14. However, the probability value is 0.6797 which is more than the significance level of 0.05 indicating that the linear regression model is not fit appropriate for the shape parameter. Additionally, the R-squared value is 0.0060 which indicates that the inappropriateness of the linear fit.

Figure 8.7 is the scatter plot of the original values of the scale parameter. Using linear regression, the best linear model for the scale parameter is 0.12 $\times$ handicap + 3.2. The probability value is $<$0.0001 which is less than the significance level of 0.05 indicating that the regression model is an appropriate fit to the values of the parameter. Additionally, the R-squared value is 0.8711 which is high enough indicating the variability in the scale parameter is highly affected by variability in the handicap values.

Figure 8.8 is the scatter plot of the original values of the location parameter. Using linear regression, the best linear model for the parameter is handicap + 73. The probability value is $<$0.0001 which is less than the significance level of 0.05 indicating that the regression model is an appropriate fit to the values of the parameter. Additionally, the R-squared value is 0.9946 very high which is indicating the variability in the location parameter is very highly affected by variability in the handicap values.

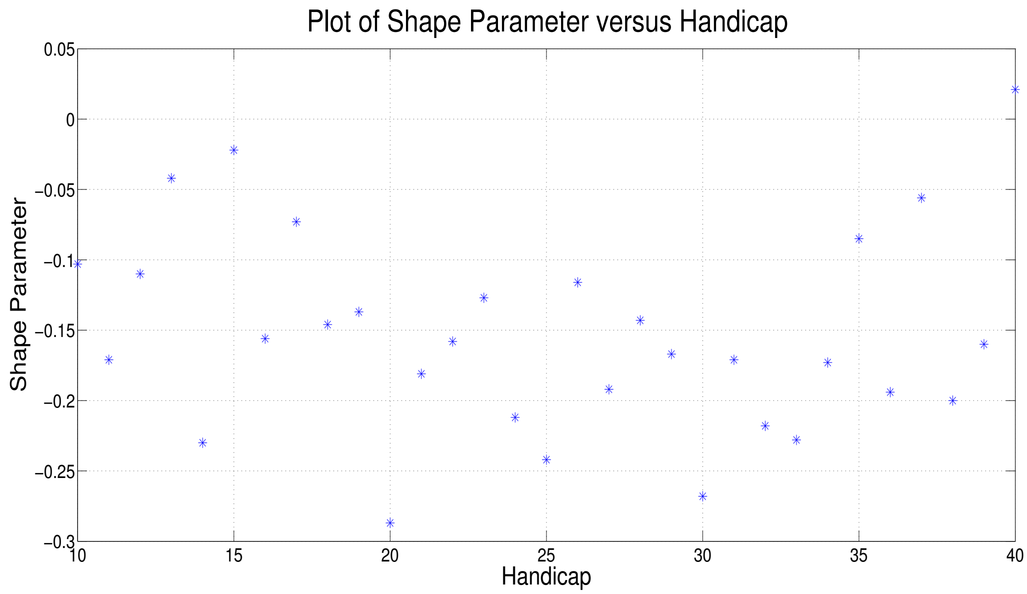


Figure 8.6 Scatter plot of values from the shape parameter of the generalized extreme value distribution.

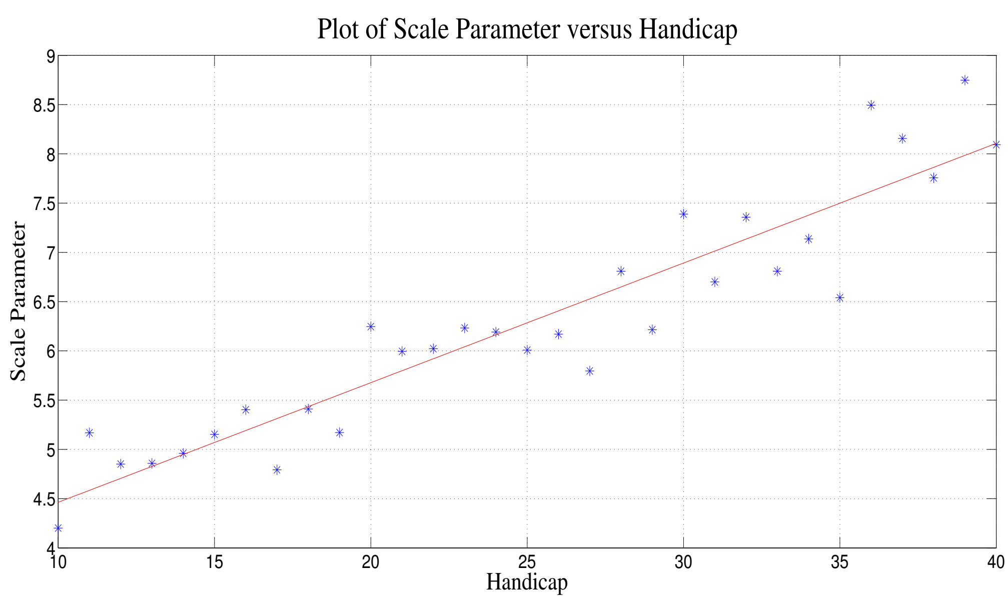


Figure 8.7 Fitting regression model to the data of values from the scale parameter of the generalized extreme value distribution. The regression model is given by scale = $0.12 \times$ handicap + 3.2.

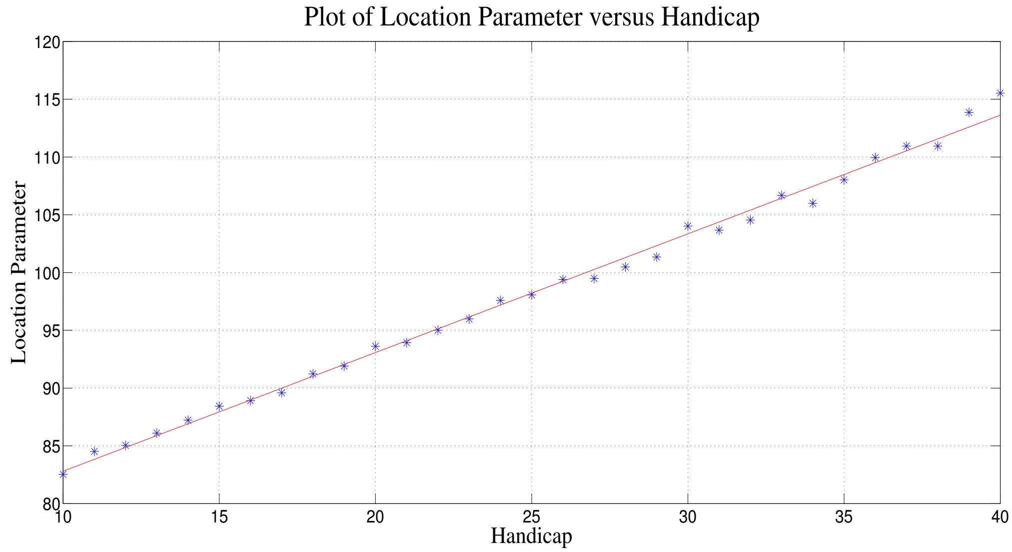


Figure 8.8 Fitting regression model to the data of values from the location parameter of the generalized extreme value distribution. The regression model is given by location = handicap + 73.

From the regression models for scale and location parameters, we will obtain the fitted values of both the parameters. Since the behavior of the shape parameter is important in the distribution yet does not follow a linear regression model, we divide our analysis into three cases : (a) we use original values of shape parameter and regression fitted values of other two, (b) we use values of the shape parameter calculated using a moving average technique over 5 adjoining values and regression fitted values of the other two parameters and (c) we use the overall mean value of shape parameter (-0.153) and regression fitted values of the other two parameters. Table 8.6 is divided into three parts displaying the GEV distribution parameter values for original, original/regression and moving average/regression.

Table 8.6 Estimated Values of Parameters of the Fitted Generalized Extreme Value Distribution to the Scores of the Handicaps

	Original			Original/Regression			Moving Average/Regression		
Handicap	Shape	Scale	Location	Shape	Scale	Location	Shape	Scale	Location
10	-0.103	4.201	82.536	-0.103	4.400	83	-0.128	4.400	83
11	-0.171	5.168	84.517	-0.171	4.520	84	-0.107	4.520	84
12	-0.110	4.851	85.039	-0.110	4.640	85	-0.131	4.640	85
13	-0.042	4.858	86.099	-0.042	4.760	86	-0.115	4.760	86
14	-0.230	4.960	87.222	-0.230	4.880	87	-0.112	4.880	87
15	-0.022	5.152	88.434	-0.022	5.000	88	-0.105	5.000	88
16	-0.156	5.403	88.908	-0.156	5.120	89	-0.125	5.120	89
17	-0.073	4.792	89.592	-0.073	5.240	90	-0.107	5.240	90
18	-0.146	5.410	91.229	-0.146	5.360	91	-0.160	5.360	91
19	-0.137	5.171	91.908	-0.137	5.480	92	-0.165	5.480	92
20	-0.287	6.246	93.613	-0.287	5.600	93	-0.182	5.600	93
21	-0.181	5.995	93.929	-0.181	5.720	94	-0.178	5.720	94
22	-0.158	6.022	95.010	-0.158	5.840	95	-0.193	5.840	95
23	-0.127	6.232	97.586	-0.127	5.960	96	-0.184	5.960	96
24	-0.212	6.191	98.076	-0.212	6.080	97	-0.171	6.080	97
25	-0.242	6.008	98.067	-0.242	6.200	98	-0.178	6.200	98
26	-0.116	6.170	99.391	-0.116	6.320	99	-0.181	6.320	99
27	-0.192	5.796	99.495	-0.192	6.440	100	-0.172	6.440	100
28	-0.143	6.809	100.490	-0.143	6.560	101	-0.177	6.560	101
29	-0.167	6.215	101.340	-0.167	6.680	102	-0.188	6.680	102
30	-0.268	7.389	104.020	-0.268	6.800	103	-0.193	6.800	103
31	-0.171	6.700	103.670	-0.171	6.920	104	-0.210	6.920	104
32	-0.218	7.357	104.540	-0.218	7.040	105	-0.212	7.040	105
33	-0.228	6.809	106.680	-0.228	7.160	106	-0.175	7.160	106
34	-0.173	7.136	105.990	-0.173	7.280	107	-0.180	7.280	107
35	-0.085	6.540	108.020	-0.085	7.400	108	-0.147	7.400	108
36	-0.194	8.494	109.950	-0.194	7.520	109	-0.142	7.520	109
37	-0.056	8.156	110.940	-0.056	7.640	110	-0.139	7.640	110
38	-0.200	7.757	110.940	-0.200	7.760	111	-0.118	7.760	111
39	-0.160	8.748	113.860	-0.160	7.880	112	-0.099	7.880	112
40	0.021	8.093	115.530	0.021	8.000	113	-0.113	8.000	113

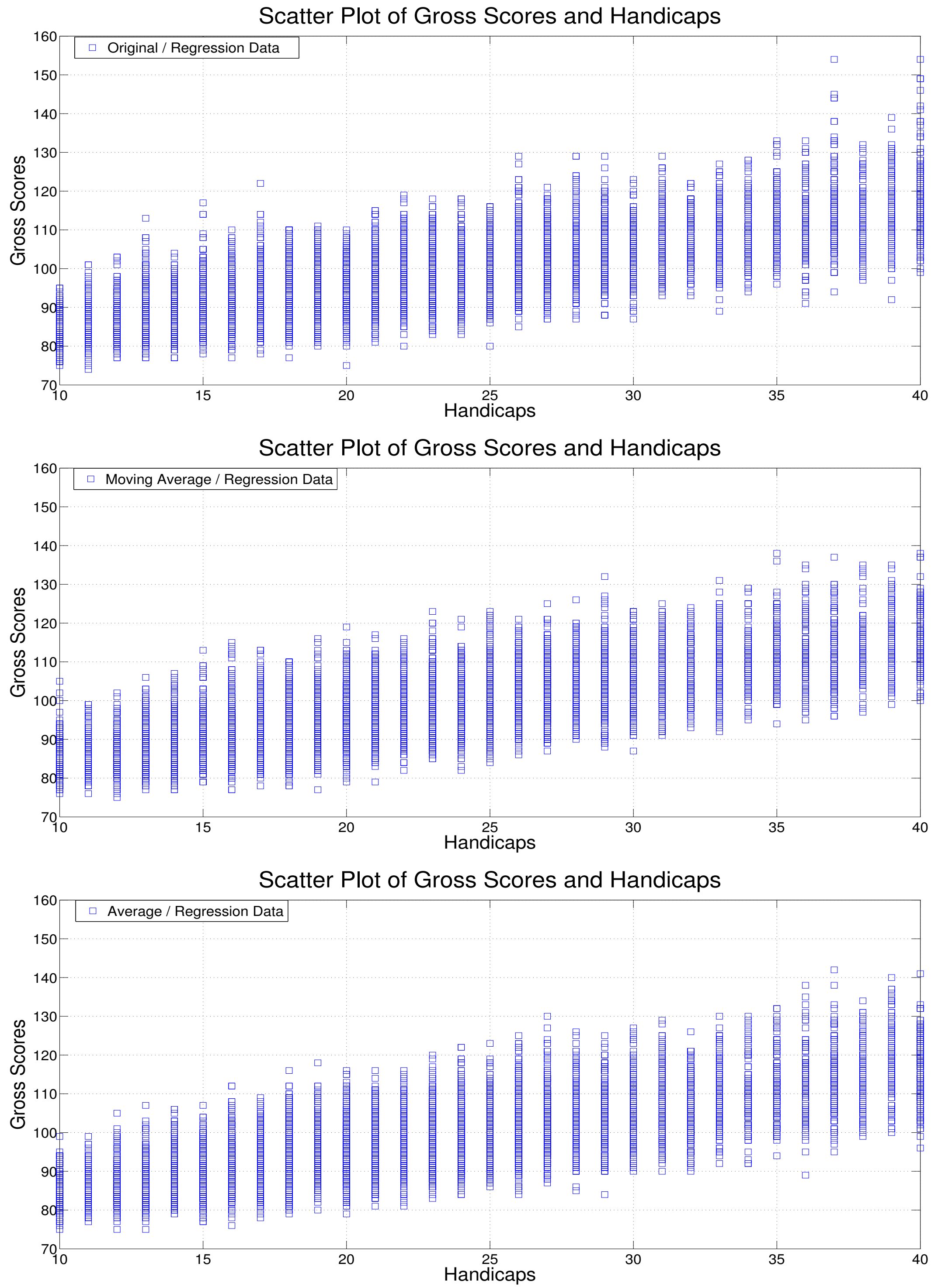


Figure 8.9 Comparing the scatter plots of gross scores for handicaps 10 to 40 depicting the three cases based on the values of the shape parameter.

Figure 8.9 contains the scatter plots for the data obtained using the regression fitted values of scale and location parameters in combination with original, moving average and average values of shape parameter. The plots based on the moving average/regression and average/regression categories have data clustered more towards center (an indicator of reduced variance) as compared to that observed in original/regression. In comparison with the scatter plot based on the fitted data shown in Figure 8.5, the plot for original/regression data matches closely as both have similar trend in the spread of data. But in comparison to the plots for moving average/regression and average/regression data, the plot for fitted data shows greater spread in terms of extreme observations. Hence, we observe that the variance in the scores is further reduced by using the averaging techniques for shape parameter in combination with regression.

8.2 Concluding Remarks

Numerical and graphical summaries have been used to describe the data spread. We have used data filtering and fitting techniques in order to reduce the variability in the golf scores. Further, we obtained a family of target distributions (GEV distribution) via the fitting procedure to help in understanding the behavior of the underlying distribution of the data. For further smoothing we used a regression procedure to fit appropriate linear model to the parameters of GEV distribution. Having this preliminary data analysis completed, we will now use this information to perform a simulation study of golf.

CHAPTER 9

SIMULATION OF GOLF HANDICAPS

Now that we have completed the statistical analysis of the data, we are ready to move forward using the data as an input into an algorithm for the handicap calculation. This is accomplished by using samples of gross scores of the golfers as input to the simulation system and then observing the behavior of the handicap indices. In this chapter, we will begin introducing the simulation procedure in detail. Further using the adopted simulation technique, the qualitative behavior of the output is analyzed for the empirical, filtered and fitted data which helps us to understand the effect of filtering and fitting.

9.1 Simulation Procedure

In Chapter 8, the exploratory analysis of the data helped us to convert the original form of data into different forms like fitted or filtered. Since statistical methods involving order statistics are generally more complicated than those based on the use of unordered data, to estimate the handicap index we will make use of Monte Carlo simulation approach. Monte Carlo methods [60] are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results by running simulations many times in order to obtain estimates of quantities, especially those for which is difficult or impossible to obtain an expression in closed form. The simulation study is conducted by using scores of the sample golfers as basic inputs. We will also study the effect of a sudden change in the distribution of the scores on the handicap.

The sequence of golf scores used in these simulations will be generated using:

- the normal distribution with assumed mean and standard deviation
- the empirical distributions of scores from the data set
- the fitted distributions of scores based upon the data set
- the fitted distributions of scores based upon the filtered version of the data set.

The purpose of including the simulation output for the case of normally distributed golf scores is to compare the sensitivity in the response of the handicap index to the form of the input data, despite our earlier data analysis which included that golf scores do not follow a normal distribution.

9.1.1 Simulation of the Scores and Handicap Indices

The algorithm of the simulation procedure for the handicap calculation follows the exact steps and formula as discussed in Chapter 7. As part of the simulations, for each simulation run we will generate scores as random samples from the handicaps 10 through 40. These scores are adjusted with respect to the handicap. Based on the moving window technique, the 20 most recent scores are used as input for calculating handicap differentials. As a result, we obtain handicap index. For our study, the values for course rating and slope rating are 70 and 113, respectively. The result is displayed graphically in terms of a plot between the scores and the handicap indices.

Let's consider the first case stated in the simulation procedure that the gross scores follow a normal distribution with mean $\mu = 85$ and standard deviation $\sigma = 2$. Figure 9.1 is the plot of the resulting handicap indices from the simulation of scores. It is observed that there is little variability in the pattern of handicap indices, which is true for a symmetric distribution.

As the next step in the simulation procedure, let us consider the golf scores from an empirical distribution of selected handicaps. As an example, we will show

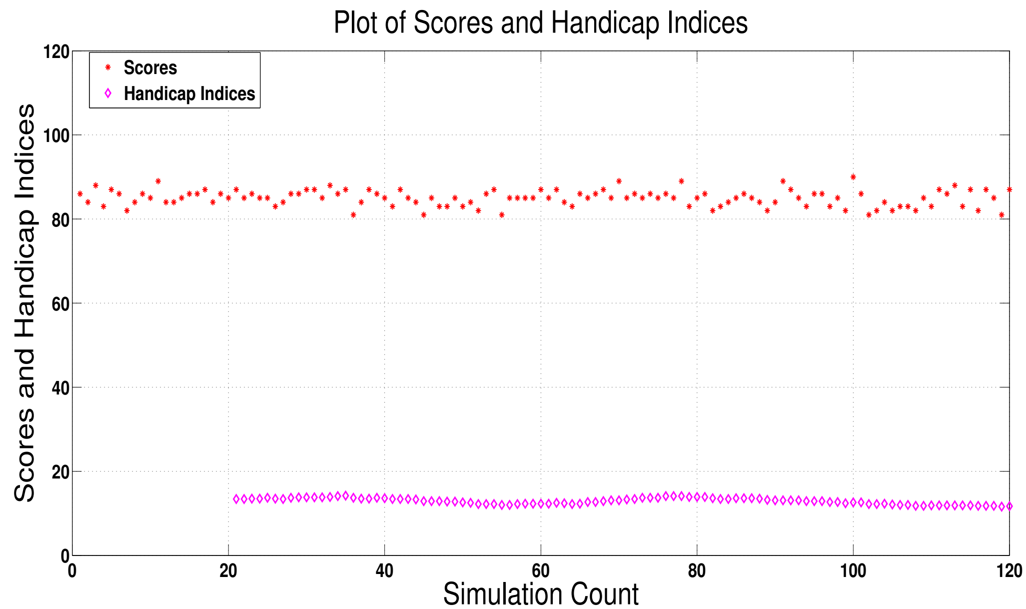


Figure 9.1 Simulation of scores from normal distribution with mean $\mu = 85$ and standard deviation $\sigma = 2$ and corresponding response of handicap indices is observed.

the results for handicap 13. Figure 9.2 is the normalized probability density function of the scores from the data. We observe that the given data set is positively skewed and contains outliers on the high end of data. Using the parameters of handicap 13, we analyze the output graphically as displayed in Figure 9.3. After comparing Figures 9.1 and 9.3, the output for an asymmetrical distribution has fluctuating output indicating variability in the data. To stabilize it we will proceed with the case of filtered data set.

Figure 9.4 is the normalized histogram for the filtered data for handicap 13. It is observed that the variance in the data is reduced thereby reducing the spread as indicated by no gaps between the bins. Now, the parameters of the algorithm are updated according to the criteria and the output is observed as in Figure 9.5. The trends in the simulation output in Figures 9.3 and 9.5 are almost similar with minor reduction in the fluctuation in the handicap indices, an indicator of reduced variance.

Now, we will conduct simulations by using the information of the fitted GEV distribution to the scores of original data set of handicap 13. From Table 8.5, the estimated parameters for handicap 13 are shape = -0.042, scale = 4.858 and location = 86.099. The histogram of the fitted distribution is shown in Figure 9.6. Updating the algorithm, the simulations are replicated and the output is observed in Figure 9.7. The response of the handicap indices to the simulated scores is similar to that in Figure 9.5.

Furthermore, considering the same fitting procedure for the filtered data set of gross scores for handicap 13. In the Table 8.5, the parameters of the filtered scores from handicap 13 are given by shape = -0.108, scale = 4.844 and location = 86.169. The normalized histogram of the fitted distribution of the filtered data set is shown in Figure 9.8. Replicating the simulation procedure using this choice of parameters, the output is observed in Figure 9.9. We observe that the output is similar to that shown in Figures 9.5 and 9.7.

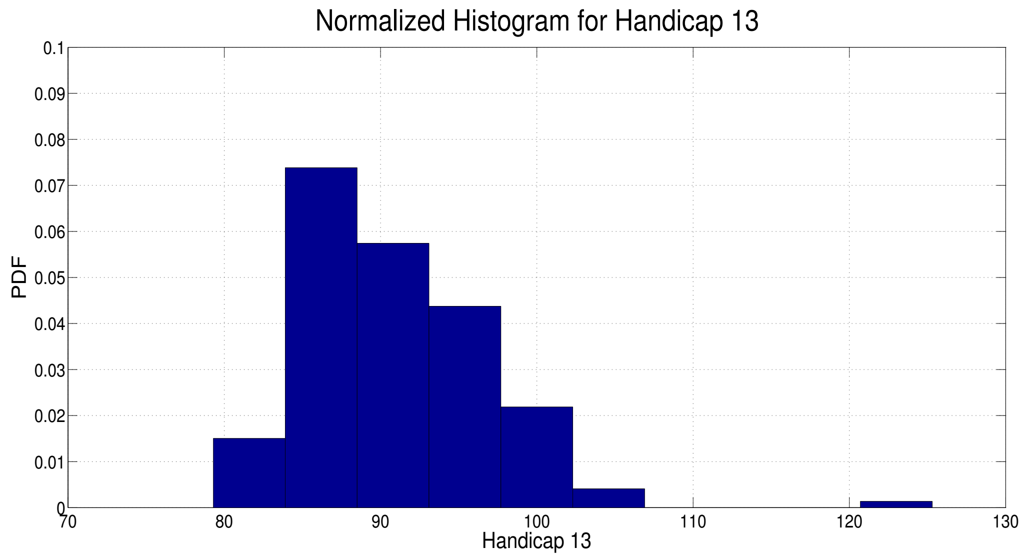


Figure 9.2 Histogram of original data set of scores for handicap 13.

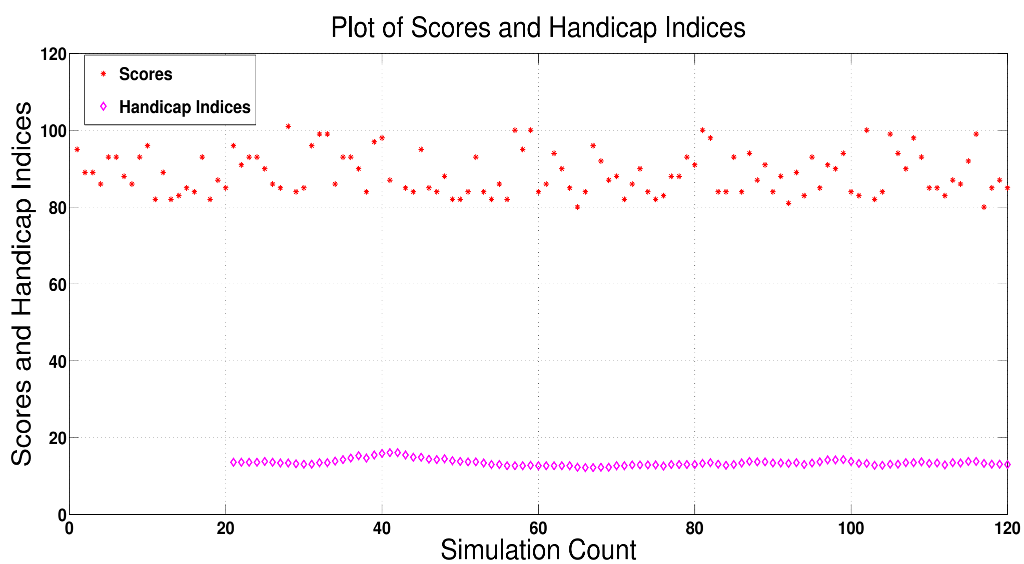


Figure 9.3 Simulation of empirical distribution of scores from handicap 13 and corresponding response of handicap indices is observed.

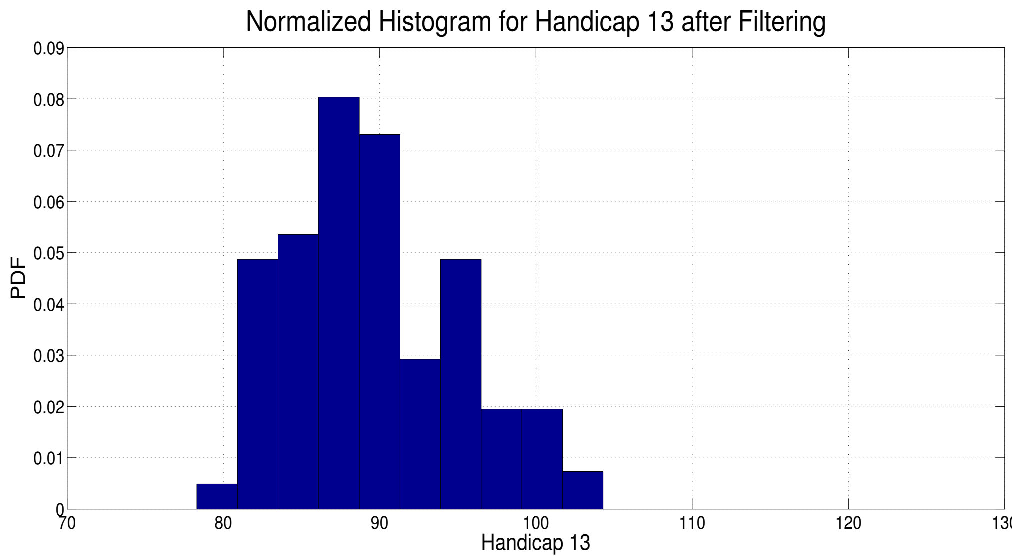


Figure 9.4 Histogram of data set of scores for handicap 13 after the filtering procedure.

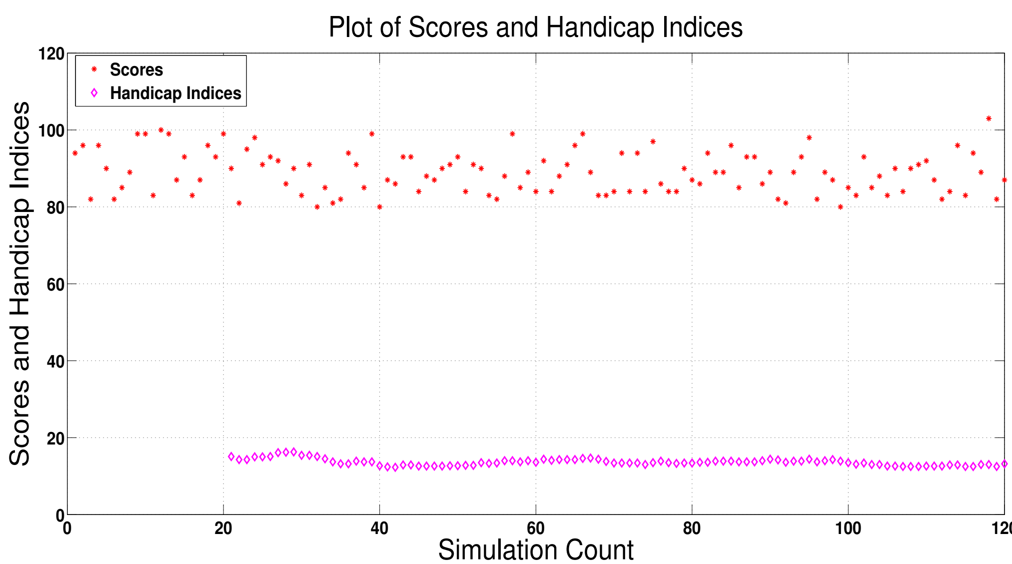


Figure 9.5 Simulation of filtered distribution of scores from handicap 13 corresponding response of handicap indices is observed.

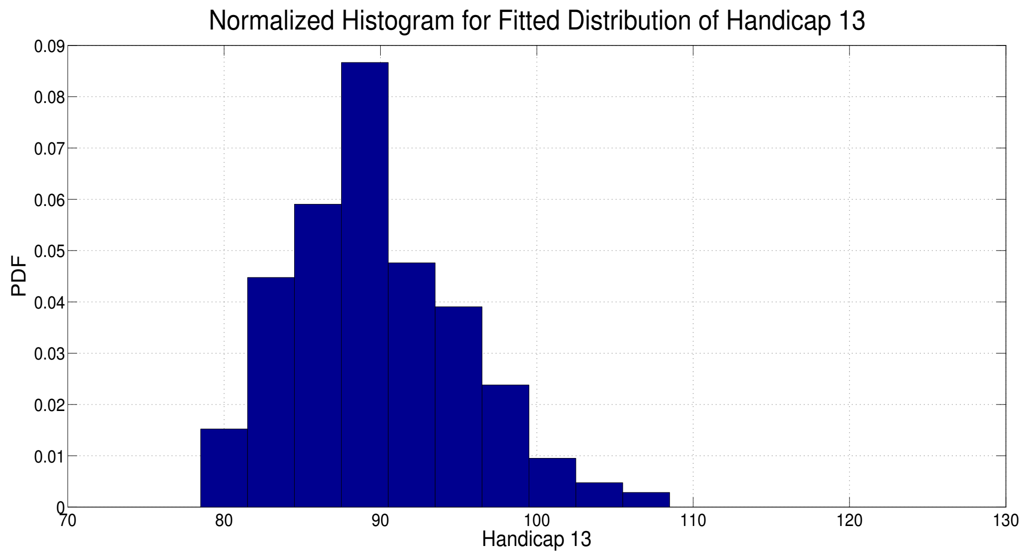


Figure 9.6 Histogram of data set of scores for handicap 13 after fitting generalized extreme value distribution.

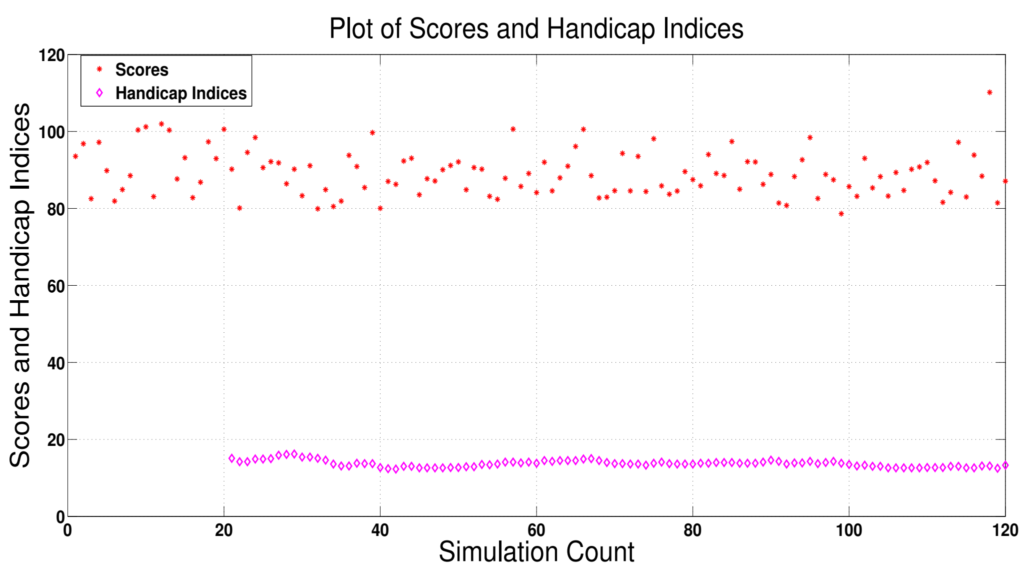


Figure 9.7 Simulation of fitted distribution of scores from handicap 13 corresponding response of handicap indices is observed.

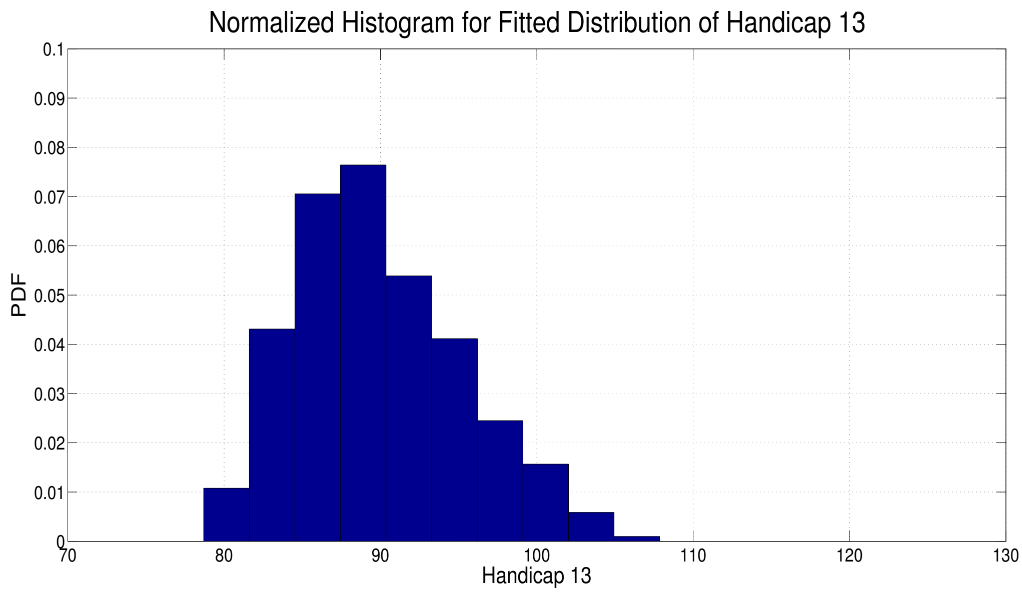


Figure 9.8 Histogram of filtered data set of scores for handicap 13 after fitting generalized extreme value distribution.

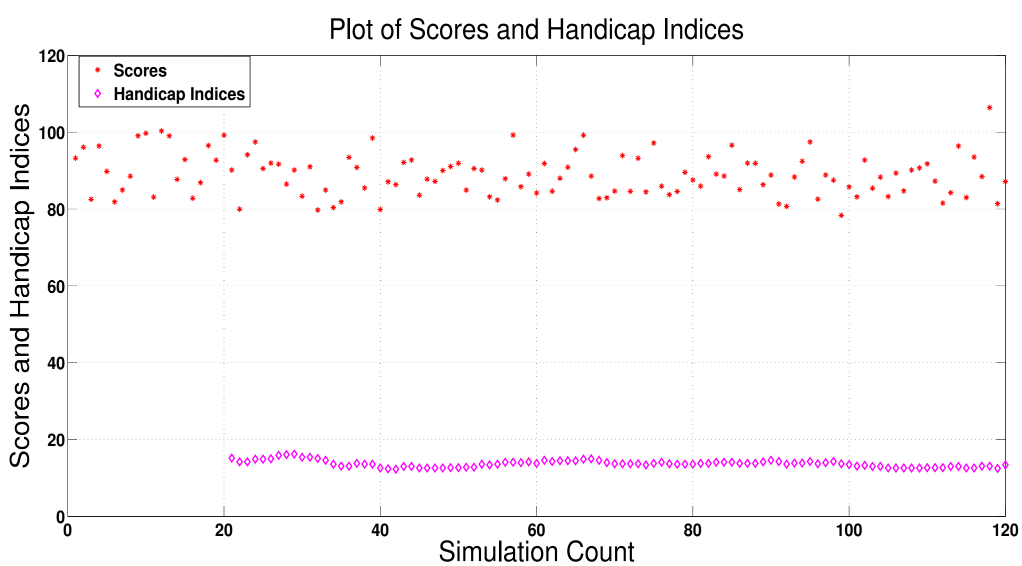


Figure 9.9 Simulation of fitted distribution of filtered data set of scores from handicap 13 and corresponding response of handicap indices is observed.

9.1.2 Effect on Handicap Indices of a Sudden Change in the Distribution of the Scores

Now, we will observe how responsive the handicap system is to a sudden change in the distribution of the scores. For example, arthritis will degrade the performance of the golfer gradually whereas major surgery might lead to a rapid decline in performance. Even though the player was a better golfer before the surgery, it will take time for the handicap to adjust to reflect the golfer's current playing potential. In fact, it will take a minimum of ten rounds of high scores before the corresponding response of the handicap system will occur. The algorithm used in this section is similar to that used in the previous section with the difference being that the parameters of the distribution are changed after certain number of games are played. For our case, we will consider the change of distribution after 50 games are played.

To begin the analysis, let us consider the gross scores which are simulated from a normal distribution with a mean of $\mu = 85$ and standard deviation $\sigma = 2$. For the same value of the standard deviation, the sudden increase in the average of the scores to mean of $\mu = 95$ will cause a delayed increase in the handicap index as seen in the Figure 9.10. The output displays that the sudden change in the handicap indices occurs smoothly because of the relative low variability in the scores.

In order to study the effect of sudden change in distribution for the given data set, the handicaps 13 and 23 are considered. With sudden change in the empirical distribution of the scores from handicap 13 to that from handicap 23, there is a delayed increase in the simulation output of the corresponding handicap indices as shown in the Figure 9.11. Compared to Figure 9.10, the transition to higher handicap index is less pronounced due to the variability existing in the data.

Similar simulations are replicated for the filtered versions of the handicap data sets 13 and 23. Figure 9.12 represents the effect of sudden change of the filtered distribution of scores from handicap 13 to 23. Figures 9.11 and 9.12 exhibit similar

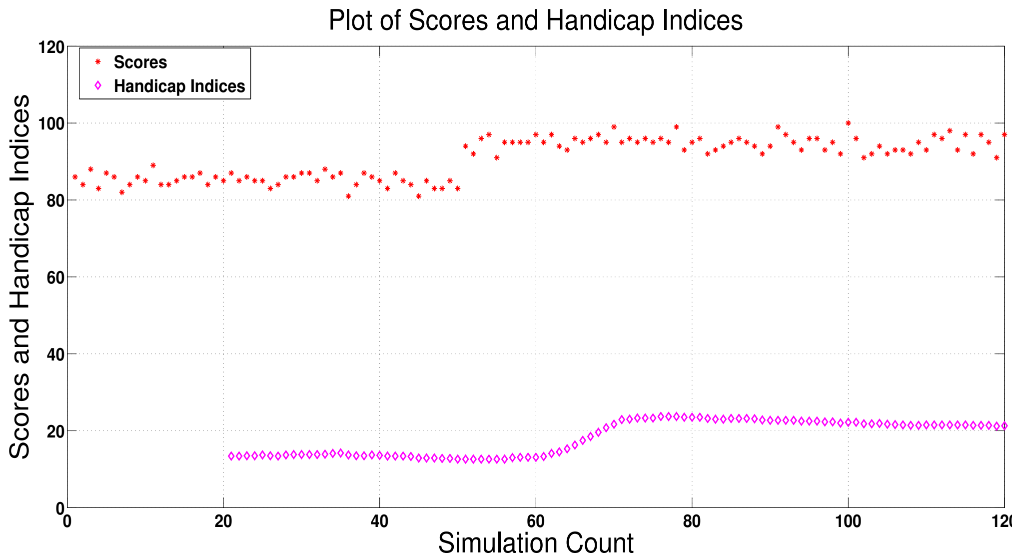


Figure 9.10 Simulation of sudden change in mean of scores from normal distribution and corresponding behavior of handicap indices is observed.

pattern in the output but former plot has lesser variability as compared to the latter plot, because the filtering reduced the spread of scores.

Next, consider the estimated parameters of the target distribution given by GEV distribution for handicap 13 and 23. From Table 8.5, parameters for handicap 13 are shape = -0.042, scale = 4.858 and location = 86.099 and the parameters of the handicap 23 are shape = -0.127, scale = 6.232 and location = 95.987. The resultant effect in the behavior of the handicap indices is displayed in Figure 9.13. The pattern in the output is similar to Figures 9.11 and 9.12.

Similarly, after fitting GEV distribution to the filtered handicap data set, the estimated parameters of the handicap 13 are given by shape = -0.108, scale = 4.844 and location = 86.1699 and for handicap 23 are given by shape = -0.208, scale = 6.084 and location = 95.973. Figure 9.14 shows the effect of sudden change in the distribution of the fitted distributions to the filtered handicap data sets.

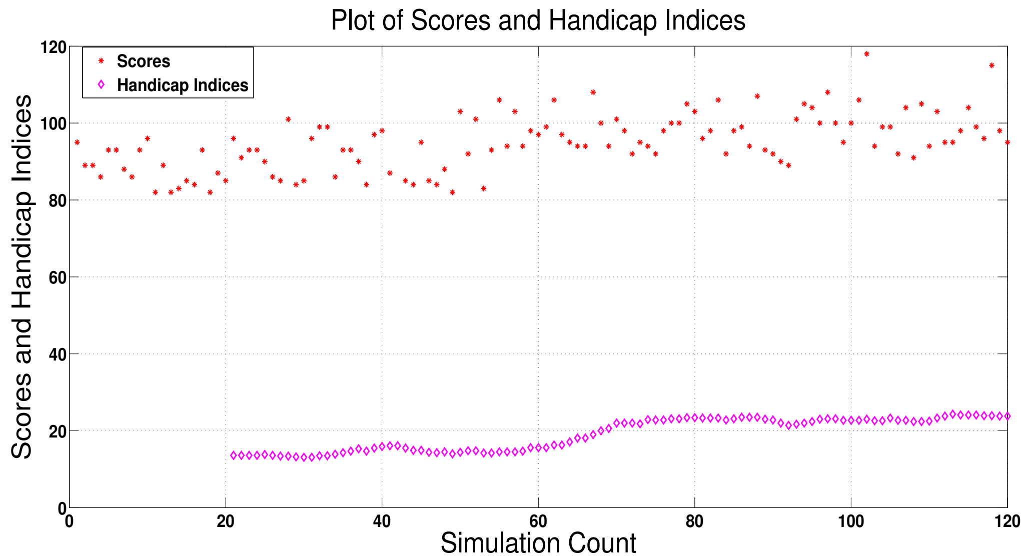


Figure 9.11 Simulation of sudden change in the empirical distribution of scores from handicap 13 to 23 and corresponding behavior of handicap indices is observed.

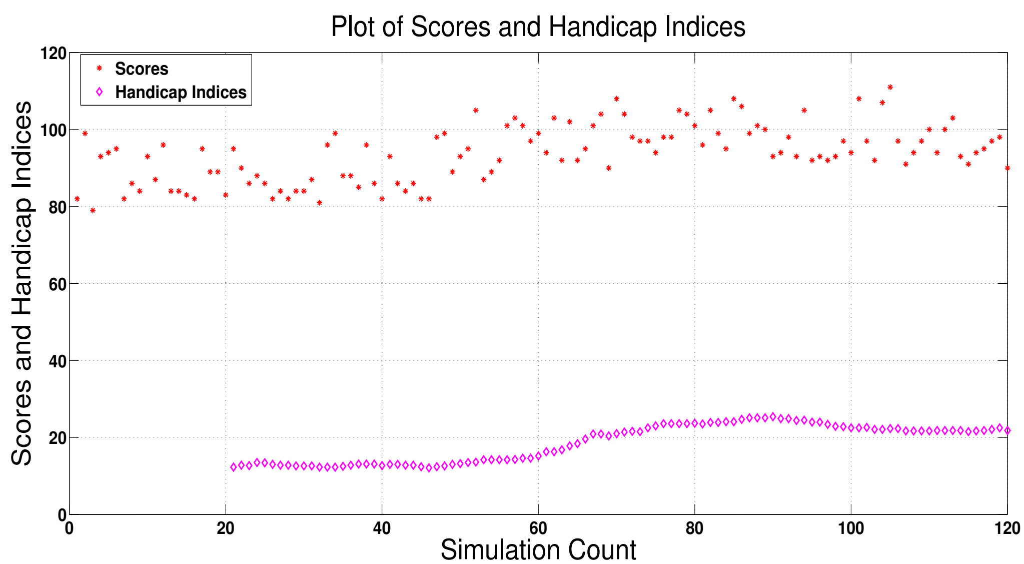


Figure 9.12 Simulation of sudden change in the filtered distribution of scores from handicap 13 to 23 and corresponding behavior of handicap indices is observed.

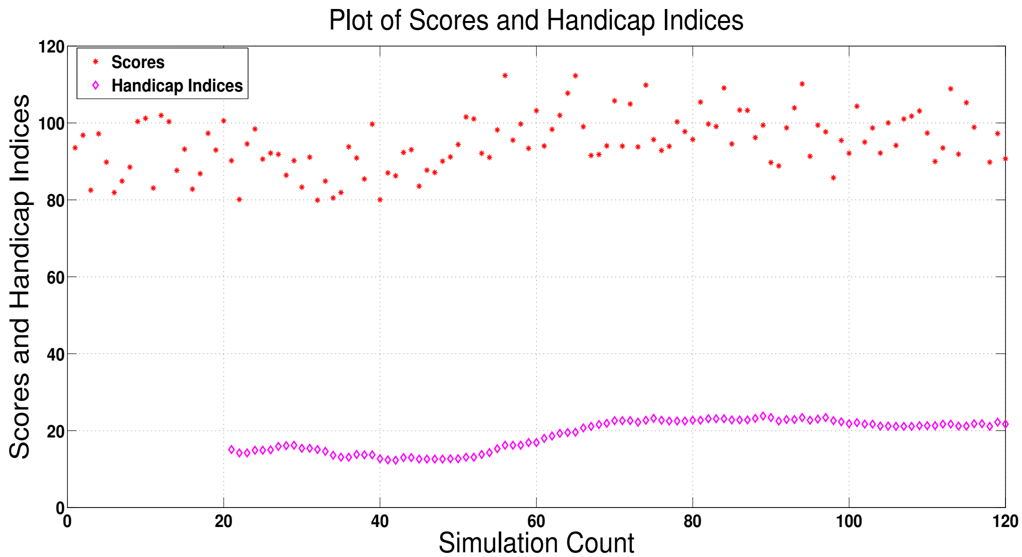


Figure 9.13 Simulation of sudden change in the fitted distribution of scores from handicap 13 to 23 and corresponding behavior of handicap indices is observed.

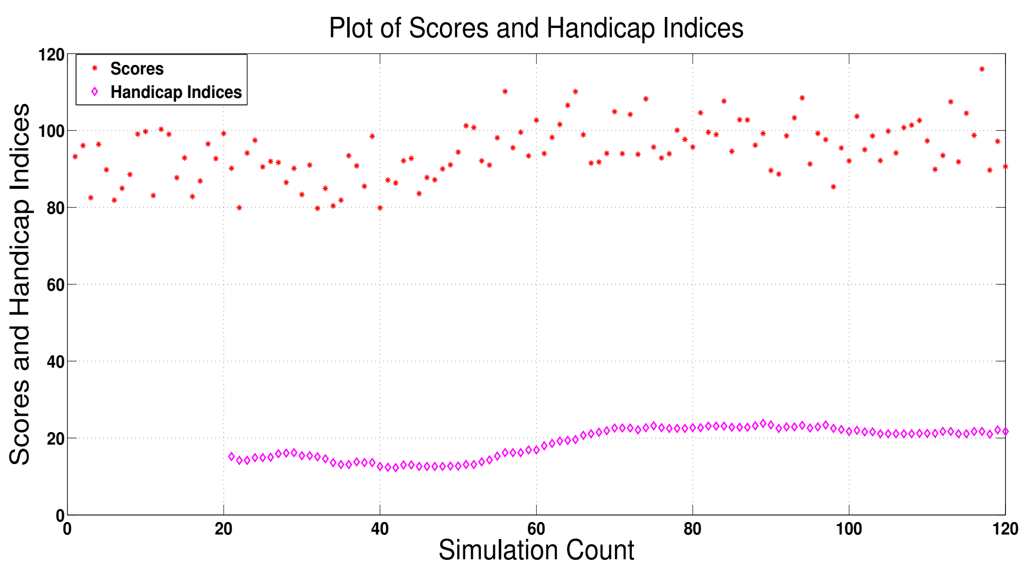


Figure 9.14 Simulation of sudden change in the fitted distribution of filtered data set of scores from handicap 13 to 23 and corresponding behavior of handicap indices is observed.

The qualitative behavior of the results from the simulations is similar for both filtered and fitted data sets but filtering has a greater impact in terms of reducing variance. In each of the simulations, a sudden increase in typical scores results in a delayed increase in the handicap index, was seen; in results not included here, a sudden decrease in typical scores led to an immediate decrease in handicap index as would be expected. We observed almost a similar pattern in the simulations based on data even though the variability in the scores tends to make the transition less obvious. Overall, we can conclude that this study indicates that the handicap system is quite responsive to changes in nature of golf scores.

9.2 Summary of Results from the Simulations

The qualitative behavior of the handicap index from the simulations is similar for both empirical and fitted. Fitting and filtering reduce the variance of the data but filtering has a greater impact. Given how handicaps are computed, a sudden increase in typical scores results in a delayed increase in the handicap index whereas a sudden decrease in typical scores causes an immediate decrease (not shown in the Subsection 9.1.2). From these simulations, we can infer that variance of the golf scores has a great impact on the handicap as high variability in the scores tends to make transition in the handicap less obvious. In order to understand the effectiveness of the handicap system in making matches between players equitable, next chapter includes the simulations of competitions between players having different handicaps.

CHAPTER 10

DISCUSSION AND ANALYSIS OF COMPETITIONS BETWEEN PLAYERS WITH DIFFERING HANDICAPS

In a statistically fair handicapping system, the chance of the better golfer and weaker golfer winning will be same. But, better golfers might suggest that the system should be designed so the stronger golfer still has an advantage over the weaker golfer [61, 62]. In order to investigate the effectiveness of the handicap system, we will conduct simulations of competitions between golfers with differing handicaps and complete competition tables comprising the results of these simulations, based on the empirical and fitted data for golf scores.

10.1 Description of Competitions Using Empirical and Fitted Data

In order to investigate the fairness of the current handicapping system, it is required to set up a simulation procedure to estimate the probability of the winning golfer in a competition between golfers. We will begin our simulation study by completing competition tables using golf scores based upon the original and fitted data sets. A competition table, used to summarize the results of the matches, consists of rows and columns both headed with handicap values 10 through 40. Each entry in the table represents the probability of a golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Each of the probabilities is obtained by averaging over 100,000 simulated golf matches. We will begin our discussion by analyzing the competitions for both the empirical and fitted data.

10.1.1 Competitions Observed Using the Data

In a two player competition, the USGA recommends that the stronger player allow the weaker to subtract from his score a number of strokes equal to the difference in their handicaps. For example for a competition between players of handicap 10 and 15, then the player with handicap 15 will have 5 strokes subtracted from the overall score. We will discuss the step wise algorithm to estimate the chance that the better golfer is the winner.

First, the data set of the gross scores of both the golfers are used as input and adjusted with respect to their handicap value. Second, the adjusted scores of both the golfers are used to conduct a tally of wins, losses and ties in the game; and determine the proportion of wins with ties excluded in the calculation of the proportion. Finally, the error bars of the winning probability are evaluated. The procedure of estimating the probability of winning the game is replicated for all possible pairs of handicaps ranging from 10 to 40. Therefore, a completed competition table will have 31 rows and 31 columns.

The competition results are shown for selected handicaps $12, 15, 18, \cdots, 39$ in form of 10×10 matrix as shown in Tables 10.1 to 10.4 for the original data, the filtered data, the GEV distribution fit to the original data and the GEV distribution fit to the filtered data, respectively. For the sake of simplicity, the full 31×31 matrices are given in the appendix. Ideally, each cell in a competition matrix should have three outputs: the probability of winning, the number of ties, and the error bars. For example, for a competition between a golfer with handicap 12 against that of handicap 15, using the original data, the three entries should be in form of probability of winning = 0.5340, number of ties = 5403 and error bars = [0.5308 0.5372]. The number of ties is small in comparison to 100,000 simulations and also the error bars are on the order of 10^{-3}. These values are completely typical of those observed for all cases; thus for the sake of brevity, they are not included in any of the tables.

Table 10.1 Competition Matrix for Original Data Set of Scores

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.5000	0.5340	0.5137	0.5072	0.5398	0.4809	0.5671	0.5471	0.5777	0.6018
15	0.4623	0.5008	0.4817	0.4738	0.5073	0.4509	0.5365	0.5162	0.547	0.5702
18	0.4849	0.5196	0.4992	0.4931	0.5215	0.4631	0.5512	0.5329	0.5621	0.5865
21	0.4952	0.5265	0.5088	0.5009	0.5305	0.4755	0.5553	0.5385	0.5663	0.5988
24	0.4568	0.4933	0.4767	0.4712	0.5022	0.4433	0.5337	0.5104	0.5473	0.5707
27	0.5194	0.553	0.5347	0.5261	0.5564	0.4984	0.5836	0.5635	0.5902	0.6182
30	0.4330	0.4652	0.4478	0.4447	0.4688	0.4174	0.5010	0.4847	0.5193	0.5527
33	0.4523	0.4856	0.4651	0.4585	0.4894	0.4388	0.5154	0.5016	0.5287	0.5672
36	0.4213	0.4535	0.4323	0.4314	0.4553	0.4077	0.4808	0.4685	0.5005	0.5352
39	0.4004	0.4322	0.4106	0.4061	0.4281	0.3819	0.4481	0.4334	0.4639	0.5006

Table 10.2 Competition Matrix for Filtered Data Set of Scores

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.5012	0.5213	0.5251	0.5244	0.5448	0.4912	0.5834	0.5660	0.5890	0.6146
15	0.4811	0.4992	0.5073	0.5040	0.5319	0.4707	0.5665	0.5457	0.5745	0.6030
18	0.4769	0.4938	0.5008	0.5015	0.5214	0.4692	0.5588	0.5393	0.5679	0.5944
21	0.4792	0.4962	0.4989	0.4994	0.5207	0.4708	0.5528	0.5414	0.5652	0.5972
24	0.4516	0.4669	0.4803	0.4780	0.4975	0.4483	0.5409	0.5199	0.5534	0.5825
27	0.5085	0.5269	0.5299	0.5269	0.5511	0.4998	0.5844	0.5675	0.5923	0.6201
30	0.4184	0.4328	0.4434	0.4485	0.4595	0.4177	0.5016	0.4862	0.5166	0.5542
33	0.4356	0.4550	0.4599	0.4608	0.4770	0.4350	0.5165	0.4974	0.5293	0.5698
36	0.4086	0.4256	0.4299	0.4343	0.4456	0.4097	0.4833	0.4727	0.5010	0.5350
39	0.3873	0.3982	0.4033	0.4053	0.4205	0.3794	0.4442	0.4309	0.4614	0.4978

Table 10.3 Competition Matrix for Fitted Data Set of Scores

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.4989	0.5341	0.5110	0.5076	0.5364	0.4812	0.5649	0.5437	0.5694	0.6065
15	0.4677	0.5006	0.4813	0.4736	0.5047	0.4507	0.5296	0.514	0.5441	0.5778
18	0.4848	0.5185	0.501	0.4907	0.5215	0.4657	0.5491	0.5312	0.5577	0.5942
21	0.4948	0.5264	0.5096	0.5010	0.5112	0.5284	0.4795	0.5384	0.5624	0.6029
24	0.4643	0.4963	0.4788	0.4740	0.4991	0.4467	0.5327	0.5102	0.5430	0.5766
27	0.5206	0.5534	0.5321	0.5234	0.5506	0.5008	0.5766	0.5613	0.5850	0.6180
30	0.4363	0.4717	0.4518	0.4465	0.4711	0.4210	0.5030	0.4842	0.5174	0.5529
33	0.4528	0.4874	0.4679	0.4627	0.4873	0.4396	0.5151	0.4970	0.5329	0.5655
36	0.4283	0.4574	0.4407	0.4318	0.4579	0.4172	0.4832	0.4690	0.5009	0.5324
39	0.3947	0.4248	0.4057	0.396	0.4222	0.3836	0.4479	0.4325	0.4638	0.4993

Table 10.4 Competition Matrix for Fitted Distribution of Filtered Data Set of Scores

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.5001	0.5201	0.5244	0.5201	0.5473	0.4929	0.5789	0.5609	0.5859	0.6243
15	0.4791	0.4976	0.5055	0.5033	0.5277	0.4753	0.5627	0.5460	0.5716	0.6085
18	0.4753	0.4966	0.4982	0.4975	0.5174	0.4697	0.5550	0.5397	0.5647	0.6032
21	0.5065	0.4942	0.4659	0.5317	0.4990	0.5134	0.4743	0.4996	0.5151	0.5176
24	0.4542	0.4706	0.4802	0.4798	0.5002	0.4520	0.5369	0.5204	0.5455	0.5890
27	0.5107	0.5254	0.5289	0.5288	0.5516	0.5001	0.5793	0.5662	0.5858	0.6239
30	0.4206	0.4376	0.4457	0.4487	0.4633	0.4216	0.5011	0.4848	0.5171	0.5543
33	0.4375	0.4543	0.4621	0.4594	0.4812	0.4386	0.5174	0.4998	0.5274	0.5654
36	0.4137	0.4288	0.4329	0.4365	0.4528	0.4102	0.4865	0.4718	0.5012	0.5431
39	0.3781	0.3914	0.3962	0.3992	0.4116	0.3777	0.4445	0.4326	0.4588	0.4995

In Tables 10.1 to 10.4, it is observed that the probabilities along the diagonal are close to 0.5 for golfers having equal handicaps. The probabilities for a competition between players with a pair of distinct handicaps and the reversed order of handicaps do not sum exactly to 1 because these results are from independent simulations. The amount that these probabilities do not sum equal to 1 is usually comparable in size to the error bar. Furthermore, it is observed that in a direct competition, a player with a lower handicap seems to have an advantage over a player with a higher handicap. In addition, as the gap between the handicaps increases, the lower-handicap golfer has an increasing advantage over the weaker golfer. This is caused due to the fact that there is high variance associated with the scores from high handicap golfers.

It is clear from the competition tables that due to high variability in the scores, there is no clear trend in the actual probability values. Results in the handicap simulations are in qualitative agreement for fitted and empirical; filtering reduces the variance in all cases. From the simulation of scores under different scenarios, we can conclude that there is an existence of bias in the handicap system which is most evident in the competition between players of greatly differing handicaps.

As none of the techniques were helpful in providing clear trends in the probability values, so we will now adopt the approach of linear regression as discussed in Chapter 8. Table 8.6 includes the lists of the regression fitted parameters. Based on the discussion in Chapter 8, we know that Table 8.6 lists the values of parameters under three categories original, original/regression and moving average/regression.

Using the information given in Table 8.6, we will now complete the competition tables. Based on the shape parameter, the competitions are completed for categories as original/regression, moving average/regression, and average and regression. Table 10.5 is the competition table using the information under the category original/regression. For the sake of brevity, the complete 31×31 table is shown in appendix. In this table, it is observed that the extreme behavior in probability values has been reduced considerably with very few fluctuations as each of those values are around 0.5.

Next, using the combination of moving average and regression to determine the parameters of the GEV distribution, the competition table is shown in Table 10.10 with complete table in the appendix for the sake of simplicity. The table has fewer fluctuations than Table 10.6 in the probability values, indicates further reduction in the variance of the scores. To clarify the trend in the probability values, it is helpful to construct a color map. In Figure 10.1, the color coding from light pink to dark pink indicates the increase in probability values around 0.45 to 0.49, values around 0.5 are given in white, while light blue to dark blue indicates probability values increasing from 0.5 to around 0.54. The color coding provides a clear indicator of the general bias observed in the handicap system.

Finally, Table 10.7 is completed using the regression fitted values of scale and location and the average value of the shape parameter in the GEV distribution. In comparison with Table 10.6, we observe that the winning probability values follow similar trend while moving across the row for each handicap. We can infer that using the regression procedure and average value of shape parameter, there is stability in values. Similar to Figure 10.1, we also obtain the color map for the given situation. We observe that the variance in Figure 10.1 is further reduced in Figure 10.2. This is indicated by the smoother transition of the colors. Hence, using the average value for shape parameter in combination with regression fitted values for the other two leads to further smoothing in the results.

Table 10.5 Competition Matrix: Using Regression Fitting Procedure and Original Values of Shape Parameter

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.4987	0.5139	0.5084	0.5104	0.5121	0.5204	0.5183	0.5268	0.5330	0.5351
15	0.4876	0.5018	0.4928	0.4969	0.4955	0.5048	0.5044	0.5098	0.5161	0.5226
18	0.4896	0.5064	0.5000	0.5050	0.5016	0.5125	0.5076	0.5079	0.5248	0.5296
21	0.4890	0.5063	0.4975	0.4991	0.5021	0.5097	0.5088	0.5150	0.5240	0.5269
24	0.4879	0.5026	0.4972	0.4985	0.4980	0.5073	0.5049	0.5127	0.5211	0.5282
27	0.4827	0.4926	0.4880	0.4921	0.4956	0.4996	0.4955	0.5050	0.5109	0.5186
30	0.4792	0.4908	0.4908	0.4923	0.4979	0.5036	0.5017	0.5100	0.5153	0.5238
33	0.4759	0.4913	0.4818	0.4836	0.4877	0.4940	0.4914	0.4989	0.5055	0.5168
36	0.4688	0.4845	0.4759	0.4812	0.4800	0.4833	0.4844	0.4900	0.5008	0.5079
39	0.4621	0.4770	0.4700	0.4722	0.4725	0.4815	0.4768	0.4862	0.4935	0.4995

Table 10.6 Competition Matrix: Using Regression Fitting Procedure and Moving Average on the Values of Shape Parameter

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.5008	0.5085	0.5089	0.5113	0.5174	0.5253	0.5251	0.5310	0.5410	0.5443
15	0.4904	0.5004	0.4988	0.4997	0.5105	0.5125	0.5175	0.5226	0.4897	0.5371
18	0.4908	0.4981	0.4983	0.5057	0.5095	0.5157	0.5179	0.5245	0.5295	0.5352
21	0.4845	0.4973	0.4987	0.4996	0.5050	0.5123	0.5169	0.5204	0.5262	0.5312
24	0.4803	0.4884	0.4911	0.4975	0.5001	0.5053	0.5078	0.5138	0.5214	0.5238
27	0.4770	0.4864	0.4868	0.4865	0.4942	0.5024	0.5011	0.5074	0.5155	0.5229
30	0.4743	0.4833	0.4820	0.4858	0.4920	0.4954	0.5007	0.5059	0.5146	0.5182
33	0.4664	0.4764	0.4789	0.4804	0.4846	0.4922	0.4919	0.5009	0.5075	0.5142
36	0.4605	0.4719	0.4710	0.4740	0.4806	0.4883	0.4876	0.4914	0.5022	0.5064
39	0.4592	0.4662	0.4629	0.4663	0.4734	0.4749	0.4774	0.4844	0.4910	0.4994

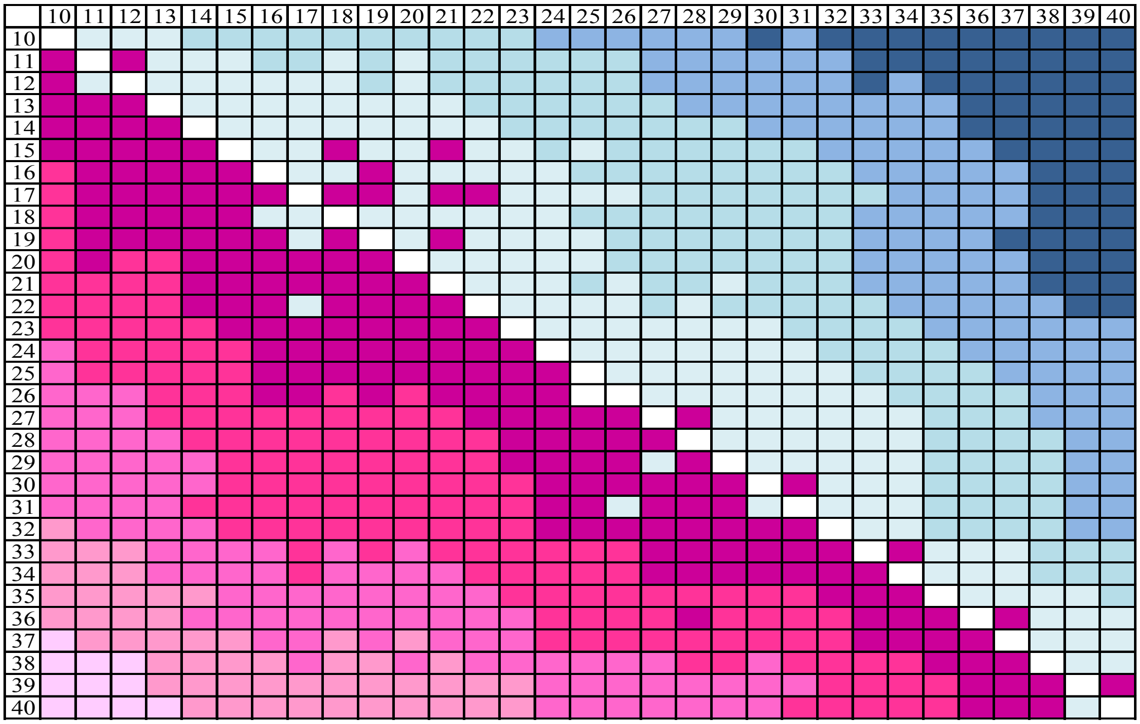


Figure 10.1 Color map of the winning probability values using regression fitted values of scale and location and moving average values for shape parameters.

Table 10.7 Competition Matrix: Using Regression Fitting Procedure and Average of the Values of Shape Parameter

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.5022	0.5066	0.5140	0.5192	0.5226	0.5272	0.5324	0.5362	0.5366	0.5411
15	0.4942	0.5004	0.5058	0.5121	0.5186	0.5209	0.5258	0.5296	0.5324	0.5376
18	0.4884	0.4943	0.5009	0.5051	0.5087	0.5156	0.5191	0.5223	0.5266	0.5325
21	0.4859	0.4864	0.4969	0.4991	0.5026	0.5116	0.5119	0.5183	0.5228	0.5245
24	0.4802	0.4832	0.4920	0.4935	0.5000	0.5050	0.5095	0.5145	0.5156	0.5208
27	0.4756	0.4821	0.4885	0.4937	0.4984	0.5028	0.5061	0.5102	0.5112	0.5169
30	0.4712	0.4742	0.4818	0.4867	0.4906	0.4959	0.4985	0.5059	0.5071	0.5109
33	0.4613	0.4706	0.4765	0.4831	0.4901	0.4891	0.4976	0.5003	0.5060	0.5087
36	0.4597	0.4649	0.4711	0.4794	0.4842	0.4861	0.4935	0.4984	0.5000	0.5040
39	0.4587	0.4636	0.4702	0.4757	0.4782	0.4830	0.4864	0.4900	0.4992	0.4994

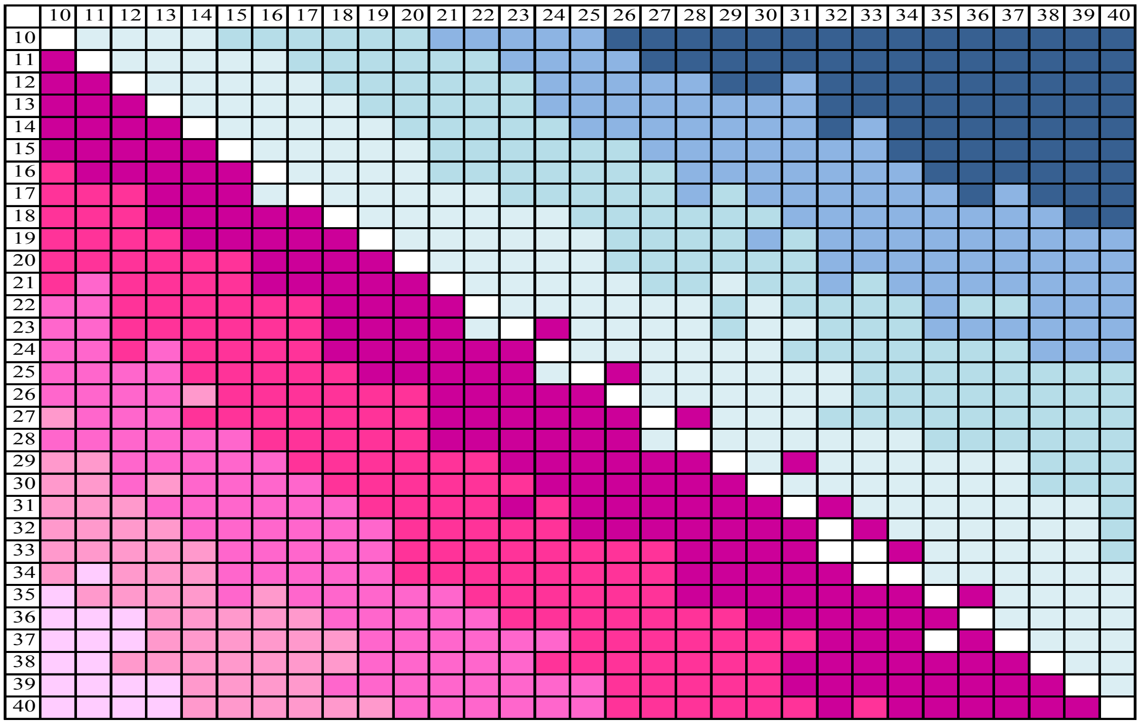


Figure 10.2 Color map of the winning probability values using regression fitted values of scale and location and overall mean value of scale parameter.

10.1.2 Competition Tables for Stratified Data

In the previous section, we observed that though the regression technique produced smoother trends in the probability values, the bias still exits. For the sake of achieving fairness in the competitions, we will restrict competitions from the data set to those between golfers playing on same day and tournament. The size of the tournaments in the data set varies from about 40 to over 100 golfers with handicaps ranging from 10 to 40. In the simulation procedure, the steps followed to complete the competition table are similar to those discussed in the previous section with the difference that, for each competition, the tournament is selected randomly. There is a possibility of having no or very few golfers corresponding to a particular handicap in the randomly selected tournament. In order to better understand the impact of this, we will consider two cases: (a) there is at least one player of each handicap in the tournament, and (b) there are at least two players of each handicap in the tournament.

To begin with, the competition table is obtained for original data set of scores. Tables 10.8 to 10.11 are the competitions for selected pairs of handicaps in form of 10×10 matrix. The complete matrix of competition of size 31×31 is given in the appendix. Table 10.8 and 10.10 show the simulation results when there is at least one player of each handicap in the selected tournament. Tables 10.9 and 10.11 contain the simulation results when there are at least two players of each handicap in the selected tournament. We observe that the winning probabilities become more extreme with improved consistency in comparison to those in Tables 10.1 to 10.4. Hence, though the fluctuations in the trend of probability values have been reduced, but a lower handicap player still seems to have an advantage over a player with a higher handicap. But in comparison to Tables 10.5 to 10.7, the results in Tables 10.8 to 10.11 do not have smoother trends in the probability values.

Table 10.8 Competition Matrix for Original Data Set of Scores with At Least One Player from Each Tournament

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.5004	0.5463	0.5222	0.5180	0.5595	0.4712	0.5805	0.5698	0.5663	0.6137
15	0.4526	0.4967	0.4858	0.4717	0.5084	0.4603	0.5506	0.5194	0.5722	0.5792
18	0.4769	0.5150	0.5039	0.4918	0.5195	0.4813	0.5556	0.5190	0.5718	0.5816
21	0.4824	0.5267	0.5083	0.5034	0.5277	0.4885	0.5566	0.5447	0.5769	0.5935
24	0.4380	0.4898	0.4807	0.4704	0.4997	0.4352	0.5219	0.5192	0.5232	0.5843
27	0.5293	0.5405	0.5179	0.5124	0.5627	0.5026	0.5675	0.5656	0.5865	0.5938
30	0.4212	0.4524	0.4461	0.4450	0.4800	0.4334	0.4977	0.4659	0.5324	0.5504
33	0.4300	0.4847	0.4784	0.4555	0.4830	0.4343	0.5326	0.4969	0.5145	0.5504
36	0.4327	0.4298	0.4282	0.4267	0.4744	0.4157	0.4645	0.4833	0.5002	0.5276
39	0.3876	0.4217	0.4164	0.4107	0.4142	0.4024	0.4515	0.4510	0.4718	0.5025

Table 10.9 Competition Matrix for Original Data Set of Scores with At Least Two Players from Each Tournament

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.5028	0.5600	0.5194	0.5235	0.5618	0.4750	0.5709	0.5689	0.6056	0.6379
15	0.4399	0.5010	0.4757	0.4779	0.5120	0.4948	0.5278	0.5213	0.5791	0.6219
18	0.4783	0.5247	0.5014	0.5082	0.5242	0.4797	0.5565	0.5517	0.6022	0.6226
21	0.4758	0.5203	0.4929	0.5000	0.5162	0.4834	0.5413	0.5404	0.5915	0.6310
24	0.4396	0.4883	0.4738	0.4832	0.4997	0.4396	0.5263	0.5070	0.5541	0.6072
27	0.5274	0.5506	0.5205	0.5189	0.5583	0.4996	0.5683	0.5721	0.6168	0.6280
30	0.4326	0.4722	0.4443	0.4611	0.4701	0.4302	0.5025	0.4704	0.5301	0.5752
33	0.4262	0.4762	0.4481	0.4558	0.4945	0.4294	0.5292	0.4996	0.5343	0.5895
36	0.3918	0.4205	0.3982	0.4113	0.4473	0.3829	0.4710	0.4660	0.4980	0.5334
39	0.3617	0.3815	0.3771	0.3678	0.3928	0.373	0.4249	0.4084	0.4625	0.5000

Table 10.10 Competition Matrix for Filtered Data Set of Scores with At Least One Player from Each Tournament

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.5044	0.5319	0.5369	0.5367	0.5586	0.4940	0.5605	0.5780	0.5970	0.6265
15	0.4711	0.4994	0.5077	0.5132	0.5295	0.4671	0.5371	0.5501	0.5790	0.6090
18	0.4654	0.4911	0.4980	0.5094	0.5155	0.4589	0.5330	0.5443	0.5683	0.6039
21	0.4597	0.4848	0.4913	0.4975	0.5157	0.4572	0.5206	0.5356	0.5620	0.5987
24	0.4434	0.4703	0.4777	0.4868	0.4988	0.4414	0.5122	0.5228	0.5531	0.5891
27	0.5088	0.5378	0.5410	0.5422	0.5594	0.5000	0.5650	0.5817	0.5999	0.6361
30	0.4374	0.4623	0.4678	0.4772	0.4861	0.4313	0.4984	0.5116	0.5380	0.5873
33	0.4250	0.4496	0.4573	0.4631	0.4764	0.4221	0.4913	0.4987	0.5268	0.5785
36	0.4038	0.4241	0.4310	0.4381	0.4423	0.4011	0.4604	0.4693	0.4999	0.5451
39	0.3832	0.4240	0.4035	0.4023	0.4242	0.3766	0.4199	0.4205	0.4571	0.5008

Table 10.11 Competition Matrix for Filtered Data Set of Scores with At Least Two Players from Each Tournament

Handicap	12	15	18	21	24	27	30	33	36	39
12	0.5023	0.5485	0.5244	0.5369	0.5645	0.4879	0.5840	0.5935	0.6260	0.6540
15	0.4509	0.5006	0.4954	0.5051	0.5270	0.4699	0.5577	0.5551	0.6172	0.6466
18	0.4741	0.5032	0.4973	0.5174	0.5237	0.4870	0.5637	0.5603	0.6118	0.6326
21	0.4643	0.4937	0.4840	0.4986	0.5117	0.4793	0.5366	0.5400	0.5874	0.6339
24	0.4350	0.4707	0.4742	0.4891	0.5003	0.4457	0.5342	0.5170	0.5625	0.6187
27	0.5148	0.5291	0.5167	0.5214	0.5537	0.4976	0.5711	0.5782	0.6187	0.6320
30	0.4126	0.4442	0.4352	0.4606	0.4623	0.4292	0.4996	0.4710	0.5252	0.5777
33	0.4085	0.4460	0.4354	0.4579	0.4811	0.4204	0.5261	0.4981	0.5290	0.5930
36	0.3749	0.3815	0.3912	0.4155	0.4425	0.3816	0.4764	0.4699	0.4985	0.5420
39	0.3445	0.3529	0.3664	0.3659	0.3790	0.3642	0.4226	0.4089	0.4551	0.4989

When the data set is stratified tournament wise, there is not enough data for each handicap within each tournament in order to complete the fitting procedure. As a result, we can not complete simulations using data from a fitted distribution.

Each of the scenarios indicate preference to the low handicap golfer in the row versus the high handicap golfer. In view of this, the competition tables discussed above still reflect the existing bias in the handicap system. In other words, these results indicate that the current handicap system is not statistically fair.

10.2 Competition Tables Using Dynamically Generating Handicaps

The purpose of this section to study the impact on the effectiveness of the system while following the performance of a single golfer through time. Under this situation, in a competition between two golfers the gross scores will be adjusted using the dynamically generated handicaps at each stage. For each run of the simulations, using the same procedure of calculating the handicap, the scores will be adjusted based on the handicap obtained in the previous run. It implies in this case that two golfers will be playing sequence of games with handicaps evolving in time. For example, in this case in a competition between golfer of "handicap 10" and "handicap 13", the corresponding scores will be adjusted by not subtracting these numbers but the current handicap. For the sake of simplicity, we have shown the results in 10×10 matrix as shown in Table 10.12. It is observed that with increase in the handicap values across the row, the golfer with lower handicap has an advantage over the higher handicap golfer. There is no smooth trend in the probability values. In order to apply fitting and filtering techniques, we would require more data. Overall, we observe that the bias in the system still exists.

Table 10.12 Competition Matrix: Using Dynamic Generation of the Handicaps

Handicap	12	15	18	20	22	24	27	28	30	36
12	0.4989	0.5153	0.5306	0.5550	0.5643	0.5585	0.6073	0.5999	0.6389	0.6390
15	0.4795	0.4998	0.5113	0.5349	0.5449	0.5372	0.5862	0.5768	0.6171	0.6183
18	0.4660	0.4828	0.4986	0.5274	0.5345	0.5268	0.5815	0.5689	0.6144	0.6165
21	0.4427	0.4603	0.4730	0.4966	0.5060	0.4967	0.5563	0.5439	0.5892	0.5956
24	0.4277	0.4487	0.4619	0.4890	0.4943	0.4861	0.5466	0.5363	0.5833	0.5874
27	0.4402	0.4575	0.4724	0.5004	0.5107	0.4991	0.5593	0.5480	0.5962	0.6001
30	0.3897	0.4121	0.4211	0.4461	0.4526	0.4405	0.4980	0.4907	0.5415	0.5542
33	0.3961	0.4173	0.4254	0.4511	0.4559	0.4476	0.5083	0.4953	0.5472	0.5606
36	0.3617	0.3832	0.3875	0.4088	0.4143	0.4070	0.4602	0.4501	0.4983	0.5180
39	0.3554	0.3750	0.3806	0.3998	0.4050	0.3962	0.4411	0.4349	0.4794	0.4977

10.3 Concluding Remarks

In Tables 10.1 to 10.4, we observe that bias is most evident in the competition between players of greatly differing handicaps. Though filtering has some impact in reducing the extreme behavior of the probability values, but there are fluctuations in the trend of those values. This is an indicator of high variance in the golf scores. In order to reduce this effect, the regression approach is adopted and results are discussed in the Tables 10.5 to 10.7 by considering various categories. It is clearly evident from these tables that the variance in the probability values for higher handicaps is reduced and fluctuations are minimized effectively. As an alternative means of studying this issue, Tables 10.8 to 10.11 are completed under the scenario that the players are restricted to the same tournament. We observe that stratifying the data causes extreme values of probability to become more extreme. Regardless of the approach taken, bias clearly exists in the USGA handicap system; the existence of bias in the handicap system is most evident in the competition between players of greatly differing handicaps.

CHAPTER 11

CONCLUSIONS AND FUTURE WORK ON GOLF HANDICAPS

For over 100 years, the USGA has been attempting to create an unbiased handicapping system. With the help of the AT&T Golf Tournament League data set, we have studied this problem. We conducted data analysis on the data set using sophisticated techniques for fitting procedure and state of the art filtering techniques; we observed that the generalized extreme value distribution is an appropriate fit for the golf scores whereas the normal distribution is not an appropriate fit for the golf scores. Simulation of handicaps demonstrated the effect of variations in golf scores on handicap. For instance, a sudden increase in typical golf scores resulted in a delayed increase in the handicap index. Simulation of direct competitions have demonstrated that a player with a lower handicap seems to have an advantage over a player with a higher handicap indicating a bias in the handicap system. This bias arises in particular due to higher variability in golf scores of players with higher handicap.

As further refinement of this work, we will explore the use of other techniques for data filtering such as the Grubbs test. Likewise, we can use other goodness of fit tests and other parameter estimation techniques. It would be worthwhile to consider variations of the current handicapping system to see the effect on the bias. Possible modifications include changing the number of scores used in calculating the handicap and modifying the procedure for integrating the course difficulty. It would also be interesting to further pursue the idea of a competition between a pair of golfers with dynamically evolving handicaps. A data set with a large temporal record of golf scores for several individuals would be necessary for such a study.

APPENDIX A

COMPETITION TABLES

A.1 Competition Matrix for Raw Data

Tables A.1 through A.3 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 having gross scores from the original data set. Each entry in this table is the value of the winning probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.1 lists competitions between handicaps 10 to 20, Table A.2 lists competitions between handicaps 21 to 30 and Table A.3 lists competitions between handicaps 31 to 40.

Table A.1 Competition Matrix for Raw Data Set

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.5029	0.5699	0.5406	0.5509	0.5441	0.5771	0.5429	0.5175	0.5633	0.5398	0.5771
11	0.4305	0.4979	0.4762	0.4805	0.4746	0.5062	0.4746	0.4498	0.4872	0.4663	0.518
12	0.4591	0.5304	0.5000	0.5083	0.5009	0.5340	0.4992	0.4790	0.5137	0.4963	0.5348
13	0.4493	0.5182	0.4863	0.4971	0.4860	0.5286	0.4868	0.4717	0.5063	0.4862	0.5268
14	0.4549	0.5304	0.4997	0.5110	0.4973	0.5356	0.4965	0.4766	0.5133	0.4959	0.5424
15	0.4202	0.4941	0.4623	0.4707	0.4649	0.5008	0.4643	0.4393	0.4817	0.4602	0.5035
16	0.4564	0.5283	0.5032	0.5136	0.5025	0.5369	0.5011	0.4819	0.5177	0.5007	0.5374
17	0.4818	0.5496	0.5221	0.5280	0.5217	0.5603	0.5200	0.4980	0.5410	0.5188	0.5573
18	0.4391	0.5106	0.4849	0.4964	0.4825	0.5196	0.4802	0.4620	0.4992	0.4801	0.5190
19	0.4610	0.5348	0.5061	0.5160	0.5047	0.5373	0.5019	0.4812	0.5189	0.4983	0.5410
20	0.4217	0.4857	0.4644	0.4760	0.4635	0.4956	0.4626	0.4454	0.4762	0.4595	0.4996
21	0.4553	0.5164	0.4952	0.505	0.4907	0.5265	0.4889	0.4739	0.5088	0.4871	0.5296
22	0.4487	0.5143	0.489	0.5013	0.4844	0.5203	0.4850	0.4737	0.5022	0.4819	0.5182
23	0.4500	0.5042	0.4845	0.4955	0.4815	0.5169	0.4785	0.4655	0.4960	0.4784	0.5091
24	0.4191	0.4833	0.4568	0.4717	0.4631	0.4933	0.4607	0.4370	0.4767	0.4556	0.501
25	0.4467	0.5101	0.4925	0.5036	0.4925	0.5237	0.4943	0.4725	0.5062	0.4885	0.5281
26	0.4142	0.4803	0.4604	0.4686	0.4594	0.4922	0.4613	0.4376	0.4777	0.4570	0.4961
27	0.4808	0.5460	0.5194	0.5302	0.5205	0.5530	0.5184	0.5007	0.5347	0.5153	0.5531
28	0.4607	0.5209	0.4994	0.5097	0.4949	0.5312	0.4936	0.4800	0.5084	0.4932	0.5277
29	0.4834	0.5406	0.5212	0.5302	0.5155	0.5534	0.5164	0.5011	0.5304	0.5172	0.5452
30	0.3959	0.4448	0.433	0.4501	0.4311	0.4652	0.4299	0.4188	0.4478	0.4295	0.4709
31	0.4537	0.5125	0.4915	0.5032	0.4908	0.5258	0.4929	0.4753	0.5051	0.4893	0.5272
32	0.4596	0.5143	0.4932	0.5037	0.4912	0.5297	0.4926	0.4766	0.5079	0.4917	0.5250
33	0.4176	0.4692	0.4523	0.464	0.4480	0.4856	0.4514	0.4377	0.4651	0.4483	0.4835
34	0.4898	0.5439	0.5188	0.5292	0.5181	0.5481	0.5164	0.5039	0.5328	0.5154	0.5427
35	0.4452	0.4975	0.4737	0.4842	0.4700	0.5125	0.4742	0.4628	0.4861	0.4710	0.5036
36	0.3849	0.4391	0.4213	0.4328	0.4206	0.4535	0.4197	0.4074	0.4323	0.4195	0.4481
37	0.3837	0.4378	0.4162	0.4249	0.4115	0.4481	0.4171	0.402	0.4306	0.4149	0.4431
38	0.4416	0.4820	0.4699	0.4846	0.4684	0.4980	0.4695	0.4608	0.4833	0.4686	0.497
39	0.3753	0.4114	0.4004	0.4099	0.3955	0.4322	0.3987	0.3854	0.4106	0.3965	0.4241
40	0.3249	0.3797	0.3596	0.3710	0.3566	0.3884	0.3550	0.3457	0.3706	0.3544	0.3837

Table A.2 Competition Matrix for Raw Data Set (Continued)

Hcp	21	22	23	24	25	26	27	28	29	30
10	0.5452	0.5534	0.5486	0.5822	0.5472	0.584	0.5184	0.5423	0.518	0.6023
11	0.4853	0.4861	0.4964	0.5175	0.4827	0.5214	0.4555	0.4808	0.4574	0.5551
12	0.5072	0.5105	0.518	0.5398	0.5086	0.5416	0.4809	0.5006	0.4784	0.5671
13	0.4957	0.5029	0.5027	0.5262	0.497	0.5289	0.4708	0.4913	0.4696	0.5525
14	0.5062	0.5131	0.5185	0.5388	0.5083	0.5399	0.4808	0.5019	0.4843	0.5693
15	0.4738	0.4779	0.4817	0.5073	0.4728	0.5098	0.4509	0.4702	0.4473	0.5365
16	0.5092	0.5171	0.5205	0.5406	0.5098	0.5411	0.4819	0.5068	0.4874	0.5674
17	0.5253	0.5315	0.5358	0.5592	0.5271	0.5591	0.5024	0.5211	0.4996	0.5792
18	0.4931	0.4992	0.5045	0.5215	0.4937	0.5241	0.4631	0.4887	0.4671	0.5512
19	0.5121	0.5149	0.5186	0.5435	0.5109	0.5436	0.4865	0.5058	0.4828	0.5676
20	0.4717	0.4803	0.4852	0.5018	0.4725	0.5052	0.4478	0.4736	0.4545	0.5348
21	0.5009	0.5036	0.5135	0.5305	0.5001	0.5297	0.4755	0.4949	0.4806	0.5553
22	0.494	0.5018	0.506	0.5215	0.4916	0.5211	0.4689	0.488	0.4707	0.5458
23	0.4886	0.4919	0.5012	0.5128	0.4844	0.5172	0.4642	0.4856	0.4723	0.541
24	0.4712	0.4797	0.4871	0.5022	0.4706	0.4991	0.4433	0.4703	0.4501	0.5337
25	0.499	0.5057	0.5165	0.5328	0.4992	0.5324	0.4722	0.498	0.4815	0.5615
26	0.47	0.4759	0.4832	0.5046	0.4683	0.5004	0.4443	0.4681	0.4522	0.5312
27	0.5261	0.5293	0.5324	0.5564	0.5241	0.5581	0.4984	0.5218	0.4992	0.5836
28	0.5046	0.5094	0.5179	0.53	0.5039	0.5306	0.4789	0.4989	0.4822	0.5562
29	0.5226	0.5298	0.5298	0.5497	0.522	0.547	0.4996	0.5164	0.5002	0.5681
30	0.4447	0.4485	0.4581	0.4688	0.4393	0.4685	0.4174	0.4457	0.4287	0.501
31	0.4985	0.5062	0.5098	0.5299	0.5007	0.5326	0.4748	0.4939	0.4762	0.5539
32	0.5001	0.5054	0.5113	0.5286	0.4989	0.5267	0.4774	0.4967	0.4807	0.5508
33	0.4585	0.4679	0.4759	0.4894	0.4576	0.4871	0.4388	0.4591	0.4428	0.5154
34	0.5237	0.5279	0.5294	0.548	0.5197	0.5487	0.4989	0.5144	0.5023	0.5746
35	0.4782	0.4782	0.4897	0.5024	0.478	0.5078	0.4579	0.4729	0.4601	0.526
36	0.4314	0.4348	0.446	0.4553	0.4294	0.4537	0.4077	0.4337	0.4122	0.4808
37	0.4198	0.425	0.4351	0.4474	0.4257	0.4536	0.4003	0.4234	0.4056	0.4748
38	0.4775	0.4865	0.4942	0.4997	0.4738	0.5003	0.4554	0.4777	0.4629	0.5255
39	0.4061	0.4087	0.4187	0.4281	0.4008	0.4317	0.3819	0.4057	0.3935	0.4481
40	0.3669	0.3681	0.3781	0.3844	0.3640	0.3877	0.3444	0.3638	0.3475	0.4145

Table A.3 Competition Matrix for Raw Data Set (Continued)

Hcp	31	32	33	34	35	36	37	38	39	40
10	0.5459	0.5389	0.5835	0.5123	0.5523	0.6144	0.6155	0.5629	0.6267	0.6749
11	0.4856	0.4879	0.5333	0.458	0.5043	0.5644	0.5669	0.5182	0.5884	0.6205
12	0.5118	0.5024	0.5471	0.4803	0.5234	0.5777	0.5844	0.5283	0.6018	0.6402
13	0.4975	0.4935	0.533	0.4667	0.5149	0.5666	0.5719	0.5186	0.5906	0.6312
14	0.5097	0.5079	0.5504	0.4830	0.5300	0.5807	0.5866	0.5332	0.6025	0.6465
15	0.4755	0.4747	0.5162	0.4529	0.4868	0.5470	0.5515	0.5003	0.5702	0.6115
16	0.5099	0.5076	0.5485	0.4853	0.5267	0.5774	0.5847	0.5306	0.6010	0.6452
17	0.5278	0.5198	0.5628	0.4992	0.5371	0.5916	0.5995	0.5399	0.6166	0.6566
18	0.4954	0.4917	0.5329	0.4700	0.5105	0.5621	0.5700	0.5169	0.5865	0.6300
19	0.5111	0.5111	0.5510	0.4831	0.5266	0.5811	0.5863	0.5322	0.6016	0.6415
20	0.4755	0.4777	0.5151	0.4553	0.4999	0.5523	0.5568	0.5005	0.5797	0.6143
21	0.5043	0.4993	0.5385	0.4767	0.5215	0.5663	0.5777	0.5231	0.5988	0.6360
22	0.4946	0.4930	0.5316	0.4717	0.5162	0.5623	0.5721	0.5174	0.5913	0.6295
23	0.4898	0.4909	0.5264	0.4651	0.5112	0.5535	0.5618	0.5053	0.5847	0.6227
24	0.4731	0.4729	0.5104	0.4519	0.4952	0.5473	0.5521	0.4989	0.5707	0.6158
25	0.5016	0.5031	0.5399	0.4778	0.5220	0.5737	0.5787	0.5276	0.5994	0.6365
26	0.4701	0.4721	0.5119	0.4504	0.4886	0.5448	0.5485	0.4984	0.5700	0.6125
27	0.5229	0.5232	0.5635	0.5015	0.5421	0.5902	0.5982	0.5477	0.6182	0.6516
28	0.5061	0.5034	0.5440	0.4835	0.5275	0.5693	0.58	0.5238	0.5959	0.6370
29	0.5248	0.5168	0.5583	0.5005	0.5431	0.5849	0.5949	0.5344	0.6102	0.6534
30	0.4486	0.4503	0.4847	0.4272	0.4739	0.5193	0.5267	0.4779	0.5527	0.5864
31	0.5012	0.4987	0.5364	0.4734	0.5200	0.5660	0.5728	0.5256	0.5928	0.6245
32	0.5023	0.4990	0.5338	0.4787	0.5167	0.5654	0.5745	0.5234	0.5957	0.6278
33	0.4630	0.4657	0.5016	0.4436	0.4880	0.5287	0.5437	0.4891	0.5672	0.5988
34	0.5251	0.5206	0.5603	0.5009	0.5427	0.5799	0.5924	0.5408	0.6100	0.6455
35	0.4799	0.4823	0.5120	0.4566	0.4998	0.5419	0.5552	0.4956	0.5785	0.6098
36	0.4345	0.4349	0.4685	0.4164	0.4560	0.5005	0.5086	0.4578	0.5352	0.5678
37	0.4276	0.4279	0.4570	0.4073	0.4429	0.4878	0.5006	0.4488	0.5193	0.5552
38	0.4767	0.4771	0.5124	0.4608	0.5004	0.5416	0.5523	0.4977	0.5791	0.6079
39	0.4069	0.4046	0.4334	0.3915	0.4216	0.4639	0.4810	0.4186	0.5006	0.5372
40	0.3744	0.3703	0.4046	0.3538	0.3937	0.4307	0.4482	0.3921	0.4651	0.4982

A.2 Competition Matrix for Filtered Data

Tables A.4 through A.6 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 having gross scores from the filtered version of the data set. Each entry in this table is the value of probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.4 lists competitions between handicaps 10 to 20, Table A.5 lists competitions between handicaps 21 to 30 and Table A.6 lists competitions between handicaps 31 to 40.

Table A.4 Competition Matrix for Filtered Data Set

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.5014	0.5677	0.5395	0.5588	0.5535	0.5628	0.5388	0.5158	0.5643	0.5372	0.5897
11	0.4323	0.5004	0.4691	0.4907	0.4813	0.4849	0.4697	0.4462	0.4924	0.4631	0.531
12	0.4602	0.5336	0.5012	0.5208	0.5157	0.5213	0.5024	0.4811	0.5251	0.4958	0.5536
13	0.4415	0.5099	0.4776	0.4996	0.4873	0.4992	0.4751	0.4574	0.5002	0.4717	0.5319
14	0.4489	0.5165	0.4851	0.5107	0.5007	0.5038	0.4854	0.4656	0.5107	0.4798	0.5406
15	0.4376	0.5135	0.4811	0.5007	0.4945	0.4992	0.4811	0.4579	0.5073	0.4742	0.539
16	0.4604	0.5301	0.4998	0.5252	0.5147	0.5194	0.5005	0.4806	0.5249	0.4986	0.5558
17	0.4842	0.5557	0.5186	0.5417	0.5311	0.5413	0.5176	0.5001	0.5443	0.5171	0.5752
18	0.4367	0.5064	0.4769	0.4963	0.4911	0.4938	0.4725	0.4554	0.5008	0.4720	0.5307
19	0.4653	0.5377	0.5030	0.5251	0.5184	0.5269	0.5052	0.4866	0.5296	0.4982	0.5610
20	0.4085	0.4691	0.4429	0.4689	0.4583	0.4611	0.4465	0.4286	0.4690	0.4403	0.4989
21	0.4459	0.5070	0.4792	0.5015	0.4923	0.4962	0.4783	0.4650	0.4989	0.4720	0.5314
22	0.4460	0.5121	0.4835	0.5071	0.4955	0.5025	0.4811	0.4695	0.5041	0.4765	0.5342
23	0.4516	0.5048	0.4795	0.5038	0.4905	0.4993	0.4769	0.4669	0.4996	0.4746	0.5305
24	0.4136	0.4781	0.4516	0.4805	0.4670	0.4669	0.4541	0.4387	0.4803	0.4453	0.5113
25	0.4482	0.5107	0.4878	0.5129	0.4986	0.5040	0.4900	0.4691	0.5140	0.4798	0.5457
26	0.4221	0.4841	0.4592	0.9100	0.4753	0.4776	0.4625	0.4413	0.4843	0.4547	0.5221
27	0.4758	0.5409	0.5085	0.5312	0.5219	0.5269	0.5086	0.4894	0.5299	0.5052	0.5606
28	0.4579	0.5181	0.4936	0.5161	0.5032	0.5075	0.4875	0.4733	0.5122	0.4883	0.5364
29	0.4761	0.5341	0.5066	0.5275	0.5218	0.5258	0.5032	0.4911	0.5308	0.5035	0.5535
30	0.3876	0.4354	0.4184	0.4457	0.4323	0.4328	0.4216	0.4088	0.4434	0.4136	0.4732
31	0.4590	0.5187	0.4956	0.5130	0.5059	0.5102	0.4959	0.4795	0.5154	0.4896	0.5467
32	0.4509	0.4982	0.4795	0.4987	0.4905	0.4982	0.4812	0.4691	0.5072	0.4778	0.5277
33	0.4053	0.4592	0.4356	0.4604	0.4486	0.4550	0.4378	0.4211	0.4599	0.4313	0.4864
34	0.4784	0.5329	0.5051	0.5299	0.5182	0.5192	0.5040	0.4919	0.5234	0.5003	0.5482
35	0.4431	0.4904	0.4651	0.4875	0.4721	0.4881	0.4672	0.4562	0.4865	0.4645	0.5149
36	0.3771	0.4304	0.4086	0.4330	0.4225	0.4256	0.4117	0.3976	0.4299	0.4054	0.4539
37	0.3830	0.4316	0.4115	0.4317	0.4218	0.4299	0.4137	0.3971	0.4335	0.4062	0.4623
38	0.4348	0.4765	0.4602	0.4813	0.4741	0.4733	0.4635	0.4552	0.4801	0.4544	0.5056
39	0.3642	0.3962	0.3873	0.4061	0.3960	0.3982	0.3853	0.3745	0.4033	0.3799	0.4244
40	0.3283	0.3820	0.3579	0.3804	0.3683	0.3732	0.3543	0.3458	0.3763	0.3515	0.4047

Table A.5 Competition Matrix for Filtered Data Set (Continued)

Handicap	21	22	23	24	25	26	27	28	29	30
10	0.5549	0.5522	0.5498	0.5862	0.5496	0.5784	0.5256	0.5441	0.5254	0.6133
11	0.4951	0.4906	0.4954	0.5189	0.4864	0.5133	0.4611	0.4828	0.4643	0.5653
12	0.5244	0.5181	0.5193	0.5448	0.5151	0.5409	0.4912	0.5100	0.4919	0.5834
13	0.4985	0.4939	0.4931	0.5197	0.4918	0.5174	0.4672	0.4857	0.4720	0.5546
14	0.5069	0.5047	0.5090	0.5315	0.4990	0.5263	0.4771	0.4958	0.4819	0.5662
15	0.504	0.4991	0.5040	0.5319	0.4963	0.5227	0.47070	0.4978	0.4719	0.5665
16	0.5196	0.5192	0.5207	0.5458	0.5106	0.5382	0.4935	0.5124	0.4950	0.5776
17	0.5368	0.5310	0.5350	0.5633	0.5284	0.5604	0.5129	0.5260	0.5084	0.5912
18	0.5015	0.4955	0.4967	0.5214	0.4878	0.5164	0.4692	0.4901	0.4688	0.5588
19	0.5268	0.5226	0.5256	0.5509	0.5192	0.5458	0.4955	0.5157	0.4961	0.5859
20	0.4690	0.4664	0.4708	0.4912	0.4564	0.4799	0.4413	0.4616	0.4484	0.5285
21	0.4994	0.4992	0.5024	0.5207	0.4882	0.5142	0.4708	0.4895	0.4777	0.5528
22	0.5055	0.4998	0.5042	0.5207	0.4901	0.5164	0.4755	0.4908	0.4773	0.5566
23	0.4990	0.4958	0.5005	0.5164	0.4869	0.5120	0.4724	0.4893	0.4798	0.5536
24	0.4780	0.4784	0.4847	0.4975	0.4670	0.4878	0.4483	0.4742	0.4556	0.5409
25	0.5092	0.5082	0.5132	0.5358	0.5012	0.5282	0.4790	0.5004	0.4925	0.5698
26	0.4860	0.4841	0.4896	0.5079	0.4701	0.5007	0.4520	0.4787	0.4643	0.5482
27	0.5269	0.5230	0.5298	0.5511	0.5188	0.5435	0.4998	0.5150	0.5033	0.5844
28	0.5121	0.5064	0.5103	0.5272	0.4987	0.5193	0.4837	0.5016	0.4862	0.5613
29	0.5219	0.5246	0.5237	0.5429	0.5140	0.5343	0.4988	0.5129	0.5016	0.5713
30	0.4485	0.4426	0.4502	0.4595	0.4275	0.4520	0.4177	0.4364	0.4245	0.5016
31	0.5183	0.5097	0.5155	0.5374	0.5041	0.5339	0.4888	0.5034	0.4923	0.5674
32	0.5026	0.4942	0.4992	0.5220	0.4914	0.5148	0.4726	0.4893	0.4787	0.5518
33	0.4608	0.4548	0.4636	0.4770	0.4476	0.4694	0.435	0.4492	0.4404	0.5165
34	0.5224	0.5202	0.5263	0.5400	0.5128	0.5328	0.4984	0.5070	0.5000	0.5738
35	0.4837	0.4762	0.4819	0.5029	0.4728	0.4968	0.4615	0.4716	0.4646	0.5302
36	0.4343	0.4323	0.4363	0.4456	0.4208	0.4378	0.4097	0.4267	0.4167	0.4833
37	0.4335	0.4287	0.4349	0.4504	0.4210	0.4449	0.4069	0.4254	0.4144	0.4850
38	0.4829	0.4774	0.4855	0.4968	0.4706	0.4888	0.4580	0.4782	0.4643	0.5271
39	0.4053	0.4012	0.4029	0.4205	0.3887	0.4118	0.3794	0.3998	0.3881	0.4442
40	0.3795	0.3743	0.3796	0.3925	0.3671	0.3846	0.3581	0.3681	0.3589	0.4288

Table A.6 Competition Matrix for Filtered Data Set (Continued)

Handicap	31	32	33	34	35	36	37	38	39	40
10	0.5403	0.5483	0.5929	0.5224	0.5558	0.6206	0.6149	0.5650	0.6357	0.6711
11	0.4816	0.5006	0.5416	0.4689	0.5107	0.5711	0.5669	0.5242	0.6018	0.6142
12	0.5080	0.5191	0.5667	0.4922	0.5353	0.5890	0.5874	0.5874	0.6146	0.6460
13	0.4874	0.4975	0.5383	0.4735	0.5132	0.5688	0.5658	0.5196	0.5951	0.6227
14	0.4953	0.5088	0.5531	0.4814	0.5258	0.5765	0.5792	0.5285	0.6063	0.6328
15	0.4883	0.5005	0.5457	0.4793	0.5094	0.5745	0.5726	0.5271	0.603	0.6307
16	0.5074	0.5187	0.5616	0.4920	0.5323	0.5871	0.5873	0.5384	0.6122	0.6443
17	0.5195	0.5322	0.5763	0.5124	0.5422	0.6025	0.6017	0.5452	0.6258	0.6534
18	0.4849	0.4982	0.5393	0.4768	0.5149	0.5679	0.568	0.5193	0.5944	0.6205
19	0.5118	0.5210	0.5678	0.4980	0.5377	0.5908	0.5906	0.5442	0.6155	0.6483
20	0.4499	0.4719	0.5107	0.4544	0.4842	0.5447	0.5384	0.4925	0.5761	0.5985
21	0.4858	0.4988	0.5414	0.4754	0.5125	0.5652	0.5699	0.5187	0.5972	0.6179
22	0.4919	0.5034	0.5427	0.4818	0.5195	0.5713	0.5719	0.5206	0.6029	0.6294
23	0.4830	0.5002	0.5335	0.4763	0.5177	0.5632	0.5649	0.5142	0.5964	0.6180
24	0.4635	0.4790	0.5199	0.4574	0.4965	0.5534	0.5480	0.5043	0.5825	0.6097
25	0.4914	0.5107	0.5536	0.4864	0.5269	0.5801	0.576	0.5310	0.6114	0.6330
26	0.4673	0.4861	0.5322	0.4671	0.5014	0.5609	0.5531	0.5098	0.5900	0.6148
27	0.5111	0.5271	0.5675	0.5040	0.5399	0.5923	0.5906	0.5431	0.6201	0.6433
28	0.4969	0.5087	0.5524	0.4900	0.5294	0.5711	0.5755	0.5240	0.6043	0.6293
29	0.5089	0.5200	0.5611	0.4990	0.5358	0.5824	0.5851	0.5364	0.6132	0.6438
30	0.4311	0.4538	0.4862	0.4287	0.4694	0.5166	0.5162	0.4725	0.5542	0.5691
31	0.5014	0.5145	0.5548	0.4864	0.5294	0.5819	0.5828	0.5346	0.6141	0.633
32	0.4847	0.4998	0.5344	0.4775	0.5135	0.5631	0.5653	0.5185	0.5904	0.6116
33	0.4465	0.4662	0.4974	0.4422	0.4854	0.5293	0.5321	0.4873	0.5698	0.5833
34	0.5088	0.5223	0.5563	0.5001	0.5355	0.5785	0.5826	0.5363	0.6082	0.6328
35	0.4691	0.4872	0.5215	0.4641	0.4996	0.5470	0.5507	0.5005	0.5871	0.6001
36	0.4176	0.4393	0.4727	0.4228	0.4539	0.5010	0.5024	0.4586	0.5350	0.5558
37	0.4187	0.4343	0.4682	0.4163	0.4476	0.5009	0.5026	0.4543	0.5339	0.5527
38	0.4647	0.4799	0.5155	0.4634	0.4983	0.5410	0.5432	0.5011	0.5856	0.5982
39	0.3874	0.4073	0.4309	0.3922	0.4157	0.4614	0.4666	0.4160	0.4978	0.5181
40	0.3679	0.3842	0.4131	0.3687	0.4000	0.4443	0.4478	0.4037	0.4813	0.4981

A.3 Competition Matrix for Fitted Data Set

Tables A.7 through A.9 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 having gross scores from the fitted version of the data set. Each entry in this table is the value of probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.7 lists competitions between handicaps 10 to 20, Table A.8 lists competitions between handicaps 21 to 30 and Table A.9 lists competitions between handicaps 31 to 40.

Table A.7 Competition Matrix for Fitted Data Set

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.5019	0.5624	0.5398	0.5478	0.5415	0.5703	0.5363	0.516	0.5543	0.5359	0.5686
11	0.4354	0.5008	0.4776	0.4859	0.4759	0.5116	0.4734	0.4546	0.4920	0.4709	0.5100
12	0.4619	0.5240	0.4989	0.5109	0.4960	0.5341	0.4996	0.4806	0.5110	0.4940	0.5340
13	0.4526	0.5151	0.4900	0.4986	0.4904	0.5237	0.4881	0.4711	0.5052	0.4838	0.5210
14	0.4595	0.5269	0.5026	0.5089	0.4992	0.5338	0.4986	0.4789	0.5155	0.4943	0.5322
15	0.4313	0.4916	0.4677	0.4783	0.4656	0.5006	0.4661	0.4474	0.4813	0.4653	0.4998
16	0.4645	0.5252	0.5023	0.5134	0.5013	0.5343	0.5002	0.4823	0.5167	0.4996	0.5333
17	0.4824	0.5440	0.5235	0.5318	0.5206	0.5492	0.5193	0.4958	0.5360	0.5151	0.5500
18	0.4465	0.5094	0.4848	0.4953	0.4839	0.5185	0.4829	0.4668	0.5010	0.4829	0.5186
19	0.4642	0.5270	0.5069	0.5154	0.5024	0.5322	0.5038	0.4860	0.5203	0.4958	0.5361
20	0.4301	0.4877	0.4666	0.4794	0.4668	0.4996	0.4691	0.4510	0.485	0.4645	0.5001
21	0.4592	0.5181	0.4948	0.5050	0.4930	0.5264	0.4926	0.4776	0.5096	0.4926	0.5229
22	0.4548	0.5107	0.4877	0.4979	0.4872	0.5166	0.4881	0.5006	0.5030	0.4857	0.5173
23	0.4498	0.5016	0.4848	0.4946	0.4842	0.5132	0.4824	0.4669	0.4953	0.4797	0.5101
24	0.4240	0.4838	0.4643	0.4748	0.4596	0.4963	0.4644	0.4454	0.4788	0.4599	0.4928
25	0.4555	0.4942	0.4933	0.5043	0.4924	0.5273	0.4924	0.4749	0.5082	0.4897	0.5247
26	0.4305	0.4844	0.4682	0.4753	0.4620	0.4985	0.4628	0.4480	0.4805	0.4594	0.4945
27	0.4829	0.5402	0.5206	0.5249	0.5206	0.5534	0.5189	0.4973	0.5321	0.5169	0.5500
28	0.4683	0.5197	0.4991	0.5089	0.4991	0.5312	0.4989	0.4839	0.5102	0.4978	0.5240
29	0.4858	0.5363	0.5182	0.5282	0.5188	0.5464	0.5175	0.5007	0.5290	0.5147	0.5448
30	0.4031	0.4559	0.4363	0.4486	0.4337	0.4717	0.4362	0.4220	0.4518	0.4342	0.4681
31	0.4617	0.5146	0.4945	0.5050	0.4936	0.5263	0.4961	0.4797	0.5093	0.4946	0.5222
32	0.4654	0.5167	0.4993	0.5074	0.4951	0.5273	0.4922	0.4817	0.5096	0.4922	0.5188
33	0.4154	0.4749	0.4528	0.4649	0.4519	0.4874	0.4535	0.438	0.4679	0.4506	0.4836
34	0.4913	0.5384	0.5180	0.5301	0.5150	0.5485	0.5179	0.5294	0.5314	0.5154	0.5419
35	0.4430	0.4956	0.4780	0.4875	0.4748	0.5056	0.4722	0.4573	0.4901	0.4768	0.5016
36	0.3958	0.4466	0.4283	0.4377	0.4255	0.4574	0.4261	0.4161	0.4407	0.4282	0.4496
37	0.3897	0.4373	0.4222	0.4340	0.4196	0.4514	0.4200	0.4082	0.4344	0.4220	0.4498
38	0.4415	0.4926	0.4748	0.4844	0.4743	0.5034	0.4707	0.4585	0.4884	0.4716	0.4974
39	0.3618	0.4094	0.3947	0.4015	0.3891	0.4248	0.3927	0.3798	0.4057	0.3882	0.4154
40	0.3262	0.3730	0.3573	0.3690	0.3571	0.3881	0.3589	0.3451	0.3706	0.3573	0.3854

Table A.8 Competition Matrix for Fitted Data Set (Continued)

Handicap	21	22	23	24	25	26	27	28	29	30
10	0.5422	0.5470	0.5503	0.5745	0.5452	0.5697	0.5130	0.5326	0.5161	0.5970
11	0.4846	0.4918	0.4958	0.5168	0.4844	0.5136	0.4590	0.4797	0.4610	0.5444
12	0.5076	0.5111	0.5139	0.5364	0.5050	0.5332	0.4812	0.5001	0.4805	0.5649
13	0.4947	0.5004	0.5068	0.5262	0.4987	0.5291	0.4739	0.4877	0.4718	0.5535
14	0.5066	0.5138	0.5194	0.5376	0.5347	0.4799	0.5669	0.5024	0.4840	0.5672
15	0.4736	0.4780	0.4841	0.5047	0.4765	0.5048	0.4507	0.4709	0.4515	0.5296
16	0.5083	0.5154	0.5168	0.5385	0.5090	0.5364	0.4849	0.5035	0.4834	0.5629
17	0.5245	0.5272	0.5337	0.5573	0.5252	0.5529	0.5020	0.5151	0.4999	0.5772
18	0.4907	0.4974	0.5035	0.5215	0.4943	0.5214	0.4657	0.4889	0.4692	0.5491
19	0.5088	0.5152	0.5202	0.5368	0.5099	0.5402	0.4851	0.5040	0.4885	0.5076
20	0.4769	0.4831	0.4885	0.5070	0.4768	0.5077	0.4539	0.4748	0.4558	0.5346
21	0.5010	0.5051	0.5130	0.5112	0.5277	0.4997	0.5284	0.4762	0.4978	0.4795
22	0.4913	0.4975	0.5039	0.5215	0.4926	0.5214	0.4692	0.4885	0.4731	0.5483
23	0.4880	0.4948	0.5028	0.5159	0.4894	0.5132	0.4652	0.4860	0.4691	0.5410
24	0.4740	0.4783	0.4815	0.4991	0.4686	0.5006	0.4467	0.4699	0.4550	0.5327
25	0.5010	0.5067	0.5095	0.5281	0.5019	0.5274	0.4779	0.4977	0.4794	0.5570
26	0.4742	0.4745	0.4851	0.4996	0.4708	0.5005	0.4487	0.4693	0.4505	0.5250
27	0.5234	0.5284	0.5344	0.5506	0.5232	0.5520	0.5008	0.5171	0.5029	0.5766
28	0.5006	0.5120	0.5155	0.5292	0.5032	0.5303	0.4832	0.5001	0.4857	0.5529
29	0.5209	0.5296	0.5305	0.5498	0.5207	0.5486	0.4977	0.5153	0.5005	0.5708
30	0.4465	0.4517	0.4602	0.4711	0.4454	0.4758	0.4210	0.4490	0.4289	0.5030
31	0.4990	0.5070	0.5103	0.5255	0.50110	0.5225	0.4789	0.4986	0.4827	0.5673
32	0.5023	0.5077	0.5125	0.5261	0.4995	0.5296	0.4801	0.4996	0.4827	0.5473
33	0.4627	0.4696	0.4912	0.4873	0.4611	0.4897	0.4396	0.4624	0.4422	0.5151
34	0.5252	0.5277	0.5344	0.5479	0.5208	0.5485	0.4991	0.5165	0.5021	0.5714
35	0.4800	0.4846	0.4935	0.5077	0.4836	0.5073	0.4577	0.4762	0.4601	0.5298
36	0.4318	0.4429	0.4485	0.4579	0.4312	0.4601	0.4172	0.4367	0.4815	0.4832
37	0.4288	0.4336	0.4379	0.4518	0.4293	0.4546	0.4112	0.4294	0.4088	0.4775
38	0.4781	0.4851	0.4900	0.5027	0.4764	0.5051	0.4587	0.4772	0.4609	0.5294
39	0.3960	0.4041	0.4114	0.4222	0.3965	0.4260	0.3836	0.4043	0.3864	0.4479
40	0.3665	0.3708	0.3771	0.3911	0.3646	0.3923	0.3455	0.3689	0.3523	0.4168

Table A.9 Competition Matrix for Fitted Data Set (Continued)

Handicap	31	32	33	34	35	36	37	38	39	40
10	0.5369	0.5371	0.5816	0.5129	0.5559	0.6030	0.6054	0.5572	0.6395	0.6727
11	0.4815	0.4864	0.5249	0.4627	0.5034	0.5561	0.5618	0.5087	0.5918	0.6228
12	0.5049	0.5065	0.5437	0.4803	0.5236	0.5694	0.5767	0.5252	0.6065	0.6384
13	0.4915	0.4935	0.5355	0.4724	0.5171	0.5605	0.5685	0.5147	0.5987	0.6320
14	0.5037	0.5076	0.5492	0.4834	0.5293	0.5731	0.5837	0.5302	0.6142	0.6476
15	0.4731	0.4728	0.5140	0.4527	0.4938	0.5441	0.5499	0.4955	0.5778	0.612
16	0.5057	0.5061	0.5476	0.4826	0.5267	0.5723	0.5800	0.5251	0.6090	0.6402
17	0.5204	0.5179	0.5614	0.4983	0.5410	0.5875	0.5884	0.5408	0.6223	0.6529
18	0.4919	0.4899	0.5312	0.4689	0.5133	0.5577	0.5667	0.5120	0.5942	0.6287
19	0.5076	0.5062	0.5482	0.4842	0.5283	0.5753	0.5798	0.5311	0.6083	0.6447
20	0.4779	0.4811	0.5178	0.4567	0.5009	0.5484	0.5555	0.5032	0.5851	0.6159
21	0.5556	0.4986	0.5384	0.4783	0.5187	0.5624	0.5729	0.5181	0.6029	0.6335
22	0.4922	0.4910	0.5304	0.4715	0.5139	0.5596	0.5647	0.5157	0.5966	0.6281
23	0.4900	0.4896	0.5258	0.4690	0.5066	0.5516	0.5582	0.5083	0.5868	0.6217
24	0.4755	0.4744	0.5102	0.4511	0.4911	0.5430	0.5462	0.4974	0.5766	0.6109
25	0.5002	0.5006	0.5391	0.4807	0.5194	0.5657	0.5748	0.5224	0.6022	0.634
26	0.4723	0.4715	0.5080	0.4533	0.4926	0.5379	0.5489	0.4943	0.5721	0.6087
27	0.5238	0.5184	0.5613	0.4992	0.5390	0.5850	0.5907	0.5403	0.618	0.6536
28	0.5042	0.503	0.5367	0.4818	0.5235	0.5643	0.5729	0.5224	0.5959	0.6294
29	0.522	0.5183	0.5556	0.4994	0.5393	0.5827	0.5875	0.5379	0.6152	0.6458
30	0.4463	0.4524	0.4842	0.4314	0.4700	0.5174	0.5236	0.4752	0.5529	0.5851
31	0.4983	0.4992	0.5381	0.4791	0.5186	0.5621	0.5675	0.5184	0.5968	0.6305
32	0.5027	0.4990	0.5322	0.4850	0.5250	0.5646	0.5698	0.5209	0.5948	0.6269
33	0.4644	0.4622	0.4970	0.4438	0.4840	0.5329	0.5410	0.4864	0.5655	0.5993
34	0.5229	0.5206	0.5556	0.4994	0.5407	0.5796	0.5865	0.5388	0.6110	0.6436
35	0.4814	0.4798	0.5133	0.4589	0.5026	0.5418	0.5515	0.5013	0.5771	0.6086
36	0.4384	0.4380	0.4690	0.4206	0.4572	0.5009	0.5074	0.4603	0.5324	0.5649
37	0.4292	0.4304	0.4619	0.4142	0.4470	0.4934	0.4983	0.4506	0.5240	0.5543
38	0.4825	0.4784	0.5161	0.4633	0.5017	0.5408	0.5498	0.5021	0.5755	0.6075
39	0.4009	0.4014	0.4325	0.3881	0.4209	0.4638	0.4765	0.4284	0.4993	0.5314
40	0.3720	0.3719	0.4011	0.3542	0.3886	0.4348	0.4446	0.3900	0.4691	0.4982

A.4 Competition Matrix for Fitted Filtered Data Set

Tables A.10 through A.12 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 having gross scores from the filtered version of the data set. Each entry in this table is the value of probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.10 lists competitions between handicaps 10 to 20, Table A.11 lists competitions between handicaps 21 to 30 and Table A.12 lists competitions between handicaps 31 to 40.

Table A.10 Competition Matrix for Fitted Filtered Data Set

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.5019	0.5663	0.536	0.5551	0.5507	0.5576	0.534	0.5145	0.5573	0.532	0.5876
11	0.4315	0.5005	0.4698	0.4917	0.4874	0.489	0.4725	0.4516	0.496	0.4647	0.5297
12	0.466	0.5295	0.5001	0.5211	0.5154	0.5201	0.5009	0.481	0.5244	0.4973	0.5525
13	0.4431	0.5075	0.4772	0.4996	0.4952	0.4992	0.477	0.4573	0.504	0.4755	0.5284
14	0.4483	0.5166	0.4861	0.5062	0.5017	0.5053	0.489	0.4663	0.5097	0.4816	0.5383
15	0.4441	0.5111	0.4791	0.5017	0.4954	0.4976	0.481	0.4594	0.5055	0.4779	0.5381
16	0.4633	0.5275	0.501	0.5225	0.5144	0.5182	0.498	0.4799	0.5223	0.495	0.5528
17	0.4833	0.5497	0.5211	0.5375	0.5318	0.5406	0.5167	0.5027	0.5391	0.5174	0.5693
18	0.4414	0.5043	0.4753	0.4981	0.4923	0.4966	0.478	0.4579	0.4982	0.4765	0.5278
19	0.4712	0.534	0.5026	0.5254	0.521	0.5246	0.502	0.4869	0.5272	0.5006	0.5594
20	0.4141	0.4724	0.445	0.4673	0.4585	0.4604	0.4476	0.431	0.4705	0.4416	0.5022
21	0.4493	0.5065	0.5065	0.4814	0.4998	0.4942	0.4962	0.4784	0.4659	0.4974	0.4772
22	0.447	0.5076	0.4835	0.5019	0.4939	0.5005	0.4829	0.4665	0.5041	0.4772	0.5355
23	0.449	0.5028	0.4792	0.5007	0.492	0.4974	0.4817	0.4667	0.4991	0.4747	0.5288
24	0.419	0.4795	0.4542	0.4717	0.47	0.4706	0.4589	0.4385	0.4802	0.4502	0.5104
25	0.4553	0.5139	0.4871	0.5099	0.5019	0.5052	0.4901	0.4717	0.5126	0.4839	0.5395
26	0.4249	0.4899	0.4598	0.4849	0.476	0.4727	0.4619	0.4455	0.4867	0.4579	0.5186
27	0.4787	0.533	0.5107	0.5292	0.5232	0.5254	0.508	0.4909	0.5289	0.504	0.5567
28	0.4626	0.5109	0.4923	0.5108	0.5028	0.5091	0.4911	0.4778	0.5111	0.4861	0.536
29	0.4769	0.5282	0.5061	0.528	0.5192	0.5225	0.5068	0.4922	0.5287	0.5025	0.5527
30	0.3917	0.4419	0.4206	0.4447	0.4365	0.4376	0.4257	0.4109	0.4457	0.419	0.4738
31	0.4667	0.5215	0.4952	0.515	0.5078	0.5118	0.4954	0.4771	0.5182	0.4898	0.546
32	0.4561	0.505	0.4827	0.5013	0.4918	0.4989	0.484	0.4704	0.5027	0.4789	0.526
33	0.4067	0.4592	0.4375	0.4585	0.4505	0.4543	0.4407	0.4269	0.4621	0.4339	0.4891
34	0.4814	0.5293	0.5079	0.5235	0.5171	0.5204	0.5062	0.4937	0.5208	0.5033	0.5474
35	0.4385	0.4861	0.4667	0.489	0.4773	0.4824	0.4625	0.4507	0.4866	0.4613	0.5129
36	0.3868	0.4326	0.4137	0.4362	0.4242	0.4288	0.4167	0.4041	0.4329	0.4093	0.4618
37	0.3828	0.4331	0.4147	0.4354	0.4241	0.4269	0.4133	0.4035	0.4295	0.41	0.46
38	0.4365	0.4828	0.4619	0.4832	0.4721	0.476	0.464	0.4516	0.4849	0.4592	0.5076
39	0.3525	0.3952	0.3781	0.397	0.3852	0.3914	0.3799	0.3657	0.3962	0.3756	0.4245
40	0.3295	0.3739	0.3544	0.3749	0.3653	0.3674	0.3579	0.3477	0.3744	0.3542	0.4039

Table A.11 Competition Matrix for Fitted Filtered Data Set (Continued)

Handicap	21	22	23	24	25	26	27	28	29	30
10	0.5523	0.5519	0.5507	0.58	0.5461	0.5727	0.5246	0.539	0.5214	0.6071
11	0.4952	0.4912	0.4986	0.5216	0.4862	0.5142	0.4662	0.4845	0.4705	0.5573
12	0.5201	0.5181	0.5212	0.5473	0.5133	0.5367	0.4929	0.5097	0.495	0.5789
13	0.4998	0.4962	0.5004	0.5241	0.4889	0.5164	0.4726	0.489	0.4742	0.5581
14	0.5071	0.5046	0.5075	0.5302	0.5002	0.525	0.4782	0.498	0.4803	0.5671
15	0.5033	0.5006	0.5022	0.5277	0.4954	0.5213	0.4753	0.4931	0.479	0.5627
16	0.5218	0.5155	0.5236	0.5434	0.5107	0.5379	0.4889	0.5104	0.4946	0.5743
17	0.5347	0.5343	0.5374	0.5614	0.5311	0.5539	0.5059	0.527	0.5094	0.5855
18	0.4975	0.4954	0.4967	0.5174	0.4904	0.5112	0.4697	0.4886	0.4735	0.555
19	0.524	0.5218	0.5248	0.5476	0.5159	0.5443	0.497	0.5146	0.4956	0.5816
20	0.4699	0.4667	0.4713	0.4898	0.4597	0.4851	0.4395	0.4514	0.4514	0.5263
21	0.5317	0.5019	0.4927	0.499	0.522	0.4926	0.5134	0.4713	0.4896	0.4743
22	0.5039	0.5008	0.504	0.5225	0.4948	0.5164	0.4727	0.4972	0.4837	0.5593
23	0.4993	0.4969	0.4993	0.5188	0.4889	0.5093	0.4741	0.4901	0.4797	0.5532
24	0.4798	0.475	0.4795	0.5002	0.4673	0.4916	0.452	0.4695	0.4552	0.5369
25	0.5102	0.5075	0.5105	0.5347	0.5001	0.5229	0.4834	0.5004	0.4878	0.5672
26	0.4846	0.4806	0.4863	0.5065	0.4737	0.5014	0.4559	0.4789	0.463	0.5421
27	0.5288	0.5247	0.527	0.5516	0.5183	0.5395	0.5001	0.5161	0.5027	0.5793
28	0.5102	0.5047	0.5094	0.5052	0.4997	0.5218	0.4837	0.4975	0.4883	0.5574
29	0.5232	0.5167	0.5232	0.5436	0.5124	0.535	0.4982	0.5119	0.5005	0.5728
30	0.4487	0.4437	0.4491	0.4633	0.4637	0.4576	0.4216	0.4398	0.4305	0.5011
31	0.5165	0.5124	0.5157	0.537	0.5038	0.5292	0.4886	0.5062	0.4931	0.5696
32	0.5039	0.5023	0.5016	0.5166	0.4907	0.5106	0.4783	0.4927	0.4804	0.5461
33	0.4594	0.4621	0.4638	0.4812	0.4505	0.47	0.4386	0.4543	0.4407	0.5174
34	0.5243	0.5205	0.5224	0.5389	0.5122	0.5334	0.498	0.5109	0.5018	0.5652
35	0.485	0.4826	0.4864	0.5025	0.472	0.4989	0.4585	0.4759	0.4621	0.536
36	0.4365	0.4318	0.4376	0.4528	0.4237	0.4429	0.4102	0.4291	0.4214	0.4865
37	0.4364	0.4286	0.4376	0.4526	0.4225	0.4475	0.4113	0.4278	0.4156	0.485
38	0.4846	0.4803	0.4838	0.4982	0.475	0.492	0.46	0.4756	0.4643	0.5313
39	0.3992	0.3958	0.3991	0.4116	0.3869	0.4075	0.3777	0.3942	0.3821	0.4445
40	0.3799	0.3726	0.3815	0.3991	0.3668	0.3866	0.3584	0.3747	0.3632	0.4317

Table A.12 Competition Matrix for Fitted Filtered Data Set (Continued)

Handicap	31	32	33	34	35	36	37	38	39	40
10	0.5372	0.5454	0.5446	0.52	0.5638	0.6138	0.6161	0.5627	0.6479	0.6703
11	0.4799	0.4971	0.5359	0.4736	0.5129	0.5685	0.5634	0.5155	0.608	0.6256
12	0.5041	0.5179	0.5609	0.4958	0.5792	0.5859	0.5846	0.5358	0.6243	0.6452
13	0.4832	0.4975	0.5395	0.4736	0.5132	0.5649	0.5672	0.5175	0.6054	0.6212
14	0.4936	0.5057	0.5489	0.4799	0.5238	0.5743	0.5728	0.5283	0.617	0.6433
15	0.4867	0.502	0.546	0.4771	0.5171	0.5716	0.5705	0.521	0.6085	0.6288
16	0.5055	0.5159	0.5614	0.4979	0.5323	0.5816	0.5854	0.5372	0.6217	0.6404
17	0.5855	0.5306	0.5738	0.5082	0.548	0.5935	0.5983	0.5497	0.6319	0.6569
18	0.4851	0.4976	0.5397	0.4736	0.5144	0.5647	0.5646	0.5174	0.6032	0.6223
19	0.5094	0.521	0.5647	0.4991	0.538	0.5899	0.5914	0.5384	0.6267	0.6459
20	0.4566	0.4718	0.5115	0.4537	0.4865	0.5403	0.5374	0.4934	0.5804	0.5956
21	0.5515	0.4867	0.4996	0.5365	0.4793	0.5151	0.564	0.5629	0.5176	0.5998
22	0.4873	0.5014	0.5419	0.4784	0.5184	0.5664	0.5702	0.5223	0.6068	0.6251
23	0.4835	0.4989	0.5388	0.4771	0.5122	0.5621	0.5633	0.5196	0.6009	0.6275
24	0.4646	0.4813	0.5204	0.4592	0.4937	0.5455	0.5489	0.5026	0.589	0.6038
25	0.4929	0.4929	0.5503	0.4882	0.5275	0.5742	0.578	0.5272	0.6123	0.6338
26	0.4703	0.4901	0.5292	0.4661	0.5022	0.5574	0.5551	0.5047	0.5967	0.6129
27	0.5076	0.5249	0.5662	0.5039	0.537	0.5858	0.5928	0.5437	0.6239	0.643
28	0.4968	0.5082	0.548	0.4879	0.5232	0.5688	0.5713	0.5243	0.608	0.6248
29	0.5064	0.5215	0.5596	0.4988	0.5367	0.5775	0.5857	0.5367	0.6161	0.6395
30	0.4325	0.4513	0.4848	0.4344	0.4619	0.5171	0.516	0.4696	0.5543	0.5709
31	0.5029	0.5137	0.5536	0.4926	0.53	0.5779	0.5797	0.5313	0.6156	0.6347
32	0.4856	0.5041	0.5349	0.4806	0.5154	0.5607	0.5655	0.5174	0.5978	0.616
33	0.4452	0.4684	0.4998	0.4421	0.4743	0.5274	0.5269	0.4825	0.5654	0.5832
34	0.5106	0.5187	0.5547	0.4995	0.5355	0.5749	0.5776	0.5307	0.6129	0.6311
35	0.4709	0.487	0.524	0.4633	0.5001	0.5455	0.5487	0.5044	0.5855	0.6018
36	0.4199	0.441	0.4718	0.4242	0.4534	0.5012	0.5008	0.4582	0.5431	0.555
37	0.4194	0.4395	0.4699	0.4172	0.4508	0.4508	0.5011	0.4552	0.5413	0.5553
38	0.4702	0.4803	0.5175	0.4644	0.4969	0.5398	0.5443	0.5016	0.5812	0.5961
39	0.3851	0.4035	0.4326	0.3895	0.4124	0.4588	0.4599	0.4197	0.4995	0.5158
40	0.3666	0.3852	0.4136	0.371	0.3976	0.4435	0.4466	0.4028	0.4865	0.4989

A.5 Competition Matrix Using Regression Fitting Procedure and Original Values of Shape Parameter

Tables A.13 through A.15 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 using regression fitting procedure on scale and location parameters and original values of shape parameter. Each entry in this table is the value of the winning probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.13 lists competitions between handicaps 10 to 20, Table A.14 lists competitions between handicaps 21 to 30 and Table A.15 lists competitions between handicaps 31 to 40.

Table A.13 Competition Matrix: Using Regression Fitting Procedure and Original Values of Shape Parameter

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.4994	0.5040	0.5054	0.5052	0.5087	0.5108	0.5146	0.5118	0.5178	0.5168	0.5183
11	0.4968	0.5010	0.4981	0.5057	0.5077	0.5100	0.5066	0.5107	0.5146	0.5181	0.5175
12	0.4945	0.4957	0.5020	0.5027	0.5050	0.5062	0.5064	0.5118	0.5111	0.5109	0.5185
13	0.4956	0.4986	0.4969	0.4992	0.5018	0.5047	0.5064	0.5071	0.5088	0.5112	0.5113
14	0.4932	0.4934	0.4964	0.4983	0.5006	0.5019	0.5055	0.5069	0.5087	0.5115	0.5101
15	0.4817	0.4801	0.4876	0.4930	0.4797	0.4975	0.4910	0.4970	0.4942	0.4966	0.4841
16	0.4903	0.4910	0.4965	0.5044	0.4888	0.5105	0.4995	0.5119	0.5048	0.5089	0.4952
17	0.4834	0.4804	0.4871	0.4961	0.4794	0.5020	0.4904	0.5015	0.4961	0.4986	0.4859
18	0.4888	0.4843	0.4915	0.5006	0.4829	0.5059	0.4957	0.5044	0.5022	0.5035	0.4905
19	0.4875	0.4822	0.4875	0.5001	0.4991	0.5059	0.4920	0.5018	0.4972	0.5004	0.4861
20	0.4948	0.4946	0.5009	0.5094	0.4927	0.5166	0.5043	0.5166	0.5101	0.5145	0.4998
21	0.4876	0.4804	0.4894	0.4989	0.4806	0.5061	0.4946	0.5048	0.4988	0.4999	0.4877
22	0.4827	0.4786	0.4888	0.4951	0.4758	0.5004	0.4873	0.4987	0.4957	0.4978	0.4820
23	0.4809	0.4740	0.4827	0.4899	0.4733	0.4953	0.4841	0.4956	0.4893	0.4943	0.4798
24	0.4815	0.4803	0.4892	0.4929	0.4768	0.5043	0.4905	0.5015	0.4938	0.4979	0.4860
25	0.4830	0.4789	0.4902	0.4981	0.4802	0.5036	0.4891	0.5033	0.4960	0.4963	0.4886
26	0.4330	0.4293	0.4408	0.4480	0.4320	0.4573	0.4464	0.4559	0.4494	0.4525	0.4355
27	0.4761	0.4697	0.4851	0.4900	0.4738	0.4937	0.4825	0.4937	0.4898	0.4924	0.4759
28	0.4746	0.4695	0.4739	0.4834	0.4669	0.4906	0.4791	0.4884	0.4822	0.4861	0.4704
29	0.4743	0.4694	0.4767	0.4833	0.4670	0.4903	0.4803	0.4939	0.4829	0.4871	0.4721
30	0.4773	0.4754	0.4843	0.4898	0.4733	0.4956	0.4832	0.4956	0.4916	0.4948	0.4806
31	0.4690	0.4661	0.4733	0.4812	0.4655	0.4876	0.4789	0.4862	0.4812	0.4901	0.4710
32	0.4728	0.4697	0.4777	0.4852	0.4668	0.4913	0.4789	0.4930	0.4848	0.4923	0.4714
33	0.4733	0.4659	0.4732	0.4833	0.4653	0.4916	0.4774	0.4889	0.4846	0.4862	0.4725
34	0.4688	0.4601	0.4685	0.4797	0.4616	0.4857	0.4720	0.4832	0.4775	0.4799	0.4660
35	0.4598	0.4566	0.4620	0.4742	0.4543	0.4780	0.4636	0.4753	0.4675	0.4714	0.4589
36	0.4620	0.4604	0.4700	0.4771	0.4585	0.4828	0.4714	0.4838	0.4744	0.4794	0.4640
37	0.4554	0.4521	0.4623	0.4664	0.4495	0.4714	0.4602	0.4709	0.4627	0.4699	0.4527
38	0.4653	0.4564	0.4675	0.4728	0.4582	0.4802	0.4668	0.4803	0.4760	0.4770	0.4620
39	0.4594	0.4521	0.4629	0.4713	0.4546	0.4783	0.4650	0.4755	0.4698	0.4720	0.4720
40	0.4491	0.4426	0.4538	0.4601	0.4414	0.4655	0.4511	0.4632	0.4554	0.4583	0.4454

Table A.14 Competition Matrix: Using Regression Fitting Procedure and Original Values of Shape Parameter (Continued)

Handicap	21	22	23	24	25	26	27	28	29	30
10	0.5225	0.5226	0.5231	0.5280	0.5277	0.5305	0.5314	0.5314	0.5330	0.5345
11	0.5176	0.5215	0.5206	0.5229	0.5257	0.5278	0.5270	0.5319	0.5321	0.5361
12	0.5213	0.5218	0.5203	0.5233	0.5225	0.5238	0.5264	0.5254	0.5301	0.5295
13	0.5126	0.5154	0.5159	0.5195	0.5224	0.5233	0.5244	0.5284	0.5290	0.5289
14	0.5132	0.5152	0.5169	0.5222	0.5179	0.5330	0.5280	0.5325	0.5340	0.5266
15	0.4951	0.5005	0.5000	0.4991	0.4986	0.5075	0.5045	0.5096	0.5089	0.5044
16	0.5068	0.5094	0.5153	0.5095	0.5064	0.5226	0.5167	0.5206	0.5230	0.5135
17	0.4961	0.5003	0.5056	0.4985	0.5000	0.5107	0.5030	0.5119	0.5096	0.5052
18	0.5039	0.5059	0.5082	0.5047	0.5033	0.5156	0.5119	0.5129	0.5171	0.5108
19	0.5012	0.5015	0.5075	0.5027	0.5012	0.5145	0.5105	0.5128	0.5138	0.5069
20	0.5125	0.5154	0.5207	0.5155	0.5125	0.5287	0.5234	0.5275	0.5299	0.5200
21	0.5009	0.5034	0.5105	0.5030	0.4992	0.5159	0.5087	0.5164	0.5152	0.5056
22	0.4980	0.5008	0.5069	0.4974	0.4967	0.5104	0.5015	0.5112	0.5105	0.5011
23	0.4918	0.4970	0.5008	0.4923	0.4925	0.5058	0.4994	0.5074	0.5019	0.4991
24	0.4972	0.5007	0.5054	0.5000	0.5002	0.5126	0.5050	0.5130	0.5128	0.5043
25	0.5030	0.5048	0.5084	0.5008	0.4993	0.5137	0.5089	0.5119	0.5140	0.5041
26	0.4530	0.4585	0.4616	0.4543	0.5005	0.5016	0.4939	0.5002	0.4970	0.4916
27	0.4485	0.4930	0.4980	0.4935	0.4875	0.5072	0.5018	0.5084	0.5077	0.4982
28	0.4390	0.4872	0.4934	0.4858	0.4866	0.4991	0.4912	0.4989	0.5028	0.4924
29	0.4868	0.4934	0.4976	0.4876	0.4863	0.4981	0.4978	0.5019	0.4996	0.4898
30	0.4916	0.4945	0.4998	0.4967	0.4951	0.5067	0.5022	0.5119	0.5064	0.5026
31	0.4835	0.4870	0.4909	0.4871	0.4849	0.4977	0.4901	0.4982	0.4974	0.4890
32	0.4850	0.4914	0.4964	0.4893	0.4849	0.5001	0.4975	0.4980	0.5016	0.4917
33	0.4858	0.4883	0.4941	0.4866	0.4839	0.5027	0.4944	0.4995	0.5016	0.4904
34	0.4798	0.4821	0.4874	0.4809	0.4784	0.4973	0.4884	0.4933	0.4935	0.4839
35	0.4772	0.4798	0.4724	0.4725	0.4860	0.4841	0.4795	0.4844	0.4821	0.4771
36	0.4768	0.4828	0.4878	0.4794	0.4930	0.4856	0.4930	0.4924	0.4930	0.4832
37	0.4679	0.4711	0.4740	0.4660	0.4783	0.4824	0.4729	0.4780	0.4796	0.4707
38	0.4755	0.4796	0.4845	0.4786	0.4814	0.4924	0.4880	0.4908	0.4913	0.4814
39	0.4689	0.4760	0.4771	0.4759	0.4752	0.4841	0.4791	0.4848	0.4872	0.4767
40	0.4602	0.4608	0.4677	0.4598	0.4566	0.4724	0.4622	0.4718	0.4713	0.4613

Table A.15 Competition Matrix: Using Regression Fitting Procedure and Original Values of Shape Parameter (Continued)

Handicap	31	32	33	34	35	36	37	38	39	40
10	0.5359	0.5364	0.5416	0.5401	0.5406	0.5451	0.5448	0.5420	0.5462	0.5470
11	0.5318	0.5348	0.5340	0.5371	0.5389	0.5401	0.5392	0.5440	0.5442	0.5436
12	0.5333	0.5352	0.5354	0.5375	0.5383	0.5373	0.5404	0.5426	0.5428	0.5445
13	0.5320	0.5336	0.5312	0.5351	0.5354	0.5368	0.5386	0.5384	0.5380	0.5420
14	0.5368	0.5355	0.5325	0.5399	0.5455	0.5400	0.5509	0.5429	0.5436	0.5581
15	0.5116	0.5069	0.5104	0.5147	0.5234	0.5165	0.5267	0.5185	0.5213	0.5360
16	0.5208	0.5220	0.5220	0.5275	0.5354	0.5254	0.5407	0.5323	0.5359	0.5445
17	0.5134	0.5283	0.5132	0.5147	0.5257	0.5157	0.5269	0.5210	0.5233	0.5351
18	0.5205	0.5131	0.5150	0.5234	0.5321	0.5326	0.5233	0.5331	0.5255	0.5278
19	0.5163	0.5157	0.5148	0.5179	0.5291	0.5205	0.5329	0.5237	0.5270	0.5410
20	0.5274	0.5285	0.5283	0.5345	0.5425	0.5320	0.5477	0.5386	0.5430	0.5515
21	0.5169	0.5145	0.5170	0.5185	0.5308	0.5208	0.5315	0.5252	0.5281	0.5403
22	0.5114	0.5119	0.5110	0.5163	0.5254	0.5165	0.5320	0.5211	0.5242	0.5384
23	0.5103	0.5071	0.5023	0.5103	0.5212	0.5152	0.5249	0.5133	0.5201	0.5318
24	0.5137	0.5151	0.5110	0.5202	0.5272	0.5161	0.5305	0.5208	0.5285	0.5409
25	0.5175	0.5151	0.5148	0.5219	0.5294	0.5200	0.5336	0.5254	0.5290	0.5435
26	0.5009	0.4974	0.4984	0.5067	0.5172	0.5102	0.5212	0.5116	0.5138	0.5283
27	0.5081	0.5045	0.5059	0.5132	0.5206	0.5136	0.5255	0.5142	0.5224	0.5364
28	0.5004	0.4995	0.4989	0.5053	0.5165	0.5036	0.5217	0.5074	0.5134	0.5318
29	0.5048	0.4975	0.5012	0.5077	0.5150	0.5067	0.5216	0.5073	0.5178	0.5310
30	0.5119	0.5082	0.5087	0.5161	0.5254	0.5193	0.5288	0.5277	0.5276	0.5399
31	0.4983	0.4971	0.4938	0.5029	0.5137	0.5057	0.5184	0.5066	0.5106	0.5238
32	0.5038	0.4964	0.5059	0.5065	0.5174	0.5104	0.5218	0.5129	0.5163	0.5319
33	0.5043	0.5006	0.4984	0.5069	0.5170	0.5105	0.5192	0.5112	0.5142	0.5290
34	0.4972	0.4938	0.4918	0.5000	0.5097	0.5031	0.5128	0.5032	0.5065	0.5217
35	0.4824	0.4843	0.4862	0.4921	0.4960	0.4889	0.5060	0.4952	0.4975	0.5142
36	0.4942	0.4927	0.4879	0.5704	0.5126	0.4999	0.5142	0.5023	0.5053	0.5217
37	0.4809	0.4793	0.4766	0.4860	0.4942	0.4868	0.4971	0.4910	0.4937	0.5110
38	0.4937	0.4878	0.4892	0.4966	0.5082	0.4994	0.5137	0.5009	0.5044	0.5172
39	0.4910	0.4853	0.4869	0.4900	0.5040	0.4938	0.5090	0.4953	0.5017	0.5196
40	0.4752	0.4660	0.4672	0.4722	0.4857	0.4770	0.4905	0.4770	0.4815	0.4986

A.6 Competition Matrix Using Regression Fitting Procedure and Average Values of Shape Parameter

Tables A.16 through A.18 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 using regression fitting procedure on scale and location parameters and average values of shape parameter. Each entry in this table is the value of the winning probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.16 lists competitions between handicaps 10 to 20, Table A.17 lists competitions between handicaps 21 to 30 and Table A.18 lists competitions between handicaps 31 to 40.

Table A.16 Competition Matrix: Using Regression Fitting Procedure and Average of the Values of Shape Parameter

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.5019	0.5007	0.5048	0.5064	0.5093	0.5105	0.5129	0.5142	0.5175	0.5180	0.5213
11	0.4966	0.5022	0.5025	0.5062	0.5068	0.5082	0.5092	0.5143	0.5148	0.5156	0.5195
12	0.4961	0.4967	0.5022	0.5028	0.5057	0.5066	0.508	0.5082	0.514	0.5145	0.5152
13	0.4956	0.4943	0.4964	0.4990	0.5000	0.5045	0.5065	0.5069	0.5091	0.5130	0.5137
14	0.4908	0.4929	0.4957	0.4993	0.4979	0.5010	0.5044	0.5043	0.5073	0.5098	0.5115
15	0.4906	0.4906	0.4942	0.4977	0.4978	0.5004	0.5004	0.5042	0.5058	0.5080	0.5092
16	0.4858	0.4901	0.4889	0.4957	0.4937	0.4996	0.4987	0.5021	0.5039	0.5057	0.5073
17	0.4884	0.4883	0.4897	0.4921	0.4944	0.4954	0.5000	0.5005	0.5033	0.5037	0.5054
18	0.4834	0.4835	0.4884	0.4898	0.4910	0.4943	0.4964	0.4993	0.5009	0.5010	0.5036
19	0.4819	0.4820	0.4846	0.4899	0.4904	0.4941	0.4982	0.4953	0.4969	0.4997	0.5026
20	0.4850	0.4836	0.4833	0.4850	0.4885	0.4899	0.4908	0.4960	0.4970	0.4994	0.4972
21	0.4801	0.4787	0.4859	0.4866	0.4862	0.4864	0.4912	0.4932	0.4969	0.4974	0.4962
22	0.4773	0.4778	0.4827	0.4817	0.4829	0.4873	0.4882	0.4893	0.4930	0.4973	0.4964
23	0.4754	0.4783	0.4821	0.4857	0.4869	0.4875	0.4865	0.4888	0.4910	0.4907	0.4970
24	0.4726	0.4775	0.4802	0.4793	0.4786	0.4832	0.4830	0.4875	0.4920	0.4913	0.4931
25	0.4733	0.4738	0.4759	0.4785	0.4816	0.4815	0.4861	0.4872	0.4892	0.4906	0.4911
26	0.4719	0.4722	0.4744	0.4770	0.4799	0.4801	0.4845	0.4854	0.4876	0.4890	0.4893
27	0.4685	0.4726	0.4756	0.4776	0.4812	0.4821	0.4802	0.4836	0.4885	0.4886	0.4909
28	0.4717	0.4716	0.4739	0.4742	0.4773	0.4775	0.4800	0.4833	0.4853	0.4874	0.4888
29	0.4685	0.4695	0.4737	0.4718	0.4748	0.4750	0.4772	0.4811	0.4822	0.4833	0.4841
30	0.4662	0.4678	0.4712	0.4699	0.4733	0.4742	0.4747	0.4793	0.4818	0.4815	0.4837
31	0.4618	0.4672	0.4697	0.4724	0.4723	0.4745	0.4741	0.4760	0.4788	0.4809	0.4842
32	0.4618	0.4645	0.4678	0.4681	0.4703	0.4751	0.4740	0.4764	0.4797	0.4796	0.4822
33	0.4622	0.4634	0.4613	0.4686	0.4669	0.4706	0.4740	0.4735	0.4765	0.4762	0.4811
34	0.4601	0.4596	0.4646	0.4662	0.4645	0.4702	0.4731	0.4750	0.4726	0.4785	0.4807
35	0.4595	0.4626	0.4609	0.4653	0.4657	0.4706	0.4675	0.4724	0.4739	0.4777	0.4796
36	0.4575	0.4595	0.4597	0.4651	0.4648	0.4649	0.4656	0.4680	0.4711	0.4754	0.4743
37	0.4561	0.4576	0.4585	0.4648	0.4608	0.4648	0.4698	0.4697	0.4681	0.4748	0.4742
38	0.4547	0.4558	0.4621	0.4619	0.4620	0.4648	0.4646	0.4680	0.4693	0.4732	0.4731
39	0.4556	0.4553	0.4587	0.4592	0.4610	0.4636	0.4660	0.4671	0.4702	0.4705	0.4728
40	0.4539	0.4551	0.4577	0.4579	0.4600	0.4623	0.4648	0.4661	0.4692	0.4693	0.4718

Table A.17 Competition Matrix: Using Regression Fitting Procedure and Average of the Values of Shape Parameter (Continued)

Handicap	21	22	23	24	25	26	27	28	29	30
10	0.5233	0.5217	0.5253	0.5260	0.5278	0.5309	0.5316	0.5303	0.5348	0.5364
11	0.5189	0.5182	0.5232	0.5241	0.5257	0.5279	0.5306	0.5318	0.5330	0.5303
12	0.5192	0.5185	0.5178	0.5226	0.5237	0.5250	0.5272	0.5297	0.5312	0.5324
13	0.5144	0.5176	0.5171	0.5223	0.5224	0.5221	0.5237	0.5228	0.5278	0.5290
14	0.5134	0.5177	0.5171	0.5166	0.5207	0.5230	0.5267	0.5258	0.5247	0.5269
15	0.5121	0.5128	0.5159	0.5186	0.5163	0.5197	0.5209	0.5227	0.5253	0.5258
16	0.5103	0.5111	0.5138	0.5163	0.5145	0.5178	0.5190	0.5208	0.5238	0.5237
17	0.5065	0.5086	0.5109	0.5102	0.5159	0.5162	0.5164	0.5202	0.5190	0.5213
18	0.5051	0.5087	0.5072	0.5087	0.5111	0.5131	0.5156	0.5148	0.5185	0.5191
19	0.5056	0.5052	0.5064	0.5082	0.5095	0.5111	0.5147	0.5137	0.5153	0.5202
20	0.5011	0.5041	0.506	0.5057	0.5080	0.5122	0.5124	0.5113	0.5148	0.5151
21	0.4991	0.5013	0.5026	0.5026	0.5087	0.5069	0.5116	0.5111	0.5090	0.5119
22	0.4971	0.5018	0.5009	0.5028	0.5056	0.5089	0.5098	0.5096	0.5131	0.5097
23	0.4968	0.5005	0.4983	0.4975	0.5016	0.5040	0.5050	0.5039	0.5120	0.5083
24	0.4935	0.4961	0.4999	0.5000	0.5015	0.5062	0.5050	0.5046	0.5058	0.5095
25	0.4936	0.4953	0.4957	0.5007	0.5001	0.4998	0.5046	0.5043	0.5065	0.5068
26	0.4925	0.4938	0.4940	0.4989	0.4985	0.4983	0.5033	0.5026	0.5056	0.5047
27	0.4937	0.4941	0.4945	0.4984	0.4992	0.4986	0.5028	0.4991	0.5053	0.5061
28	0.4913	0.4921	0.4922	0.4952	0.4962	0.4978	0.5000	0.4966	0.5039	0.5059
29	0.4875	0.4887	0.4909	0.4903	0.4914	0.4990	0.4970	0.4977	0.4995	0.5020
30	0.4867	0.4870	0.4891	0.4906	0.4904	0.4962	0.4959	0.4970	0.4980	0.4985
31	0.4850	0.4868	0.4903	0.4859	0.4903	0.4910	0.4946	0.4937	0.4970	0.4967
32	0.4849	0.4877	0.4860	0.4884	0.4904	0.4909	0.4923	0.4935	0.4955	0.4964
33	0.4831	0.4829	0.4854	0.4901	0.4881	0.4886	0.4891	0.4901	0.4945	0.4976
34	0.4805	0.4823	0.4812	0.4854	0.4874	0.4882	0.4881	0.4923	0.4915	0.4930
35	0.4781	0.4821	0.4825	0.4818	0.4857	0.4899	0.4883	0.4914	0.4913	0.4918
36	0.4794	0.4782	0.4822	0.4842	0.4829	0.4842	0.4861	0.4889	0.4885	0.4935
37	0.4767	0.4766	0.4791	0.4796	0.4855	0.4830	0.4858	0.4851	0.4908	0.4894
38	0.4747	0.4777	0.4798	0.4815	0.4804	0.4844	0.4848	0.4878	0.4867	0.4884
39	0.4757	0.4749	0.4772	0.4782	0.4791	0.4845	0.4830	0.4844	0.4879	0.4864
40	0.4747	0.4737	0.4762	0.4772	0.4778	0.4834	0.4817	0.4832	0.4866	0.4851

Table A.18 Competition Matrix: Using Regression Fitting Procedure and Average of the Values of Shape Parameter (Continued)

Handicap	31	32	33	34	35	36	37	38	39	40
10	0.5372	0.5380	0.5387	0.5408	0.5410	0.5430	0.5418	0.5432	0.5474	0.5475
11	0.5318	0.5369	0.5386	0.5377	0.5382	0.5406	0.5391	0.5400	0.5459	0.5464
12	0.5298	0.5312	0.5362	0.5378	0.5370	0.5366	0.5394	0.5408	0.5411	0.5429
13	0.529	0.5343	0.5360	0.5362	0.5343	0.5352	0.5360	0.5366	0.5402	0.5390
14	0.5272	0.5316	0.5293	0.5333	0.5358	0.5365	0.5367	0.5352	0.5394	0.5368
15	0.5247	0.5289	0.5296	0.5308	0.5324	0.5324	0.5349	0.5348	0.5376	0.5359
16	0.5224	0.5267	0.5269	0.5289	0.5305	0.5305	0.5330	0.5348	0.5341	0.5366
17	0.5236	0.5263	0.5251	0.5296	0.5253	0.5309	0.5296	0.5349	0.5315	0.5339
18	0.5211	0.5223	0.5223	0.5202	0.5280	0.5266	0.5277	0.5273	0.5325	0.5347
19	0.5184	0.5207	0.5212	0.5219	0.5230	0.5255	0.5296	0.5253	0.5259	0.5299
20	0.5191	0.5204	0.5219	0.5190	0.5235	0.5266	0.5296	0.5283	0.5272	0.5283
21	0.5141	0.5203	0.5183	0.5216	0.5207	0.5228	0.5209	0.5254	0.5245	0.5269
22	0.5153	0.5133	0.5159	0.5171	0.5201	0.5198	0.5199	0.5236	0.5263	0.5274
23	0.5096	0.5140	0.5106	0.5193	0.5208	0.5218	0.5218	0.5220	0.5247	0.5227
24	0.5123	0.5133	0.5145	0.5139	0.5137	0.5156	0.5195	0.5210	0.5208	0.5223
25	0.5078	0.5071	0.5112	0.5118	0.5134	0.5170	0.5182	0.5173	0.5191	0.5188
26	0.5077	0.5091	0.5123	0.5091	0.5144	0.5132	0.5143	0.5179	0.5177	0.5176
27	0.5079	0.5101	0.5102	0.5124	0.5122	0.5112	0.5138	0.5173	0.5169	0.5179
28	0.5057	0.5099	0.5078	0.5078	0.5105	0.5107	0.5122	0.5157	0.5173	0.5158
29	0.4996	0.5055	0.5088	0.5074	0.5080	0.5093	0.5095	0.5120	0.5114	0.5131
30	0.5052	0.5064	0.5059	0.5058	0.5054	0.5071	0.5082	0.5108	0.5109	0.5146
31	0.4995	0.4990	0.5006	0.5033	0.5066	0.5059	0.5090	0.5099	0.5091	0.5111
32	0.4963	0.4971	0.4996	0.5032	0.5028	0.5067	0.5059	0.5047	0.5079	0.5124
33	0.4975	0.5001	0.5003	0.4990	0.5006	0.5060	0.5064	0.5058	0.5087	0.5113
34	0.4931	0.4979	0.5002	0.4985	0.5019	0.5050	0.5065	0.5044	0.5055	0.5078
35	0.4936	0.4986	0.4986	0.4992	0.4986	0.5034	0.5022	0.5038	0.5081	0.5071
36	0.4939	0.4940	0.4984	0.4989	0.4958	0.5000	0.5008	0.5007	0.5040	0.5071
37	0.4894	0.4932	0.4948	0.4964	0.5004	0.4997	0.4992	0.5001	0.5040	0.5036
38	0.4901	0.4944	0.4938	0.4939	0.4927	0.4969	0.4988	0.5003	0.5034	0.5029
39	0.4909	0.4918	0.4900	0.4941	0.4947	0.4992	0.4978	0.4974	0.4994	0.5008
40	0.4897	0.4905	0.4885	0.4930	0.4936	0.4977	0.4963	0.4990	0.4989	0.4995

A.7 Competition Matrix Using Regression Fitting Procedure and Moving Average Values of Shape Parameter

Tables A.19 through A.21 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 using regression fitting procedure on scale and location parameters and moving average values of shape parameter. Each entry in this table is the value of the winning probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.19 lists competitions between handicaps 10 to 20, Table A.20 lists competitions between handicaps 21 to 30 and Table A.21 lists competitions between handicaps 31 to 40.

Table A.19 Competition Matrix: Using Regression Fitting Procedure and Moving Average on the Values of Shape Parameter

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.5019	0.5027	0.5042	0.5078	0.5104	0.5134	0.5144	0.5143	0.5133	0.5149	0.5151
11	0.4955	0.4998	0.4990	0.5031	0.5053	0.5064	0.5106	0.5122	0.5089	0.5123	0.5094
12	0.4945	0.5022	0.5008	0.5062	0.5047	0.5085	0.5087	0.5087	0.5089	0.5104	0.5094
13	0.4923	0.4970	0.4940	0.5001	0.5010	0.5052	0.5061	0.5087	0.5053	0.5070	0.5054
14	0.4876	0.4939	0.4976	0.4987	0.4988	0.5031	0.5024	0.5055	0.5050	0.5012	0.5057
15	0.4904	0.4921	0.4904	0.4952	0.4961	0.5004	0.5019	0.5025	0.4988	0.5016	0.5033
16	0.4869	0.4937	0.4933	0.4946	0.4992	0.4966	0.5009	0.5024	0.5040	0.4994	0.5003
17	0.4827	0.4912	0.4868	0.4908	0.4950	0.4974	0.4959	0.4989	0.4961	0.4976	0.5013
18	0.4881	0.4922	0.4908	0.4950	0.4972	0.4981	0.5019	0.5033	0.4983	0.5030	0.5010
19	0.4852	0.4920	0.4900	0.4929	0.4971	0.4990	0.4966	0.5019	0.4978	0.4989	0.5015
20	0.4854	0.4900	0.4865	0.4899	0.4980	0.4977	0.4967	0.4989	0.4956	0.4982	0.5011
21	0.4814	0.4885	0.4845	0.4909	0.4927	0.4973	0.4944	0.4998	0.4987	0.4984	0.4994
22	0.4814	0.4906	0.4875	0.4879	0.4955	0.4975	0.4953	0.5006	0.4954	0.4965	0.4989
23	0.4804	0.4877	0.4836	0.4884	0.4888	0.4917	0.4933	0.5008	0.4938	0.4945	0.4970
24	0.4781	0.4805	0.4803	0.4838	0.4885	0.4884	0.4882	0.4916	0.4911	0.4922	0.4946
25	0.4768	0.4819	0.4803	0.4833	0.4873	0.4880	0.4911	0.4927	0.4907	0.4913	0.4915
26	0.4775	0.4790	0.4795	0.4842	0.4874	0.4896	0.4900	0.4900	0.4888	0.4908	0.4885
27	0.4730	0.4791	0.4770	0.4810	0.4813	0.4864	0.4878	0.4890	0.4868	0.4894	0.4873
28	0.4741	0.4759	0.4756	0.4784	0.4800	0.4848	0.4843	0.4871	0.4852	0.4880	0.4863
29	0.4710	0.4736	0.4753	0.4792	0.4797	0.4853	0.4858	0.4882	0.4832	0.4868	0.4845
30	0.4709	0.4730	0.4743	0.4757	0.4799	0.4833	0.4815	0.4868	0.4820	0.4841	0.4845
31	0.4710	0.4739	0.4731	0.4783	0.4809	0.4829	0.4845	0.4870	0.4837	0.4854	0.4835
32	0.4691	0.4755	0.4724	0.4766	0.4792	0.4806	0.4813	0.4849	0.4828	0.4814	0.4841
33	0.4634	0.4698	0.4664	0.4716	0.4762	0.4764	0.4793	0.4818	0.4789	0.4812	0.4789
34	0.4640	0.4708	0.4693	0.4722	0.4762	0.4786	0.4759	0.4814	0.4752	0.4753	0.4792
35	0.4599	0.4655	0.4605	0.4699	0.4693	0.4731	0.4724	0.4764	0.4748	0.4731	0.4762
36	0.4607	0.4640	0.4605	0.4676	0.4702	0.4719	0.4713	0.4739	0.4710	0.4734	0.4734
37	0.4535	0.5048	0.4621	0.4595	0.4657	0.4682	0.4718	0.4740	0.4698	0.4712	0.4686
38	0.4550	0.5000	0.4592	0.4644	0.4654	0.4665	0.4689	0.4727	0.4674	0.4694	0.4710
39	0.4529	0.4556	0.4592	0.4613	0.4657	0.4662	0.4625	0.4698	0.4629	0.4657	0.4681
40	0.4541	0.4565	0.4602	0.4590	0.4615	0.4625	0.4619	0.4675	0.4662	0.4664	0.4662

Table A.20 Competition Matrix: Using Regression Fitting Procedure and Moving Average on the Values of Shape Parameter (Continued)

Handicap	21	22	23	24	25	26	27	28	29	30
10	0.5192	0.5180	0.5174	0.5229	0.5229	0.5251	0.5274	0.5273	0.5292	0.5311
11	0.5101	0.5133	0.5156	0.5177	0.5198	0.5168	0.5237	0.5252	0.5254	0.5226
12	0.5113	0.5150	0.5155	0.5174	0.5195	0.5196	0.5253	0.5253	0.5247	0.5251
13	0.5083	0.5102	0.5133	0.5156	0.5176	0.5170	0.5189	0.5209	0.5210	0.5216
14	0.5055	0.5068	0.5121	0.5105	0.5155	0.5161	0.5169	0.5189	0.5161	0.5207
15	0.4997	0.5062	0.5082	0.5105	0.5094	0.5136	0.5125	0.5153	0.5165	0.5175
16	0.5044	0.5025	0.5036	0.5086	0.5104	0.5119	0.5155	0.5176	0.5158	0.5157
17	0.4974	0.4977	0.5043	0.5049	0.5068	0.5067	0.5132	0.5125	0.5139	0.5123
18	0.5057	0.5051	0.5052	0.5095	0.5108	0.5137	0.5157	0.5142	0.5148	0.5179
19	0.4985	0.5040	0.5072	0.5042	0.5075	0.5110	0.5151	0.5137	0.5133	0.5160
20	0.5058	0.5008	0.5052	0.5073	0.5074	0.5122	0.5123	0.5132	0.5140	0.5144
21	0.4996	0.5031	0.5027	0.5050	0.5105	0.5084	0.5123	0.5102	0.5084	0.5169
22	0.4991	0.4985	0.5050	0.5067	0.5064	0.5093	0.5119	0.5076	0.5105	0.5129
23	0.4959	0.4999	0.5000	0.5037	0.5032	0.5064	0.5086	0.5072	0.5065	0.5095
24	0.4975	0.4950	0.4991	0.5001	0.4997	0.5470	0.5053	0.5473	0.5053	0.5078
25	0.4935	0.4934	0.4970	0.4981	0.4996	0.5449	0.5492	0.5479	0.5063	0.5073
26	0.4919	0.4940	0.4942	0.4947	0.5014	0.5002	0.5017	0.5048	0.5056	0.5062
27	0.4865	0.4924	0.4924	0.4942	0.4972	0.4989	0.5024	0.4997	0.5014	0.5011
28	0.4872	0.4899	0.4906	0.4929	0.4956	0.4974	0.4982	0.4971	0.5036	0.5035
29	0.4862	0.4880	0.4916	0.4930	0.4909	0.4990	0.5004	0.4984	0.5008	0.5015
30	0.4858	0.4883	0.4870	0.4920	0.4943	0.4951	0.4954	0.4984	0.4959	0.5007
31	0.4894	0.4864	0.4877	0.4928	0.4943	0.5004	0.4972	0.4997	0.4997	0.5027
32	0.4867	0.4831	0.4889	0.4910	0.4953	0.4910	0.4993	0.4971	0.4970	0.4991
33	0.4804	0.4835	0.4849	0.4846	0.4888	0.4842	0.4922	0.4939	0.4950	0.4919
34	0.4759	0.4810	0.4853	0.4869	0.4828	0.4869	0.4916	0.4946	0.4923	0.4916
35	0.4756	0.4776	0.4810	0.4829	0.4830	0.4848	0.4854	0.4863	0.4875	0.4852
36	0.4740	0.4747	0.4744	0.4806	0.4789	0.4836	0.4883	0.4900	0.4857	0.4876
37	0.4723	0.4741	0.4772	0.4808	0.4805	0.4817	0.4810	0.4844	0.4852	0.4833
38	0.4687	0.4702	0.4721	0.4762	0.4754	0.4767	0.4795	0.4815	0.4831	0.4778
39	0.4663	0.4703	0.4699	0.4734	0.4714	0.4742	0.4749	0.4765	0.4797	0.4774
40	0.4678	0.4651	0.4683	0.4741	0.4727	0.4742	0.4791	0.4785	0.4797	0.4775

Table A.21 Competition Matrix: Using Regression Fitting Procedure and Moving Average on the Values of Shape Parameter (Continued)

Handicap	31	32	33	34	35	36	37	38	39	40
10	0.5278	0.5317	0.5324	0.5343	0.5409	0.5404	0.5391	0.5464	0.5460	0.5475
11	0.5235	0.5275	0.5325	0.5310	0.5335	0.5364	0.5353	0.5387	0.5423	0.5433
12	0.5225	0.5281	0.5310	0.5295	0.5376	0.5410	0.5410	0.5406	0.5443	0.5415
13	0.5221	0.5228	0.5273	0.5265	0.5290	0.5330	0.5342	0.5379	0.5398	0.5399
14	0.5201	0.5207	0.5247	0.5272	0.5294	0.5329	0.5343	0.5362	0.5369	0.5377
15	0.5173	0.5206	0.5226	0.5200	0.4853	0.4897	0.5306	0.5319	0.5371	0.5339
16	0.5161	0.5170	0.5226	0.5244	0.5291	0.5284	0.5287	0.5314	0.5348	0.5356
17	0.5119	0.5143	0.5194	0.5207	0.5211	0.5263	0.5270	0.5286	0.5315	0.5319
18	0.5165	0.5169	0.5245	0.5237	0.5288	0.5295	0.5295	0.5325	0.5352	0.5351
19	0.5156	0.5172	0.5217	0.5205	0.5281	0.5286	0.5325	0.5295	0.5349	0.5338
20	0.5157	0.5183	0.5229	0.5206	0.5266	0.5287	0.5260	0.5313	0.5342	0.5368
21	0.5118	0.5125	0.5204	0.5210	0.5221	0.5262	0.5288	0.5313	0.5312	0.5329
22	0.5139	0.5138	0.5174	0.5204	0.5225	0.5209	0.5278	0.5284	0.5332	0.5320
23	0.5102	0.5123	0.5160	0.5176	0.5204	0.5220	0.5255	0.5290	0.5293	0.5280
24	0.5093	0.5100	0.5138	0.5157	0.5189	0.5214	0.5223	0.5246	0.5238	0.5257
25	0.5068	0.5086	0.5133	0.5128	0.5147	0.5177	0.5221	0.5227	0.5276	0.5283
26	0.5037	0.5057	0.5096	0.5151	0.5163	0.5187	0.5180	0.5237	0.5258	0.5247
27	0.5063	0.5046	0.5074	0.5080	0.5149	0.5155	0.5160	0.5218	0.5229	0.5217
28	0.5007	0.5035	0.5072	0.5069	0.5111	0.5148	0.5164	0.5189	0.5245	0.5237
29	0.5009	0.5006	0.5055	0.5079	0.5122	0.5141	0.5192	0.5199	0.5201	0.5211
30	0.4997	0.5044	0.5059	0.5067	0.5117	0.5146	0.5153	0.5175	0.5182	0.5212
31	0.4967	0.5027	0.5057	0.5070	0.5108	0.5146	0.5135	0.5198	0.5205	0.5235
32	0.4988	0.4995	0.5075	0.5063	0.5122	0.5129	0.5138	0.5162	0.5201	0.5208
33	0.4930	0.4951	0.5009	0.4988	0.5068	0.5075	0.5087	0.5116	0.5142	0.5130
34	0.4927	0.4914	0.4982	0.5017	0.5054	0.5052	0.5097	0.5117	0.5166	0.5130
35	0.4895	0.4924	0.4967	0.4947	0.5003	0.5024	0.5029	0.5054	0.5078	0.5101
36	0.4860	0.4868	0.4914	0.4942	0.4958	0.5022	0.4997	0.5056	0.5064	0.5060
37	0.4861	0.4889	0.4930	0.4914	0.4961	0.4956	0.5000	0.5008	0.5070	0.5070
38	0.4811	0.4825	0.4879	0.4892	0.4938	0.4934	0.4960	0.5013	0.5041	0.5030
39	0.4798	0.4814	0.4844	0.4857	0.4874	0.4910	0.4942	0.4960	0.4994	0.4992
40	0.4804	0.4809	0.4864	0.4857	0.4899	0.4941	0.4944	0.4974	0.5005	0.4993

A.8 Competition Matrix for Original Data Set of Scores with At Least One Player from Each Tournament

Tables A.22 through A.24 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 using for original data set of scores with at least one player from each tournament. Each entry in this table is the value of the winning probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.22 lists competitions between handicaps 10 to 20, Table A.23 lists competitions between handicaps 21 to 30 and Table A.24 lists competitions between handicaps 31 to 40.

Table A.22 Competition Matrix for Original Data Set of Scores with At Least One Player from Each Tournament

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.4976	0.5685	0.5027	0.5503	0.5305	0.5440	0.5211	0.4953	0.5533	0.5237	0.5515
11	0.4304	0.4994	0.4347	0.4466	0.4742	0.5081	0.4549	0.4439	0.4871	0.4712	0.5219
12	0.4981	0.5617	0.5004	0.5188	0.5233	0.5463	0.5069	0.4920	0.5222	0.5108	0.5386
13	0.4497	0.5555	0.4809	0.4988	0.5075	0.5387	0.5045	0.4781	0.5206	0.5100	0.5324
14	0.4706	0.5260	0.4739	0.4921	0.5000	0.5289	0.5126	0.4762	0.5256	0.4991	0.5383
15	0.4578	0.4925	0.4526	0.4571	0.4722	0.4967	0.4703	0.4557	0.4858	0.4854	0.4917
16	0.4777	0.5479	0.4934	0.4925	0.4872	0.5281	0.5000	0.4859	0.5196	0.5157	0.5441
17	0.506	0.5519	0.5105	0.5227	0.5254	0.5447	0.5113	0.5014	0.5418	0.5279	0.5616
18	0.4504	0.5113	0.4769	0.4788	0.4747	0.5150	0.4811	0.4557	0.5039	0.4929	0.5270
19	0.4744	0.5300	0.4908	0.4895	0.5011	0.5151	0.4855	0.4713	0.5051	0.5007	0.5275
20	0.4466	0.4895	0.4603	0.4676	0.4599	0.5092	0.4552	0.4367	0.4728	0.4741	0.4963
21	0.4476	0.5114	0.4824	0.4885	0.4949	0.5267	0.4829	0.4731	0.5083	0.5079	0.5277
22	0.4693	0.5014	0.4900	0.4722	0.4828	0.5105	0.4744	0.4614	0.4982	0.4833	0.5000
23	0.4365	0.4995	0.4636	0.4635	0.4780	0.4957	0.4790	0.4544	0.4850	0.4900	0.5023
24	0.4508	0.5077	0.4380	0.4779	0.4556	0.4898	0.4621	0.4184	0.4807	0.4678	0.4909
25	0.4470	0.5085	0.4730	0.4824	0.4700	0.5072	0.4830	0.4757	0.5011	0.4817	0.5191
26	0.4265	0.4955	0.4352	0.4506	0.4543	0.4637	0.4389	0.4413	0.4796	0.4582	0.4954
27	0.4805	0.5508	0.5293	0.5257	0.5059	0.5405	0.5129	0.5051	0.5179	0.5230	0.5457
28	0.4738	0.5041	0.4805	0.5053	0.4857	0.4977	0.4743	0.4669	0.5081	0.497	0.5219
29	0.4892	0.5581	0.504	0.5154	0.5221	0.5334	0.5049	0.5034	0.5454	0.5162	0.5409
30	0.4093	0.4720	0.4212	0.4270	0.4302	0.4524	0.4346	0.4300	0.4461	0.4421	0.4420
31	0.4544	0.5348	0.4790	0.4627	0.4849	0.5230	0.4913	0.4748	0.4993	0.5076	0.5261
32	0.4700	0.5497	0.4877	0.4804	0.5000	0.5473	0.4984	0.4678	0.5045	0.5086	0.5252
33	0.4188	0.4953	0.4300	0.4276	0.4485	0.4847	0.4536	0.4311	0.4784	0.4559	0.4929
34	0.5194	0.5683	0.5662	0.5669	0.5522	0.5792	0.5313	0.5398	0.5823	0.565	0.5829
35	0.4892	0.5000	0.4867	0.5014	0.4826	0.5288	0.4986	0.4716	0.4991	0.4922	0.5249
36	0.4182	0.4371	0.4327	0.4172	0.3887	0.4298	0.4243	0.4135	0.4282	0.4291	0.4407
37	0.4509	0.5242	0.4579	0.4721	0.4411	0.5109	0.4418	0.4716	0.4706	0.4663	0.4997
38	0.4336	0.4841	0.4577	0.4605	0.4420	0.4759	0.4564	0.4350	0.4928	0.4514	0.4928
39	0.3976	0.4158	0.3876	0.3983	0.4017	0.4217	0.4237	0.3616	0.4164	0.3911	0.4103
40	0.3499	0.4245	0.3688	0.3687	0.3805	0.4126	0.3802	0.3541	0.3692	0.3753	0.4113

Table A.23 Competition Matrix for Original Data Set of Scores with At Least One Player from Each Tournament (Continued)

Handicap	21	22	23	24	25	26	27	28	29	30
10	0.5567	0.5315	0.5644	0.5478	0.5519	0.5703	0.5192	0.5304	0.5078	0.5925
11	0.4875	0.5000	0.5014	0.4908	0.4889	0.5064	0.4473	0.4969	0.4422	0.5276
12	0.5180	0.5114	0.5336	0.5595	0.5238	0.5622	0.4712	0.5184	0.4921	0.5805
13	0.5123	0.5266	0.5356	0.5249	0.5180	0.5484	0.4748	0.4897	0.4858	0.5743
14	0.5051	0.5198	0.5171	0.5418	0.5290	0.5463	0.4981	0.5104	0.4763	0.5683
15	0.4717	0.4891	0.5037	0.5084	0.4936	0.5411	0.4603	0.5004	0.4620	0.5506
16	0.5141	0.5265	0.5206	0.5385	0.5187	0.5586	0.4869	0.5288	0.4948	0.5630
17	0.5258	0.5410	0.5427	0.5827	0.5270	0.5581	0.4953	0.5333	0.4955	0.5700
18	0.4918	0.5018	0.5141	0.5195	0.5004	0.5231	0.4813	0.4909	0.4569	0.5556
19	0.4927	0.5194	0.5104	0.5267	0.5199	0.5414	0.4761	0.5007	0.4852	0.5594
20	0.472	0.4959	0.4974	0.5071	0.4799	0.5046	0.4527	0.4797	0.4547	0.5548
21	0.5034	0.5060	0.5189	0.5277	0.5182	0.5229	0.4885	0.5112	0.4704	0.5566
22	0.4963	0.4977	0.5046	0.5173	0.4882	0.5093	0.4821	0.5021	0.4552	0.5422
23	0.4771	0.4965	0.4994	0.5211	0.4750	0.5251	0.4639	0.4961	0.4669	0.5298
24	0.4704	0.4832	0.4780	0.4997	0.4674	0.4929	0.4352	0.4727	0.4688	0.5219
25	0.4800	0.5149	0.5256	0.5289	0.5012	0.5368	0.4677	0.5020	0.4727	0.5331
26	0.4804	0.4904	0.4767	0.5082	0.4644	0.4986	0.4377	0.4786	0.4665	0.5076
27	0.5124	0.5181	0.5374	0.5627	0.5358	0.5608	0.5026	0.5084	0.4985	0.5675
28	0.4929	0.4971	0.5013	0.5272	0.4988	0.5233	0.4906	0.4982	0.4612	0.5563
29	0.5297	0.5396	0.5334	0.5281	0.5261	0.5315	0.5025	0.5353	0.4985	0.556
30	0.4450	0.4598	0.4734	0.4800	0.4640	0.4931	0.4334	0.4415	0.4419	0.4977
31	0.5148	0.514	0.5136	0.5323	0.5023	0.5404	0.4941	0.5012	0.4895	0.5378
32	0.5058	0.5104	0.5173	0.5539	0.5127	0.5408	0.4916	0.5066	0.4556	0.5354
33	0.4555	0.4774	0.4786	0.4830	0.4668	0.5119	0.4343	0.4577	0.4823	0.5326
34	0.5739	0.5685	0.5750	0.5810	0.5717	0.5684	0.5612	0.5717	0.5428	0.5938
35	0.4949	0.5108	0.5047	0.5406	0.5009	0.5490	0.4958	0.5154	0.4868	0.5222
36	0.4267	0.4424	0.4437	0.4744	0.4202	0.4560	0.4157	0.4613	0.4147	0.4645
37	0.4536	0.4642	0.4555	0.4779	0.4568	0.5027	0.4413	0.4641	0.4297	0.5393
38	0.4679	0.4804	0.4754	0.5050	0.4798	0.5016	0.4472	0.4978	0.4339	0.5263
39	0.4107	0.3953	0.4511	0.4142	0.4062	0.4401	0.4024	0.3974	0.4003	0.4515
40	0.3788	0.3742	0.4110	0.4036	0.3839	0.3835	0.3645	0.3956	0.3841	0.4574

Table A.24 Competition Matrix for Original Data Set of Scores with At Least One Player from Each Tournament (Continued)

Handicap	31	32	33	34	35	36	37	38	39	40
10	0.5455	0.5306	0.5795	0.4811	0.5085	0.5828	0.5493	0.5634	0.6006	0.6537
11	0.4634	0.4514	0.5059	0.4282	0.5030	0.5620	0.4803	0.5181	0.5828	0.5763
12	0.5218	0.5051	0.5698	0.4316	0.5131	0.5663	0.5437	0.5450	0.6137	0.6312
13	0.5375	0.5193	0.5700	0.4354	0.4956	0.5831	0.5303	0.5409	0.6014	0.6320
14	0.5197	0.5033	0.5485	0.4506	0.5178	0.6085	0.5583	0.5576	0.6026	0.6198
15	0.4776	0.4557	0.5194	0.4207	0.4703	0.5722	0.4900	0.5213	0.5792	0.5869
16	0.5073	0.5008	0.5500	0.4697	0.5005	0.5759	0.5568	0.5425	0.5812	0.6210
17	0.5245	0.5369	0.5661	0.4601	0.5323	0.5879	0.5278	0.5611	0.6404	0.6440
18	0.4989	0.4935	0.519	0.4167	0.4966	0.57180	0.5284	0.5084	0.5816	0.6315
19	00.4907	0.488	0.5431	0.4338	0.5099	0.5716	0.5292	0.5493	0.6071	0.6234
20	0.4745	0.4754	0.5085	0.4168	0.4716	0.5602	0.5039	0.5082	0.5913	0.6412
21	0.4830	0.4909	0.5447	0.4248	0.5069	0.5769	0.5480	0.5366	0.5935	0.6224
22	0.4886	0.4899	0.5236	0.4311	0.4914	0.5555	0.5362	0.5154	0.6061	0.6225
23	0.4816	0.4795	0.5257	0.4238	0.4964	0.5574	0.5471	0.5238	0.5502	0.5942
24	0.4666	0.4452	0.5192	0.4197	0.4612	0.5232	0.5205	0.4939	0.5843	0.5953
25	0.4932	0.4850	0.5337	0.4307	0.4965	0.5772	0.5470	0.5226	0.5932	0.6167
26	0.4625	0.4534	0.4886	0.4266	0.4471	0.5410	0.4980	0.5012	0.5595	0.6139
27	0.5064	0.5063	0.5656	0.4398	0.5024	0.5865	0.5576	0.5513	0.5938	0.6348
28	0.4934	0.4950	0.5429	0.4279	0.4826	0.5389	0.5343	0.4999	0.6029	0.6084
29	0.5130	0.5434	0.5176	0.4553	0.5165	0.5864	0.5713	0.5668	0.5987	0.6171
30	0.4623	0.4701	0.4659	0.4071	0.4757	0.5324	0.4639	0.4725	0.5504	0.5404
31	0.4994	0.4777	0.5310	0.4604	0.4947	0.5811	0.5036	0.5082	0.5972	0.6248
32	0.5201	0.4979	0.5404	0.4369	0.5112	0.5666	0.5056	0.5384	0.6036	0.6052
33	0.4705	0.4581	0.4969	0.4001	0.4683	0.5145	0.4865	0.5104	0.5504	0.5855
34	0.5435	0.5629	0.6041	0.4978	0.5719	0.6544	0.6035	0.6247	0.6317	0.6621
35	0.5043	0.4902	0.5329	0.4236	0.5051	0.5489	0.5381	0.5381	0.6027	0.6321
36	0.4146	0.4329	0.4833	0.3448	0.4560	0.5002	0.4811	0.4721	0.5276	0.5802
37	0.4973	0.4962	0.5156	0.3969	0.4641	0.5206	0.5028	0.4466	0.5572	0.6336
38	0.4906	0.4657	0.4843	0.3779	0.4632	0.5264	0.5539	0.4964	0.5890	0.6031
39	0.4016	0.3965	0.4510	0.3700	0.3966	0.4718	0.4402	0.4104	0.5025	0.5204
40	0.3776	0.3941	0.4122	0.3375	0.3687	0.4169	0.3676	0.4033	0.4789	0.4971

A.9 Competition Matrix for Original Data Set of Scores with At Least Two Players from Each Tournament

Tables A.25 through A.27 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 using for original data set of scores with at least two players from each tournament. Each entry in this table is the value of the winning probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.25 lists competitions between handicaps 10 to 20, Table A.26 lists competitions between handicaps 21 to 30 and Table A.27 lists competitions between handicaps 31 to 40.

Table A.25 Competition Matrix for Original Data Set of Scores with At Least Two Players from Each Tournament

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.4973	0.5830	0.5623	0.5380	0.5691	0.5840	0.5524	0.5435	0.5891	0.5690	0.5985
11	0.4179	0.4993	0.4408	0.4426	0.4643	0.4857	0.4742	0.4260	0.4838	0.4453	0.5124
12	0.4381	0.5622	0.5028	0.5142	0.5215	0.5600	0.5145	0.4780	0.5194	0.4909	0.5451
13	0.4598	0.5544	0.4833	0.4996	0.5090	0.5526	0.5066	0.4859	0.5277	0.5002	0.5377
14	0.4354	0.5345	0.4810	0.4927	0.5012	0.5324	0.5152	0.4757	0.5115	0.4947	0.5415
15	0.4124	0.5125	0.4399	0.4476	0.4689	0.5010	0.4734	0.4496	0.4757	0.4726	0.4990
16	0.4428	0.5256	0.4842	0.4942	0.4903	0.5261	0.5011	0.4752	0.5205	0.5008	0.5332
17	0.4603	0.5783	0.5212	0.5147	0.5241	0.5515	0.5238	0.5015	0.5469	0.5364	0.5650
18	0.4126	0.5158	0.4783	0.4685	0.4859	0.5247	0.4822	0.4514	0.5014	0.4880	0.5237
19	0.4284	0.5531	0.5092	0.5040	0.5028	0.5250	0.5018	0.4685	0.5132	0.5005	0.5346
20	0.4047	0.4921	0.4536	0.4589	0.4623	0.4999	0.4689	0.4362	0.4732	0.4640	0.4993
21	0.4129	0.5109	0.4758	0.4812	0.4907	0.5203	0.4813	0.4653	0.4929	0.4879	0.5280
22	0.4367	0.5172	0.4786	0.4594	0.4745	0.5104	0.4740	0.4561	0.4906	0.4815	0.5180
23	0.4232	0.4854	0.4829	0.4690	0.4823	0.5105	0.4719	0.4503	0.4929	0.4772	0.5102
24	0.3852	0.5105	0.5604	0.4567	0.4586	0.4833	0.4572	0.4300	0.4738	0.4522	0.4890
25	0.4008	0.5113	0.4826	0.4852	0.4635	0.5174	0.4917	0.4737	0.5049	0.4759	0.5167
26	0.3953	0.4924	0.4526	0.4523	0.4637	0.4931	0.4507	0.4446	0.4830	0.4475	0.4992
27	0.4571	0.5474	0.5274	0.5088	0.5122	0.5506	0.5101	0.4999	0.5205	0.5054	0.5433
28	0.4091	0.5127	0.4947	0.4941	0.4781	0.5033	0.4909	0.4648	0.5035	0.4952	0.5212
29	0.4622	0.5360	0.5081	0.5105	0.5236	0.5543	0.5045	0.5040	0.5435	0.5198	0.5479
30	0.4022	0.4646	0.4326	0.4279	0.4267	0.4722	0.4420	0.4191	0.4443	0.4260	0.4615
31	0.4589	0.5102	0.4697	0.4701	0.4723	0.5135	0.4823	0.4689	0.4922	0.4761	0.5196
32	0.4221	0.5049	0.4499	0.4504	0.4667	0.4991	0.4699	0.4434	0.4843	0.4782	0.5035
33	0.3722	0.4754	0.4262	0.4159	0.4407	0.4762	0.4478	0.4297	0.4481	0.4327	0.4933
34	0.4731	0.5393	0.5191	0.5104	0.5068	0.5348	0.5131	0.5071	0.5373	0.5113	0.5468
35	0.4110	0.4752	0.4628	0.4496	0.4399	0.4961	0.4509	0.4349	0.4720	0.4474	0.4891
36	0.3024	0.4142	0.3918	0.3963	0.3696	0.4205	0.4028	0.3800	0.3982	0.3688	0.4174
37	0.3495	0.4302	0.3844	0.3814	0.3644	0.4348	0.3707	0.3810	0.3866	0.3911	0.4196
38	0.3540	0.4411	0.4141	0.4197	0.4216	0.4392	0.4306	0.4153	0.4376	0.4116	0.4587
39	0.3178	0.3917	0.3617	0.3549	0.3377	0.3815	0.3832	0.3202	0.3771	0.3452	0.3753
40	0.3049	0.3876	0.3291	0.3127	0.3292	0.3665	0.3312	0.3147	0.3208	0.3246	0.3608

Table A.26 Competition Matrix for Original Data Set of Scores with At Least Two Players from Each Tournament (Continued)

Handicap	21	22	23	24	25	26	27	28	29	30
10	0.5868	0.5653	0.5776	0.6132	0.6002	0.6050	0.5440	0.5931	0.5388	0.5977
11	0.4864	0.4837	0.5141	0.4920	0.4841	0.5080	0.4509	0.4855	0.4655	0.5350
12	0.5235	0.5225	0.5173	0.5618	0.5168	0.5478	0.4750	0.5105	0.4930	0.5709
13	0.5177	0.5407	0.5346	0.5373	0.5166	0.5473	0.4875	0.5066	0.4862	0.5672
14	0.5117	0.5236	0.5189	0.5413	0.5356	0.5354	0.4930	0.5248	0.4784	0.5754
15	0.4779	0.4872	0.4868	0.5120	0.4798	0.5044	0.4489	0.4948	0.4506	0.5278
16	0.5168	0.5287	0.5269	0.5442	0.5082	0.5484	0.4895	0.5087	0.4951	0.5563
17	0.5344	0.5431	0.5460	0.5746	0.5254	0.5605	0.4997	0.5338	0.4978	0.5821
18	0.5082	0.5074	0.5068	0.5242	0.4921	0.5144	0.4797	0.4981	0.4556	0.5565
19	0.5175	0.5196	0.5217	0.5509	0.5226	0.5507	0.4961	0.5021	0.4771	0.5728
20	0.4760	0.4827	0.4910	0.5126	0.4822	0.5012	0.4528	0.4801	0.4518	0.5382
21	0.5000	0.5117	0.5140	0.5162	0.5049	0.5133	0.4834	0.4973	0.4615	0.5413
22	0.4887	0.5017	0.5021	0.5122	0.4876	0.5102	0.4632	0.4927	0.4573	0.5355
23	0.4872	0.4955	0.4989	0.5287	0.4916	0.5289	0.4616	0.4803	0.4658	0.5156
24	0.4832	0.4879	0.4763	0.4997	0.4760	0.4990	0.4396	0.4631	0.4473	0.5263
25	0.4965	0.5138	0.5087	0.5231	0.4982	0.5333	0.4660	0.5102	0.4686	0.5347
26	0.4824	0.4884	0.4728	0.5025	0.4666	0.5001	0.4390	0.4846	0.4431	0.5234
27	0.5189	0.5369	0.5386	0.5583	0.5323	0.5585	0.4996	0.5270	0.5105	0.5683
28	0.4982	0.5088	0.5230	0.5379	0.4886	0.5176	0.4740	0.4987	0.4566	0.5467
29	0.5384	0.5458	0.5344	0.5478	0.5328	0.5597	0.4874	0.5465	0.4971	0.5510
30	0.4611	0.4614	0.4855	0.4701	0.4635	0.4755	0.4302	0.4561	0.4484	0.5025
31	0.5032	0.5012	0.5067	0.5281	0.4889	0.5259	0.4663	0.4999	0.4860	0.5331
32	0.4791	0.4824	0.5181	0.5176	0.4899	0.5042	0.4592	0.4878	0.4435	0.5102
33	0.4558	0.4617	0.4769	0.4945	0.4510	0.4849	0.4294	0.4648	0.4462	0.5292
34	0.5285	0.5245	0.5442	0.5512	0.5395	0.5307	0.4920	0.5375	0.4959	0.5660
35	0.4737	0.4652	0.4870	0.5065	0.4818	0.4966	0.4463	0.4854	0.4426	0.5195
36	0.4113	0.4188	0.4198	0.4473	0.4044	0.4275	0.3829	0.4312	0.3857	0.4710
37	0.3913	0.3917	0.3916	0.4053	0.3774	0.4174	0.3449	0.3969	0.3637	0.4740
38	0.4300	0.4565	0.4487	0.4678	0.4513	0.4575	0.4106	0.4683	0.4160	0.4764
39	0.3678	0.3793	0.4009	0.3928	0.3694	0.4025	0.3730	0.3813	0.3671	0.4249
40	0.3346	0.3430	0.3506	0.3667	0.3508	0.3529	0.3245	0.3646	0.3214	0.4071

Table A.27 Competition Matrix for Original Data Set of Scores with At Least Two Players from Each Tournament (Continued)

Handicap	31	32	33	34	35	36	37	38	39	40
10	0.5430	0.5786	0.6305	0.5285	0.5862	0.6974	0.6539	0.6449	0.6835	0.7000
11	0.4931	0.4973	0.5206	0.4560	0.5250	0.5891	0.5703	0.5590	0.6083	0.6155
12	0.5314	0.5513	0.5689	0.4769	0.5411	0.6056	0.6170	0.5847	0.6379	0.6724
13	0.5305	0.5483	0.5820	0.4900	0.5496	0.6058	0.6171	0.5802	0.6447	0.6893
14	0.5275	0.5353	0.5596	0.4911	0.5627	0.6318	0.6362	0.5775	0.6627	0.6722
15	0.4835	0.4998	0.5213	0.4617	0.5034	0.5791	0.5627	0.5621	0.6219	0.6334
16	0.5165	0.5311	0.5516	0.4879	0.5484	0.5956	0.6314	0.5680	0.6194	0.6705
17	0.5292	0.5579	0.5696	0.4935	0.5656	0.6218	0.6189	0.5856	0.6785	0.6852
18	0.5080	0.5165	0.5517	0.4681	0.5251	0.6022	0.6137	0.5661	0.6226	0.6833
19	0.5209	0.5275	0.5632	0.4903	0.5533	0.6304	0.6127	0.5893	0.6555	0.6758
20	0.4810	0.5009	0.5082	0.4505	0.5105	0.5843	0.5810	0.5372	0.6233	0.6374
21	0.4954	0.5197	0.5404	0.4681	0.5303	0.5915	0.6084	0.5669	0.6310	0.6607
22	0.4992	0.5177	0.5386	0.4779	0.5350	0.5832	0.6106	0.5446	0.6230	0.6564
23	0.4964	0.4820	0.5233	0.4551	0.5184	0.5827	0.6111	0.5506	0.5950	0.6488
24	0.4723	0.4797	0.5070	0.4508	0.4914	0.5541	0.5949	0.5362	0.6072	0.6330
25	0.5108	0.5085	0.5511	0.4632	0.5199	0.5964	0.6230	0.5527	0.6269	0.6492
26	0.4744	0.4966	0.5147	0.4695	0.5018	0.5754	0.5821	0.5424	0.5939	0.6460
27	0.5375	0.5411	0.5721	0.5122	0.5547	0.6168	0.6561	0.5900	0.6280	0.6731
28	0.4962	0.5073	0.5350	0.4567	0.5138	0.5692	0.6027	0.5330	0.6175	0.6359
29	0.5106	0.5553	0.5530	0.5046	0.5568	0.6110	0.6376	0.5846	0.6317	0.6777
30	0.4639	0.4936	0.4704	0.4375	0.4840	0.5301	0.5261	0.5196	0.5752	0.5937
31	0.4997	0.5207	0.5456	0.4779	0.5187	0.5851	0.5917	0.5441	0.6115	0.6400
32	0.4807	0.5012	0.5011	0.4537	0.4988	0.5527	0.5681	0.5462	0.6335	0.6150
33	0.4558	0.4957	0.4996	0.4245	0.4961	0.5343	0.5717	0.5369	0.5895	0.5897
34	0.5196	0.5489	0.5737	0.5006	0.5715	0.6320	0.6068	0.5991	0.6569	0.6577
35	0.4819	0.5033	0.5055	0.4311	0.4988	0.5544	0.5910	0.5324	0.5969	0.6233
36	0.4183	0.4410	0.4660	0.3670	0.4430	0.4980	0.5179	0.4557	0.5334	0.5698
37	0.4090	0.4314	0.4287	0.3964	0.4087	0.4831	0.5012	0.3829	0.5267	0.5828
38	0.4552	0.4525	0.4612	0.3979	0.4650	0.5419	0.6138	0.4986	0.5907	0.6077
39	0.3891	0.3623	0.4084	0.3403	0.4041	0.4625	0.4744	0.4078	0.5000	0.5235
40	0.3607	0.3829	0.4111	0.3435	0.3762	0.4310	0.4166	0.3883	0.4814	0.5015

A.10 Competition Matrix for Filtered Data Set of Scores with At Least One Player from Each Tournament

Tables A.28 through A.30 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 using for filtered data set of scores with at least one player from each tournament. Each entry in this table is the value of the winning probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.28 lists competitions between handicaps 10 to 20, Table A.29 lists competitions between handicaps 21 to 30 and Table A.30 lists competitions between handicaps 31 to 40.

Table A.28 Competition Matrix for Filtered Data Set of Scores with At Least One Player from Each Tournament

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.5003	0.5677	0.5360	0.5566	0.5648	0.5696	0.5353	0.5236	0.5738	0.5466	0.6038
11	0.4312	0.4987	0.4714	0.4892	0.5026	0.4996	0.4783	0.45890	0.5074	0.4766	0.5482
12	0.4672	0.5287	0.5044	0.5182	0.5308	0.5319	0.5037	0.4928	0.5369	0.5076	0.5739
13	0.4441	0.5084	0.4811	0.5009	0.5078	0.5091	0.4819	0.4708	0.5138	0.4846	0.5533
14	0.4329	0.4973	0.4709	0.4916	0.5034	0.4975	0.4756	0.4626	0.5025	0.4769	0.5425
15	0.4288	0.5003	0.4711	0.4881	0.5030	0.4994	0.4765	0.4578	0.5077	0.4778	0.5474
16	0.4583	0.5231	0.4945	0.5150	0.5269	0.5196	0.5006	0.4864	0.5333	0.4998	0.5636
17	0.4786	0.5404	0.5078	0.5272	0.5406	0.5391	0.5143	0.5004	0.5443	0.5170	0.5769
18	0.4267	0.4943	0.4654	0.4881	0.4950	0.4911	0.4710	0.4550	0.4980	0.4723	0.5402
19	0.4577	0.5223	0.4943	0.5115	0.5218	0.5212	0.4976	0.4825	0.5282	0.4983	0.5608
20	0.3973	0.4517	0.4304	0.4483	0.4573	0.4571	0.4363	0.4218	0.4629	0.4361	0.5012
21	0.4277	0.4882	0.4597	0.4797	0.4871	0.4848	0.4662	0.4539	0.4913	0.4673	0.5310
22	0.4360	0.4937	0.4713	0.4890	0.4930	0.4906	0.4706	0.4634	0.4991	0.4764	0.5322
23	0.4367	0.4863	0.4663	0.4888	0.4892	0.4945	0.4740	0.4583	0.4953	0.4727	0.5317
24	0.4086	0.4663	0.4434	0.4695	0.4759	0.4703	0.4481	0.4363	0.4777	0.4524	0.5186
25	0.4526	0.5043	0.4796	0.5029	0.5094	0.5079	0.4887	0.4743	0.5134	0.4897	0.5539
26	0.4176	0.4770	0.4527	0.4756	0.4863	0.4791	0.4605	0.4438	0.4912	0.4581	0.5263
27	0.4775	0.5375	0.5088	0.5294	0.5362	0.5378	0.5142	0.5027	0.5410	0.5138	0.5720
28	0.4434	0.4938	0.4715	0.4921	0.4961	0.4970	0.4764	0.4655	0.5012	0.4797	0.5328
29	0.4654	0.5168	0.4949	0.5153	0.5215	0.5205	0.4963	0.4855	0.5274	0.5030	0.5585
30	0.4086	0.4560	0.4374	0.4595	0.4657	0.4623	0.4477	0.4354	0.4678	0.4444	0.5049
31	0.4704	0.5213	0.4974	0.5150	0.5287	0.5274	0.5049	0.4924	0.5320	0.5069	0.5639
32	0.4414	0.4835	0.4696	0.4873	0.4936	0.4941	0.4754	0.4658	0.5002	0.4761	0.5278
33	0.3929	0.4397	0.4250	0.4450	0.4481	0.4496	0.4306	0.4208	0.4573	0.4305	0.4919
34	0.4622	0.5090	0.4886	0.5081	0.5126	0.5112	0.4905	0.4825	0.5136	0.4944	0.5441
35	0.4172	0.4620	0.4432	0.4568	0.4628	0.4682	0.4468	0.4389	0.4723	0.4475	0.5037
36	0.3746	0.4229	0.4038	0.4221	0.4223	0.4241	0.4070	0.3976	0.4310	0.4070	0.4587
37	0.3859	0.4226	0.4072	0.4226	0.4215	0.4476	0.4166	0.4031	0.4356	0.4191	0.4525
38	0.3850	0.4091	0.4146	0.4231	0.4252	0.4536	0.4241	0.4144	0.4352	0.4259	0.4602
39	0.3684	0.3865	0.3832	0.3928	0.3901	0.4240	0.3877	0.3796	0.4035	0.3934	0.4179
40	0.3605	0.4046	0.3888	0.3957	0.3920	0.4279	0.3886	0.3844	0.4075	0.3934	0.4236

Table A.29 Competition Matrix for Filtered Data Set of Scores with At Least One Player from Each Tournament(Continued)

Handicap	21	22	23	24	25	26	27	28	29	30
10	0.5687	0.5646	0.5625	0.5903	0.5529	0.5834	0.5234	0.5596	0.5344	0.5882
11	0.5152	0.5101	0.5113	0.5318	0.4970	0.5248	0.4626	0.5063	0.4851	0.5459
12	0.5367	0.5341	0.5338	0.5586	0.5190	0.5474	0.4940	0.5301	0.5079	0.5605
13	0.5178	0.5111	0.5121	0.5297	0.4967	0.5266	0.4732	0.5080	0.4840	0.5395
14	0.5116	0.5067	0.5076	0.5246	0.4896	0.5142	0.4608	0.5025	0.4813	0.5367
15	0.5132	0.5059	0.5075	0.5295	0.4923	0.5205	0.4671	0.5075	0.4769	0.5371
16	0.5339	0.5254	0.5305	0.5466	0.5124	0.5406	0.4864	0.5257	0.5027	0.5544
17	0.5447	0.5363	0.5365	0.5652	0.5246	0.5580	0.5028	0.5328	0.5134	0.5642
18	0.5094	0.4995	0.5011	0.5155	0.4859	0.5116	0.4589	0.4971	0.4748	0.533
19	0.5319	0.5267	0.5271	0.5475	0.5124	0.5385	0.4868	0.5241	0.4971	0.556
20	0.4685	0.4672	0.4693	0.4821	0.8582	0.4736	0.4251	0.4657	0.4443	0.4950
21	0.4975	0.4944	0.4982	0.5157	0.4793	0.5069	0.4572	0.4937	0.4687	0.5206
22	0.5056	0.4988	0.5023	0.5146	0.4826	0.5045	0.4630	0.4961	0.4732	0.5253
23	0.5003	0.4988	0.5002	0.5149	0.4776	0.5046	0.4558	0.4924	0.4737	0.5197
24	0.4868	0.4858	0.4845	0.4988	0.4664	0.4877	0.4414	0.4843	0.4579	0.5122
25	0.5200	0.5168	0.5214	0.5360	0.4983	0.5265	0.4733	0.5147	0.4948	0.5433
26	0.5001	0.4953	0.4957	0.5144	0.4754	0.5023	0.4478	0.4919	0.4658	0.5213
27	0.5422	0.5366	0.5376	0.5594	0.5251	0.5540	0.5000	0.5369	0.5169	0.5650
28	0.5071	0.5041	0.5069	0.5180	0.4871	0.5074	0.4646	0.4977	0.4763	0.5300
29	0.5306	0.5268	0.5250	0.5445	0.5071	0.5302	0.4862	0.5257	0.5013	0.5489
30	0.4772	0.4761	0.4798	0.4861	0.4568	0.4785	0.4313	0.4713	0.4536	0.4984
31	0.5348	0.5277	0.5320	0.5530	0.5172	0.5431	0.4919	0.5272	0.5064	0.5547
32	0.5064	0.5012	0.5016	0.5187	0.4851	0.5089	0.4646	0.5012	0.4796	0.5261
33	0.4631	0.4562	0.4641	0.4764	0.4453	0.4649	0.4221	0.4582	0.4365	0.4913
34	0.5198	0.5156	0.5174	0.5293	0.5016	0.5244	0.4811	0.5085	0.4925	0.5379
35	0.4770	0.4663	0.4663	0.4838	0.4506	0.4817	0.4379	0.4680	0.4520	0.4924
36	0.4381	0.4360	0.4370	0.4423	0.4210	0.4400	0.4011	0.4390	0.4146	0.4604
37	0.4312	0.4255	0.4385	0.4561	0.4216	0.4429	0.4028	0.4301	0.4104	0.4507
38	0.4372	0.4417	0.4564	0.4636	0.4325	0.4474	0.4076	0.4426	0.4207	0.4660
39	0.4023	0.3994	0.4114	0.4242	0.3907	0.4096	0.3766	0.4006	0.3805	0.4199
40	0.4071	0.4032	0.4119	0.4221	0.4000	0.4124	0.3826	0.4015	0.3835	0.4241

Table A.30 Competition Matrix for Filtered Data Set of Scores with At Least One Player from Each Tournament(Continued)

Handicap	31	32	33	34	35	36	37	38	39	40
10	0.5353	0.5558	0.6037	0.5394	0.5797	0.6268	0.6140	0.6192	0.6421	0.643
11	0.4816	0.5163	0.5589	0.4901	0.5349	0.5776	0.5707	0.5899	0.6171	0.5956
12	0.4998	0.5285	0.5780	0.5132	0.5601	0.5970	0.5893	0.5967	0.6265	0.6161
13	0.4818	0.5094	0.5544	0.4892	0.5383	0.5383	0.5720	0.5754	0.6095	0.6007
14	0.4794	0.5083	0.5494	0.4884	0.5365	0.5757	0.5653	0.5719	0.6092	0.5978
15	0.4744	0.5071	0.5501	0.4889	0.5299	0.5790	0.5646	0.5749	0.6090	0.5948
16	0.4949	0.5242	0.5674	0.5067	0.5528	0.5922	0.5822	0.5864	0.6237	0.6178
17	0.5085	0.5357	0.5803	0.5180	0.5619	0.6007	0.5949	0.5908	0.6294	0.6180
18	0.4673	0.5029	0.5443	0.4870	0.5265	0.5683	0.5599	0.5690	0.6039	0.5915
19	0.4935	0.5211	0.5699	0.5076	0.5534	0.5911	0.5869	0.5874	0.6203	0.6138
20	0.4339	0.4659	0.5069	0.4525	0.4967	0.5413	0.5297	0.5330	0.5837	0.5587
21	0.4623	0.4965	0.5356	0.4836	0.5289	0.5620	0.5561	0.5582	0.5987	0.5839
22	0.4714	0.5000	0.5423	0.4839	0.5320	0.5631	0.5577	0.5571	0.6056	0.5904
23	0.4691	0.4974	0.5356	0.4840	0.5306	0.5649	0.5595	0.5532	0.6029	0.5879
24	0.4512	0.4786	0.5228	0.4715	0.5126	0.5531	0.5448	0.5458	0.5891	0.5824
25	0.4815	0.5161	0.5564	0.5023	0.5476	0.5820	0.5744	0.5761	0.6206	0.6044
26	0.4536	0.4900	0.5355	0.4786	0.5193	0.5632	0.5493	0.5570	0.5964	0.5855
27	0.5088	0.5404	0.5817	0.5231	0.5632	0.5999	0.5975	0.5990	0.6361	0.6198
28	0.4707	0.4981	0.5420	0.4925	0.5320	0.5628	0.5608	0.5518	0.6038	0.5951
29	0.4931	0.5222	0.5644	0.5097	0.5447	0.5853	0.5788	0.5771	0.6230	0.6098
30	0.4455	0.4742	0.5116	0.4616	0.5095	0.5380	0.5348	0.5335	0.5873	0.5710
31	0.4984	0.5298	0.5675	0.5127	0.5586	0.5942	0.5889	0.5934	0.6312	0.6135
32	0.4721	0.5012	0.5344	0.4872	0.5258	0.5656	0.5606	0.5556	0.6006	0.5841
33	0.4280	0.4640	0.4987	0.4474	0.4964	0.5268	0.5262	0.5249	0.5785	0.5538
34	0.4881	0.5139	0.5512	0.5002	0.5434	0.5696	0.5711	0.5659	0.6071	0.5949
35	0.4423	0.4755	0.5022	0.4595	0.4981	0.5340	0.5352	0.5275	0.5815	0.5588
36	0.4077	0.4358	0.4693	0.4277	0.4662	0.4999	0.4939	0.4866	0.5451	0.5296
37	0.4116	0.4280	0.4575	0.4171	0.4532	0.4919	0.5022	0.4788	0.5287	0.5166
38	0.4163	0.4399	0.4660	0.4300	0.4715	0.5140	0.5219	0.4986	0.5621	0.5385
39	0.3839	0.3972	0.4205	0.3930	0.4195	0.4571	0.4754	0.4385	0.5008	0.4910
40	0.3985	0.4063	0.4409	0.3946	0.4388	0.4675	0.4844	0.4622	0.5072	0.5011

A.11 Competition Matrix for Filtered Data Set of Scores with At Least Two Players from Each Tournament

Tables A.31 through A.33 are the competition matrices reflecting the competitions between players of handicaps 10 through 40 using for filtered data set of scores with at least two players from each tournament. Each entry in this table is the value of the winning probability the golfer with handicap as indicated by the row emerging as the winner of a competition with a golfer with handicap as indicated by the column. Due to space limitation, it is divided into three tables where Table A.31 lists competitions between handicaps 10 to 20, Table A.32 lists competitions between handicaps 21 to 30 and Table A.33 lists competitions between handicaps 31 to 40.

Table A.31 Competition Matrix for Filtered Data Set of Scores with At Least Two Players from Each Tournament

Handicap	10	11	12	13	14	15	16	17	18	19	20
10	0.5023	0.5875	0.5615	0.5464	0.574	0.5826	0.5521	0.5311	0.5938	0.5623	0.6153
11	0.4143	0.4983	0.4435	0.4570	0.4744	0.4793	0.4824	0.4265	0.4930	0.4478	0.5284
12	0.4400	0.5559	0.5023	0.5306	0.5241	0.5485	0.5221	0.4854	0.5244	0.4923	0.5664
13	0.4517	0.5442	0.4686	0.4970	0.4953	0.5186	0.4811	0.4707	0.5103	0.4925	0.5477
14	0.4243	0.5225	0.4733	0.5000	0.4964	0.5088	0.5053	0.4709	0.5083	0.4790	0.5479
15	0.4206	0.5234	0.4509	0.4809	0.4900	0.5006	0.4880	0.4606	0.4954	0.4813	0.5307
16	0.4490	0.5185	0.4797	0.5211	0.4998	0.5108	0.5020	0.4735	0.5208	0.4954	0.5491
17	0.4704	0.5710	0.5181	0.5304	0.5317	0.5400	0.5260	0.5013	0.5510	0.5333	0.5811
18	0.4073	0.5075	0.4741	0.4871	0.4880	0.5032	0.4795	0.4500	0.4973	0.4777	0.5368
19	0.4374	0.5512	0.5096	0.5079	0.5165	0.5174	0.5063	0.4663	0.5229	0.5001	0.5532
20	0.3865	0.4722	0.4346	0.4530	0.4526	0.4682	0.4532	0.4191	0.4602	0.4453	0.4995
21	0.4034	0.4939	0.4643	0.4894	0.4845	0.4937	0.4714	0.4538	0.4840	0.4689	0.5284
22	0.4335	0.5215	0.4781	0.4716	0.4814	0.4976	0.4790	0.4587	0.4951	0.4737	0.5371
23	0.4149	0.4844	0.4744	0.4925	0.4887	0.4966	0.4674	0.4530	0.4931	0.4730	0.5226
24	0.3792	0.5025	0.4350	0.4723	0.4600	0.4707	0.4475	0.4224	0.4742	0.4391	0.5029
25	0.4007	0.5116	0.4829	0.5054	0.4741	0.4976	0.4836	0.4775	0.5097	0.4706	0.5339
26	0.3879	0.5002	0.4506	0.4791	0.4763	0.4765	0.4498	0.4398	0.4905	0.4550	0.5190
27	0.4506	0.5338	0.5148	0.5211	0.5125	0.5291	0.5050	0.4900	0.5167	0.4974	0.5528
28	0.4017	0.5122	0.4889	0.4962	0.4800	0.4889	0.4875	0.4660	0.5027	0.4893	0.5346
29	0.4578	0.5157	0.4947	0.5288	0.5200	0.5220	0.4925	0.4926	0.5379	0.5123	0.555
30	0.3876	0.4396	0.4126	0.4281	0.4253	0.4442	0.4326	0.4029	0.4352	0.4082	0.4669
31	0.4677	0.5157	0.4737	0.4973	0.4876	0.5133	0.4948	0.4750	0.4986	0.4795	0.5466
32	0.4084	0.4790	0.4377	0.4558	0.4644	0.4678	0.4563	0.4300	0.4693	0.4613	0.5022
33	0.3583	0.4579	0.4085	0.4006	0.4367	0.4460	0.4326	0.4198	0.4354	0.4144	0.4969
34	0.4659	0.5301	0.5047	0.5044	0.5057	0.5125	0.5000	0.4915	0.5309	0.4984	0.5518
35	0.4099	0.4683	0.4512	0.4606	0.4444	0.4778	0.4431	0.4339	0.4726	0.4410	0.5032
36	0.2934	0.3901	0.3749	0.3866	0.3684	0.3815	0.3848	0.3714	0.3912	0.3642	0.4248
37	0.3581	0.4310	0.3852	0.4193	0.3874	0.4170	0.3878	0.3846	0.3974	0.3977	0.4447
38	0.3443	0.4353	0.4090	0.4356	0.4268	0.4230	0.4252	0.4036	0.4362	0.4050	0.4711
39	0.2957	0.3743	0.3445	0.3657	0.3310	0.3529	0.3707	0.3078	0.3664	0.3288	0.3771
40	0.2936	0.3661	0.3166	0.3075	0.3319	0.3384	0.3296	0.3058	0.3121	0.3126	0.3709

Table A.32 Competition Matrix for Filtered Data Set of Scores with At Least Two Players from Each Tournament (Continued)

Handicap	21	22	23	24	25	26	27	28	29	30
10	0.5970	0.5669	0.5813	0.6163	0.6000	0.6128	0.5493	0.5995	0.5451	0.6134
11	0.5052	0.4814	0.5151	0.4981	0.4853	0.5003	0.4654	0.4881	0.4822	0.5603
12	0.5369	0.5223	0.5251	0.5645	0.5167	0.5482	0.4879	0.5135	0.5055	0.5840
13	0.5102	0.5288	0.5055	0.5240	0.4967	0.5219	0.4814	0.5011	0.4708	0.5721
14	0.5153	0.5201	0.5107	0.5386	0.5276	0.5238	0.4905	0.5178	0.4786	0.5758
15	0.5051	0.5048	0.5032	0.5270	0.4982	0.5216	0.4699	0.5072	0.4790	0.5577
16	0.5258	0.5232	0.5329	0.5548	0.5161	0.5498	0.4941	0.5130	0.5075	0.5694
17	0.5501	0.5404	0.5478	0.5766	0.5230	0.5603	0.5131	0.5359	0.5083	0.5943
18	0.5174	0.5045	0.5071	0.5237	0.4874	0.5110	0.4870	0.4989	0.4627	0.5637
19	0.5324	0.5253	0.5287	0.5626	0.5250	0.5478	0.5076	0.5115	0.4887	0.5900
20	0.4706	0.4633	0.4785	0.4988	0.4664	0.4848	0.4478	0.4639	0.4483	0.5307
21	0.4986	0.4976	0.5039	0.5117	0.4937	0.5024	0.4793	0.4910	0.4621	0.5366
22	0.5028	0.4984	0.5037	0.5172	0.4862	0.5140	0.4726	0.4930	0.4681	0.5540
23	0.4960	0.4933	0.5009	0.5263	0.4910	0.5225	0.4706	0.4816	0.4765	0.5265
24	0.4891	0.4801	0.4730	0.5003	0.4702	0.4918	0.4457	0.4616	0.4573	0.5342
25	0.5109	0.5157	0.5111	0.5312	0.4988	0.5301	0.4755	0.5162	0.4769	0.5439
26	0.4996	0.4877	0.4763	0.5071	0.4686	0.4998	0.4515	0.4859	0.4534	0.5477
27	0.5214	0.5281	0.5278	0.5537	0.5249	0.5457	0.4976	0.5212	0.5152	0.5711
28	0.5109	0.5041	0.5218	0.5393	0.4874	0.5142	0.4790	0.4991	0.4608	0.5568
29	0.5392	0.5300	0.5220	0.5449	0.5206	0.5457	0.4837	0.5401	0.4994	0.5507
30	0.4606	0.4446	0.4726	0.4623	0.4540	0.4508	0.4292	0.4402	0.4508	0.4996
31	0.5222	0.5013	0.5251	0.5425	0.5045	0.5296	0.4837	0.5138	0.5163	0.5590
32	0.4829	0.4638	0.5025	0.5098	0.4729	0.4836	0.4578	0.4776	0.4439	0.5050
33	0.4579	0.4456	0.4670	0.4811	0.4346	0.4615	0.4204	0.4548	0.4447	0.5261
34	0.5312	0.5131	0.5312	0.5387	0.5246	0.5196	0.4845	0.5356	0.4960	0.5648
35	0.4828	0.4552	0.4812	0.5074	0.4727	0.4866	0.4442	0.4807	0.4504	0.5221
36	0.4155	0.4025	0.4058	0.4425	0.3923	0.4073	0.3816	0.4232	0.3897	0.4764
37	0.4157	0.3917	0.3960	0.4135	0.3873	0.4150	0.3562	0.3999	0.3851	0.4987
38	0.4354	0.4445	0.4462	0.4645	0.4346	0.4525	0.4047	0.4627	0.4203	0.4792
39	0.3659	0.3598	0.3880	0.3790	0.3499	0.3845	0.3642	0.3680	0.3649	0.4226
40	0.3416	0.3315	0.3386	0.3584	0.3428	0.3362	0.3278	0.3579	0.3288	0.4173

Table A.33 Competition Matrix for Filtered Data Set of Scores with At Least Two Players from Each Tournament (Continued)

Handicap	31	32	33	34	35	36	37	38	39	40
10	0.5313	0.5923	0.6419	0.5348	0.5936	0.7020	0.6408	0.6554	0.7037	0.7070
11	0.4813	0.5177	0.5446	0.4686	0.5343	0.6090	0.5719	0.5682	0.6242	0.6355
12	0.5225	0.5608	0.5935	0.4977	0.5476	0.6260	0.6140	0.5926	0.6540	0.6778
13	0.4991	0.5486	0.5997	0.4903	0.5380	0.6121	0.5802	0.5627	0.6379	0.6965
14	0.5117	0.5339	0.5629	0.4931	0.5543	0.6320	0.6096	0.5721	0.6696	0.6697
15	0.4862	0.5285	0.5551	0.4854	0.5249	0.6172	0.5799	0.5795	0.6466	0.6631
16	0.5069	0.5429	0.5692	0.5026	0.5559	0.6156	0.6168	0.5730	0.6350	0.6729
17	0.5259	0.5698	0.5825	0.5051	0.5668	0.6295	0.6169	0.5988	0.6918	0.6910
18	0.5018	0.5294	0.5603	0.4734	0.5259	0.6118	0.6013	0.5633	0.6326	0.6875
19	0.5212	0.5409	0.5845	0.5035	0.5591	0.6391	0.5995	0.5948	0.6698	0.6883
20	0.4550	0.4975	0.5041	0.4482	0.5009	0.5767	0.5539	0.5327	0.6240	0.6272
21	0.4744	0.5183	0.5400	0.4700	0.5197	0.5874	0.5820	0.5619	0.6339	0.6560
22	0.4971	0.5383	0.5567	0.4878	0.5473	0.5958	0.6094	0.5586	0.6403	0.6703
23	0.4711	0.4980	0.5331	0.4659	0.5160	0.5908	0.6075	0.5554	0.6092	0.6594
24	0.4579	0.4902	0.5170	0.4631	0.4909	0.5625	0.5858	0.5356	0.6187	0.6409
25	0.4945	0.5262	0.5637	0.4725	0.5294	0.6055	0.6110	0.5630	0.6517	0.6563
26	0.4704	0.5176	0.5362	0.4795	0.5175	0.5928	0.5843	0.5453	0.6150	0.6628
27	0.5225	0.5441	0.5782	0.5162	0.5535	0.6187	0.6436	0.5925	0.632	0.6708
28	0.4876	0.5227	0.5501	0.4666	0.5206	0.5768	0.5964	0.5369	0.6284	0.6423
29	0.4844	0.5587	0.5568	0.5076	0.5476	0.6075	0.6193	0.5798	0.6324	0.6738
30	0.4404	0.4927	0.4710	0.4377	0.4786	0.5252	0.4977	0.5178	0.5777	0.5821
31	0.4994	0.5492	0.5718	0.5009	0.5363	0.6077	0.6000	0.5699	0.6503	0.6557
32	0.4483	0.4991	0.5016	0.4517	0.4878	0.5580	0.5496	0.5436	0.6391	0.6093
33	0.4270	0.4972	0.4981	0.4254	0.4855	0.5290	0.5439	0.5339	0.5930	0.5786
34	0.4995	0.5445	0.5756	0.4986	0.5600	0.6248	0.5793	0.5933	0.6597	0.6510
35	0.4645	0.5118	0.5127	0.4409	0.4983	0.5576	0.5826	0.5422	0.6049	0.6305
36	0.3940	0.4437	0.4699	0.3749	0.4422	0.4985	0.5003	0.4584	0.5420	0.5696
37	0.4015	0.4516	0.4555	0.4236	0.4196	0.4958	0.4975	0.3992	0.5538	0.5950
38	0.4265	0.4522	0.4675	0.4071	0.4633	0.5463	0.6041	0.4975	0.5993	0.6103
39	0.3490	0.3631	0.4089	0.3404	0.3946	0.4551	0.4430	0.4004	0.4989	0.5145
40	0.3437	0.3927	0.4232	0.3454	0.3697	0.4319	0.3994	0.3913	0.4833	0.5028

BIBLIOGRAPHY

[1] D. N. P. Murthy and I. Djamaludin. New product warranty - a literature review. *International Journal of Production Economics*, 79 (3): 231-260, 2002.

[2] W. R. Blischke, M. R. Karim and D. N. P. Murthy. *Warranty Data Collection and Analysis*. Heidelberg, NY: Springer, 2011.

[3] D. N. P. Murthy. Product Warranty and Reliability. *Annals of Operations Research*, 143 (1): 133-146, 2006.

[4] W. R. Blischke and D. N. P. Murthy. *Warranty Cost Analysis*. New York, NY: Marcel Dekker, Inc., 1994.

[5] W. R. Blischke and D. N. P. Murthy. *Product Warranty Handbook*. New York, NY: Marcel Dekker, Inc., 1996.

[6] W. R. Blischke. Mathematical models for analysis of warranty policies. *Mathematical and Computer Modelling*, 13 (7): 1-16, 1990.

[7] W. R. Blischke and D. N. P. Murthy. Product warranty management - I: A taxonomy for warranty policies. *European Journal of Operational Research*, 62 (2): 27-148, 1992.

[8] D. N. P. Murthy and W. R. Blischke. Product warranty management - II: An integrated framework for study. *European Journal of Operational Research*, 62 (3): 261-281, 1992.

[9] D. N. P. Murthy and W. R. Blischke. Product warranty management - III: A review of mathematical models. *European Journal of Operational Research*, 63 (1): 1-34, 1992.

[10] F. M. Biedenweg. *Warranty Analysis: Consumer Value Versus Manufacturers Cost*. Unpublished PhD thesis, Stanford University, USA, 1981.

[11] D. G. Nguyen and D. N. P. Murthy. An optimal policy for servicing warranty. *Journal of Operational Research Society*, 37 (11): 1081-1088, 1986.

[12] D. G. Nguyen and D. N. P. Murthy. Optimal replace-repair strategy for servicing items sold with warranty. *European Journal of Operational Research*, 39 (2): 206-212, 1989.

[13] N. Jack and F. A. Van der Duyn Schouten. Optimal repair-replace strategies for a warranted product. *International Journal of Production Economics*, 67 (1): 95-100, 2000.

[14] X. Jiang, A. K. S. Jardine, and D. Lugitigheid. On a conjecture of optimal repair replacement strategies for warranted products. *Mathematical and Computer Modelling*, 44 (11-12): 963-972, 2006.

[15] N. Jack and D. N. P. Murthy. A servicing strategy for items sold under warranty. *Journal of Operational Research Society*, 52 (11): 1284-1288, 2001.

[16] R. Banerjee. *Some Contributions to Modeling Usage Sensitive Strategies and their Analysis*. Ph.D. Thesis, New Jersey Institute of Technology and Rutgers University, 2010.

[17] B. P. Iskandar, D. N. P. Murthy, and N. Jack. A new repair-replacement strategy for items sold with a two-dimensional warranty. *Computers and Operations Research*, 32 (3): 669-682, 2005.

[18] N. Jack, B. P. Iskandar, and D. N. P. Murthy. A repair-replace strategy based on usage rate for items sold with a two-dimensional warranty. *Reliability Engineering and System Safety*, 94 (2): 611-617, 2009.

[19] W. Y. Yun, D. N. P. Murthy, and N. Jack. Warranty servicing with imperfect repair. *International Journal of Production Economics*, 111 (1): 159-169, 2008.

[20] S. Chukova and M. R. Johnston. Two-dimensional warranty repair strategy based on minimal and complete repairs. *Mathematical and Computer Modelling*, 44 (11-12): 1133-1143, 2006.

[21] S. Varnosafaderani and S. Chukova. 2D warranty: An imperfect repair strategy. In McGraw Hill, editor, *Proceedings of 4th Asia-Pacific International Symposium on Advanced Reliability and Maintenance Modeling (APARM 2010)*, 716-723, 2010.

[22] B. P. Iskandar. *Modelling and Analysis of Two-Dimensional Warranties Policies*. Doctoral Dissertation, The University of Queensland, Australia, 1993.

[23] D. N. P. Murthy, I. Djamaludin and R. J. Wilson. A consumer incentive warranty policy and servicing strategy for products with uncertain quality. *Quality and Reliability Engineering International*, 11 (3): 155-163, 1995.

[24] J. Lawless, J. Hu and J. Cao. Methods for estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Analysis*, 1 (3): 227-240, 1995.

[25] R. B. Nelsen. *An Introduction to Copulas*. New York, NY: Springer Series in Statistics, 2 edition, 2006.

[26] G. A. Fredricks and R. B. Nelsen. On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. *Journal of Statistical Planning and Inference*, 137 (7): 2143-2150, 2007.

[27] B. Seymour. MTTF, fail rate, reliability and life testing. *Burr-Brown Corporation*, 1993.

[28] E. Aarts and J. Korst. *Simulated Annealing and Boltzmann Machines*. New York, NY: John Wiley, 1990.

[29] W. Y. Yang, W. Cao, T. S. Chung and J. Morris. *Applied Numerical Methods Using MATLAB*. Hoboken, NJ: Wiley Interscience, 2005.

[30] J. W. Jawitz. Moments of truncated continuous univariate distributions. *Advances in Water Resources*, 27 (3): 269-281, 2004.

[31] D. R. Barr and E. T. Sherrill. Mean and variance of truncated normal distributions. *The American Statistician*, 53 (4), 357-361, 1999.

[32] U. J. Dixit and K. D. Phal. On estimating scale parameter of a truncated gamma distribution. *Soochow Journal of Mathematics*, 31 (4): 515-523, 2005

[33] T. Zhang and M. Xie. On the upper truncated Weibull distribution and its reliability implications. *Reliability Engineering and System Safety*, 96 (1): 194-200, 2011.

[34] K. Shahanaghi, R. Noorossana, S. G. Jalali - Naini and M. Heydari. Failure modeling and optimizing preventive maintenance strategy during two dimensional extended warranty contracts. *Engineering Failure Analysis*, 28: 90-102, 2013.

[35] J. Eliashberg, N. D. Singpurwalla and S. P. Wilson. Calculating the reserve for a time and usage indexed warranty. *Management Science*, 43 (7): 966-975, 1997.

[36] W. Nelson. *Applied Life Data Analysis*. New York, NY: Wiley, 1982.

[37] J. W. Tukey. *Exploratory Data Analysis*. Menlo Park, CA: Addison-Wesley Publishing Company, 1997.

[38] K. Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London*, 185 (A): 71-110, 1894.

[39] J. R. M. Hosking. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of Royal Statistical Society*, 52 (1): 105-124, 1990.

[40] M. A. Stephens. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69 (347): 730737, 1974.

[41] M. A. Stephens. Test of Fit for the Logistic Distribution Based on the Empirical Distribution Function. *Biometrika*, 66 (3): 591-595, 1979.

[42] R. B. D'Agostino and M. A. Stephens. *Goodness-of-Fit Techniques*. New York, NY: Marcel Dekker, Inc., 1986.

[43] D. J. Sheskin. *Parametric and Nonparametric Statistical Procedures*. Boca Raton, Fl: CRC Press, 1997.

[44] C. Dytham. *Choosing and Using Statistics, A Biologist's Guide*. Hoboken, NJ: Blackwell Publishing Company, 2 edition, 2003.

[45] F. W. Scholz and M. A. Stephens. K-sample Anderson-Darling tests. *Journal of the American Statistical Association*, 82 (399): 918-924, 1987.

[46] W. R. Blischke and D. N. P. Murthy. *Reliability: Modeling, Prediction and Optimization*. New York, NY: Wiley, 2000.

[47] *The USGA Handicap System*, United States Golf Association, New Jersey, 2002-2005.

[48] L. L. Kupper, L. B. Hearne, S. L. Martin and J. M. Griffin. Is the USGA golf handicap system equitable? *Newsletter of the Statistics in Sports Section of the American Statistical Association*, 14 (1): 30-35, 2001.

[49] F. Scheid. You're not getting enough strokes. *Golf Digest*, June (reprinted in *The Best of Golf Digest*, 1975, pp. 32-33), 1971.

[50] S. M. Pollock. A model of the USGA handicap system and fairness of medal and match play. In Ladany, S. P. and Machol, R. E. (eds.) *Optimal Strategies in Sports*, Amsterdam: North Holland, 141-150, 1977.

[51] G. M. Tallis. A stochastic model for team golf competitions with applications to handicapping. *Australian Journal of Statistics*, 36 (3): 257-269, 1994.

[52] P. D. Larkey. *"Statistics in Golf," in Statistics in Sport, ed. J. M. Bennett*. London: Arnold, 121-140, 1998.

[53] D. R. Bingham and T. B. Swartz. Equitable handicapping in golf. *The American Statistician*, 54 (3): 170-177, 2000.

[54] J. R. Lackrtiz. A new analysis of the golf handicap system: does the better player have an advantage? *Chance*, 24 (2): 24-34, 2011.

[55] B. C. Arnold, N. Balakrishnan and H. N. Nagaraja. *A First Course in Order Statistics*. New York, NY: John Wiley and Sons, Inc., 1992.

[56] N. Balakrishnan and A. C. Cohen. *Order Statistics and Inference: Estimation Methods*. Millbrae, CA: Academic Press, 1991.

[57] A. J. Cochran. On the normality and independence of golf scores, with applications in science and golf. *Proceedings of the First World Scientific Congress of Golf*, London: E. and F. N. Spon, 147-152, 1990.

[58] J. W. Tukey. *Data Analysis and Regression*. Menlo Park, CA: Addison-Wesley Publishing Company, 1997.

[59] S. Kotz and S. Nadarajah. *Extreme Value Distributions: Theory and Applications.* London: Imperial College Press, 2000.

[60] S. M. Ross. *A Course in Simulation.* New York, NY: Macmillan Publishing Company, 1990.

[61] F. Scheid. An evaluation of the handicap system of the USGA, in Ladany, S.P. and Machol, R.E. (eds.). *Optimal Strategies in Sports*, 151-155, 1977.

[62] F. Scheid. The search for the perfect handicap, in Cochran, A. J. and Farrally, M. R. (eds.). *Science and Golf II*, London: E. and F. N. Spon, 222-227, 1974.

[63] F. Scheid and L. Calvin. Adjusting golf handicaps for the difficulty of the course. *Proceedings of the Joint Statistical Meetings*, Orlando, 1995.