Indian Statistical Institute Series

Hemanta Saikia
Dibyojyoti Bhattacharjee
Diganta Mukherjee

Cricket Performance Management

Mathematical Formulation and Analytics

Springer

Indian Statistical Institute Series

The *Indian Statistical Institute Series* publishes high-quality content in the domain of mathematical sciences, bio-mathematics, financial mathematics, pure and applied mathematics, operations research, applied statistics and computer science and applications with primary focus on mathematics and statistics. Editorial board comprises of active researchers from major centres of Indian Statistical Institute. Launched at the 125th birth Anniversary of P.C. Mahalanobis, the series will publish textbooks, monographs, lecture notes and contributed volumes. Literature in this series will appeal to a wide audience of students, researchers, educators, and professionals across mathematics, statistics and computer science disciplines.

More information about this series at http://www.springer.com/series/15910

Hemanta Saikia · Dibyojyoti Bhattacharjee ·
Diganta Mukherjee

Cricket Performance Management

Mathematical Formulation and Analytics

Springer

Hemanta Saikia
Agriculture Statistics Discipline
Assam Agricultural University
Jorhat, Assam, India

Dibyojyoti Bhattacharjee
Department of Statistics
Assam University
Silchar, Assam, India

Diganta Mukherjee
Sampling and Official Statistics Unit
Indian Statistical Institute
Kolkata, West Bengal, India

ISSN 2523-3114 ISSN 2523-3122 (electronic)
Indian Statistical Institute Series
ISBN 978-981-15-1353-4 ISBN 978-981-15-1354-1 (eBook)
https://doi.org/10.1007/978-981-15-1354-1

Cover photo: Reprography & Photography Unit, Indian Statistical Institute, Kolkata, India

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Cricket is one of those few games in which an official statistician is involved from its very initial days. Though, in other sports like soccer, tennis, etc., several performance-related statistics are displayed these days for the sake of television viewers, but in cricket the performance-related statistics come naturally. It is evident that several statisticians and data scientists try different tools of modeling and pattern reorganization to solve various cricketing issues. This has been attempted successfully by several researchers around the globe. Of late, with the advances in computer technology, in terms of both hardware and software, application of statistical and data mining tools in cricket has reached its zenith.

Though statisticians and data scientists have applied different multivariate data analysis tools in solving different issues in cricket like rescheduling targets in rain-truncated matches, measuring team and individual performances, team selection, predicting match outcome, match scheduling and on several other decision-making aspects of cricket, yet there is no unified approach to compile all these works and place them in the form of an ordered text. Such an approach, on the one hand, shall document the huge literature on cricket analytics and, on the other hand, provide a comprehensive text explaining how complex data sets can be analyzed and how such large data sets can be utilized for decision making. Hence, this book is planned.

This book focuses on the application of statistical and data mining techniques in cricket. More specifically, we have compiled as well as worked on several quantitative features related to Twenty20 cricket, the latest and the most popular format of the game. Highlights of the work include performance quantification of cricketers (batsmen, bowlers, all-rounders, and wicket keepers), determining the market valuation of cricketers based on their on-field performances, effect of age on performance of the cricketers, etc.

The application of the techniques discussed in this book shall not only be limited to cricket data but an experimental scientist can also apply them following the lines of the discussion to a wide variety of areas like medical sciences, managerial decision making, marketing research, and so on.

Since the text provides a detailed overview of different aspects of the game, where quantitative techniques are gainfully applied for decision making, analysis of multivariate data related to cricket is to be done. Data analysis while interesting with one variable becomes realistic and challenging when several variables are considered together. Computer packages available these days can provide numerical results to complex statistical analyses. This book shall not confine its discussion to the analysis of cricket data only, but shall also provide the readers supporting knowledge necessary for making proper interpretations of the results available from computer packages and select appropriate techniques.

This book is expected to have a wide readership including postgraduate students of statistics/mathematics, data analysts, sports libraries, sports managing bodies, data scientists including research scholars in statistics, mathematics, operations research, and computer science. Most of the available texts address the issues of data mining/data analysis based on hypothetical data sets, the outcome of which does not have much sense for the readers. Cricket being a very popular topic in this part of the globe, with which most readers can associate themselves, the analysis shall become more meaningful to the reader. This shall attract readers toward the utility of the application of complex data analysis tools and how to take decisions on analysis of large data sets.

This book is divided into eight chapters. Chapter 1 discusses ‘cricket, statistics, and data mining’ and provides an extended overview of the issue-based significant works on cricket analytics. Cricket has suddenly been blessed by Twenty20 tournaments sponsored by franchisees where domestic as well as foreign players participate. This has given birth to a new era in cricket and a feast for cricket analysts. Franchisee-based cricket has led to the formation of cricket teams based on cricketer’s auction from a pool of domestic and international stars. This has opened a new area of decision making in cricket—like which player to bid for and for how much? This forms content of Chap. 2. Quantification of the performance of cricketers, based on their performance in different dexterities of the game namely batting, bowling, fielding, and wicket keeping, is approached by different authors. A discussion on such measures and then a simpler but effective linear model of performance computation is provided in Chap. 3. Chapter 4 provides a formula for measuring the fielding performance of cricketers—an approach that was never attempted by any other cricket analyst as quantifying the fielding performance always remained a challenge as it never generated any performance statistics directly. On combining the performance measures defined in the previous chapters, Chap. 5 describes ways to compute the market valuation of players. The market valuation of players when compared with their bid prices provided interesting insights. Chapter 6 quantifies how the performance of cricketers is associated with their age. Chapter 7 introduces an interesting concept of measuring the performance of cricketers considering the match situation in which the feat was achieved. The difficulty level, under which the team is performing, is quantified by introducing a measure called the *pressure index* in this chapter. The pressure index is an inclusive measure that can be utilized to measure batting performance both individually and between partners, bowling performance, determining the turning point of the match, and predicting the outcome of the match. The last and the concluding chapter tells the readers how optimization technique—the binary integer programming—can be utilized for selecting the optimum balanced cricket team from a host of cricketers of varied expertise.

Data sets used in this book are live data collected from international matches and from franchisee leagues. Several websites these days provide reliable data, and they are properly referred to in the text as and when data from such sites is used for calculations. These days, using technology, cricket coaches and analysts are able to collect tremendous amount of data with relative ease. Indeed, many of them have more data than they can handle. However, such data is usually meaningless until they are utilized for trends, patterns, relationships, and other useful information. This book brings about practical presentation of how statistical and data mining tools, from simple to complex, can be used to extract information from large data sets. The data mining and statistical techniques used in this book are not limited to aggregation, weighting, multiple regression, logistic regression, Bayesian data analysis, classification, fitting of probability distributions, application of several parametric and nonparametric tests, binary integer programming, binomial pricing option, etc. Description of output available from the application of statistical packages on cricket data and their interpretation are provided in detail. This will help the reader to understand the way of reporting the findings of statistical/data mining techniques for any data, not only those related to cricket. This increases the utility of this book beyond cricket analysts and makes it important to data scientists and statisticians involved in any type of data-oriented work. On reading this text, one will be able to understand the purpose of the technique applied and critically evaluate the application of the techniques to live and large data sets. We have tried to maintain an applied spirit throughout this book.

We welcome comments/suggestions on our presentation style, choice of topics, and any other aspect of this book that needs attention. We thank all the researchers who have contributed toward the field of sports analytics and have produced sufficient literature for us to read, get interested and take up research in this field. Special thanks are due to Prof. Hermanus H. Lemmer, Department of Statistics, University of Johannesburg, South Africa, who has inspired us through his work and advice. We also acknowledge the anonymous referees who have improved our research papers in this field by providing constructive suggestions. We would like to thank the reviewers of Springer Nature for their valuable advice. Finally, we would like to thank Shamim Ahmad, Senior Editor, Mathematics and Statistics, Springer Nature, New Delhi, for inspiring us to compile the text and provide necessary input in making this book more presentable. This book is possible only because of the support of Mr. Ahmad.

Jorhat, India Hemanta Saikia
Silchar, India Dibyojyoti Bhattacharjee
Kolkata, India Diganta Mukherjee

Contents

About the Authors

Hemanta Saikia is Assistant Professor at the Department of Agricultural Statistics, Assam Agricultural University, India. He previously taught statistics at the School of Business at Kaziranga University, Assam, India. He completed his Ph.D. at Assam University, Silchar, on the topic "Application of Bayesian Data Analysis and Neural Network for Evaluating Cricketers' Performance and Compensation Packages in Twenty20 Cricket". He is a life member of the Indian Bayesian Society, Indian Association for Social Sciences, and Health and North Eastern Economic Association. He has published over 21 research articles in various respected national and international journals.

Dibyojyoti Bhattacharjee is Professor and Head of the Department of Statistics, and Director of Internal Quality Assurance Cell, Assam University, India. He was an advisor to SPSS South East Asia for North East India and was also a visiting scholar at the Sampling and Official Statistics Unit, Indian Statistical Institute, Kolkata, India, and the University of Maryland Baltimore County, USA. He has worked with institutes like the Central Statistical Organisation, G. C. College, Gauhati University, and Assam University, India. He completed his Ph.D. in Statistical Graphics at Gauhati University, India. His research interests include data mining in sports (specifically cricket analytics) and statistical graphics. Professor Bhattacharjee completed his postdoctoral research at XLRI, Xavier School of Management, Jamshedpur, India. He has authored 15 books and 70 research papers in reputed national/international journals.

Diganta Mukherjee is Associate Professor at the Sampling and Official Statistics Unit, Indian Statistical Institute, Kolkata, India. He also serves as a technical advisor to Multi Commodity Exchange, the Reserve Bank of India, the Securities and Exchange Board of India, National Sample Survey Office, National Accounts Division, United Nations Development Programme, and Ministry of Electronics and Information Technology. He was previously associated with Jawaharlal Nehru University, Essex University, and ICFAI Business School, India. He completed his Ph.D. at the Indian Statistical Institute, Kolkata. He has published three books and over 80 research articles in various respected international journals. He is a life member of the Indian Econometric Society and Indian Statistical Institute, Kolkata. His research interests are welfare and development economics, particularly in multidimensional poverty.

Chapter 1
Cricket, Statistics, and Data Mining

1.1 Overview of Cricket Analytics

Cricket is a relatively leisurely outdoor game played in a circular ground, where the interaction between the bat and the ball takes place in a 22-yard hard surface called the pitch. It is believed to have its origin in the thirteenth century, yet no authentic information is available about when or where did cricket start. The first-ever reference to the game is found in a court case documentation in 1598 over an ownership dispute of a plot of land in Guildford, Surrey (Shah, 2012). In England, by the end of the sixteenth century, several ball games were played and became popular like cricket, stoolball, baseball, stickball, trap ball, and others. Some of these games were region-specific. Cricket, for example, originated in the woodland area of counties of Kent, Surrey, and Sussex, to the southeast of London (Underdown, 2000). At the end of the seventeenth century, cricket became a substantial gambling sport in England as evidenced by newspaper reporting of a game that was played in Sussex, England, in 1697 (Leach, 2018). The first inter-country match was played between the Londoners (Middlesex) versus the Kentish in 1719, but at that time there were no limits concerning the size and shape of the bat and ball. Further, its publicity was spread in the southern countries of England, but ultimately it was London where it gained immense popularity among the wealthy patrons. Although the first written laws of cricket were established in 1744, yet certain rules and regulations were in existence even before that. The first international cricket match was played between the USA and Canada on September 24–26, 1844, at Bloomingdale Park in Manhattan. However, there are some disputes regarding the validity of the claim that the game was international, as players were mostly drawn from two clubs of the respective countries (Williamson, 2007). But, Canada came into existence as a country only in 1867; till that time, it was a part of England, so technically (if the match is approved as an international match) the first international match was between England and USA. Thereafter, the inaugural match of test cricket was played between England and Australia in 1877. In England, in the year 1890, the first country championship was recognized, and instantaneously, several national competitions were constituted

H. Saikia et al., *Cricket Performance Management*, Indian Statistical Institute Series,
https://doi.org/10.1007/978-981-15-1354-1_1

in other parts of the world such as Currie Cup in South Africa, Plunket Shield in New Zealand, and Ranji Trophy in India (Encyclopedia Britannica, 2009).

The first limited overs cricket match was played at Melbourne Cricket Ground in 1971, and it was a time-filler match after a test match had been deserted as rain interrupted the play. Gradually, limited overs cricket (i.e., one-day cricket) became an ideal spectacle for the audience. The Board of Control for Cricket in India (BCCI) conveniently made no effort to correct this misconception. Instead, the BCCI steadily kept its distance from test cricket and poured their efforts to promote and capitalize one-day cricket in India (Ray, 2007). The ultimate result was that one-day cricket in India expanded beyond its recognition. The Indian commercial firms were also increasingly inclined to invest colossal sums of money in it, and the honeycomb of the game came to nest in India (Ray, 2007). With the passage of time, numerous changes occurred in the nature of the game, and today, cricket has become a major team sport in terms of participants, spectators, and media interests.

Innovation is a vital aspect of the game's continued development to remain relevant and to attract new fans and supporters (Mani, 2009). One suitable example of this is the Twenty20 format of cricket, and probably, it is the most significant development in cricket in the twenty-first century. In 2001, Stuart Robertson, the marketing manager of the England and Wales Cricket Board, performed a market survey on the decline of cricket spectators in the field and proposed a new format with 20 over a side, and titled it Twenty20 (Daily Mail, 2008). This new format of Twenty20 cricket could be finished in three hours, therefore not requiring even the day-long promise by spectators. Twenty20 cricket was later introduced in 2003 in England through a domestic tournament named Twenty20 cup. This tournament had the punch line 'I don't like cricket, I love it.' It was a splendid achievement in England, and hordes of spectators relished and welcome the modern version of the game.

On June 13, 2005, the first international Twenty20 cricket was played between England and Australia at Rose Bowl in Hampshire. Recognizing the popularity of this format of cricket, the International Cricket Council (ICC) well-thought-out to held the first Twenty20 World Cup in South Africa in the year 2017. India won the competition leading to a farfetched lift to this format of cricket back home. Later on, Twenty20 game spread around the world creating a lovely spectacle of sport enjoyed by spectators at the ground and viewers on TV.

The game of cricket got a new dimension in April 2008 when BCCI initiated the Indian Premier League (IPL). It is a Twenty20 cricket tournament to be played among eight domestic teams. These teams were named after eight Indian states or cities but owned by the franchisee. The franchisee formed their teams by competitive bidding from a group of Indian and international players and the best of Indian imminent talent. The IPL was modeled in the line of the English Premier Football League (EPFL). In EPFL, international players enhance talent and charm to the local teams (Hayden, 2011). However, though Twenty20 is the latest format of cricket; IPL has emerged as a center of attention for the researchers of a number of disciplines, viz. economics, management, decision science, and human resource. This is primarily for the intention that players' remunerations are decided through an auction process. It has led to a rare opportunity of obtaining a valuation of the players in monetary terms.

The monetary valuation of the players determined through auction and accessibility of players' performance has allowed researchers to deduce on different aspects of the game (Saikia, Bhattacharjee, & Lemmer, 2012a, 2012b). For detailed knowledge of the history of cricket, one may read H. S. Altham's *History of Cricket* (1926), Rowland Bowen's *Cricket* (1970), and Derek Birley's *Social History of English Cricket* (1999).

The term data mining has often been used by statisticians, data analysts, and the management information system (MIS) communities. It is the application of some specific tools and techniques for extracting hidden information or patterns from the data (Fayyad, Shapiro, & Smyth, 1996). Recently, most of the sports organizations have realized that there is a huge wealth of knowledge hidden in their data and hence are employing statisticians and data analysts to retrieve the useful information for future prospects (Solieman, 2006). Data mining and statistical analysis in sports came to the notice of the mass with the book *Moneyball: The Art of Winning an Unfair Game* by Michael Lewis that was published in 2003. The non-fiction book tells of a manager of a baseball team-'Oakland Athletics' who with the help of a Yale University graduate applied analytical tools for a new method of team selection, which ultimately led the team to victory, despite Oakland's disadvantageous financial condition. The everyday vast amount of sports data are routinely collected about players, coaching decisions, and game events. Making sense of this data is important to those seeking an edge. By transforming this data into actionable knowledge, scouts, managers, and coaches can have a better idea of what to expect from opponents and be able to use a player draft more effectively (Schumaker, Solieman, & Chen, 2010a, 2010b).

Cricket is a team game, and in every cricket match, an enormous amount of statistics is generated in terms of player, team, game, and season. Therefore, the application of advanced data mining technique is yet to blossom to its full potential in this sport. Statistical analysis of the updated cricket data has become both necessary and beneficial for policy-formation and decision-making process. But unlike other sports such as basketball and baseball which are well researched from a sports analytics perspective, for cricket, these tasks are yet to be investigated in depth (Sankaranarayanan, Sattar, & Lakshmanan, 2014). Application of analytical methods in the field of cricket is termed as 'cricket analytics.' It empowers cricket teams to make the right strategy and accurate decision about a game that is well known for its unpredictability. Thus, this book focuses on how data mining techniques, viz. Bayesian, neural network, and regression, can be applied to sports (especially in limited overs cricket) in the form of performance measurement, pattern discovery as well as outcome prediction.

1.2 Different Dimensions of Quantitative Analysis of Cricket Data

The main focus of this book is the performance measurement of cricketers using statistical and data mining tools. The other objectives are derivatives of the performance measurement of cricketers. Many statisticians and data analysts have used different statistical tools and techniques to analyze cricket data. The other aspects of cricket where the significant application of statistical and data mining tools is used include team selection, target resetting in rain-interrupted limited over matches, the impact of technological advances, the impact of home advantage, and toss on the outcome of the match.

1.2.1 Performance Measurement

The first work exemplifying the quantitative analysis of cricket data is probably that of Elderton and Elderton (1909). Taking data sets from the performance of cricketers in test matches, the authors tried to explain several basic concepts of statistics. Their work recommends the use of standard deviation to quantify the performance of batsmen. In Elderton (1927), cricket data was utilized to fit the exponential distribution. Wood (1945, 1945a) advocated the application of the coefficient of variation to quantify the performance of cricketers. Later, Pollard (1977) and Clarke (1994) used the coefficient of variation to measure the consistency of batsman's performance and distributional patterns of runs scored taking the ball-by-ball data from test matches. Pollard, Benjamin, and Reep (1977) successfully applied negative binomial distribution to model the runs scored by different batting partnerships. Pollard (1977) tried to bring a comparison between the application of geometric and negative binomial distribution to model cricket scores. Another work that successfully analyzed the performance of batting partnership scores was that of Croucher (1979), who also applied the negative binomial distribution for the purpose. Schofield (1988) applied production function, used popularly in Economics, to English county cricket teams. In his study, it was indicated that, although batting and bowling are essential skills to decide the achievement of the teams, bowling seemed to be the more important factor between the two.

Further, several authors visited the area of quantitative analysis of cricket data through measuring the player's performance, run-scoring policies, batting strategies, etc., by applying dynamic programming in the late 80s. Clarke (1988) applied dynamic programming to examine the association between runs scored and wickets lost in ODI cricket. The study showed that a team in order to optimize their batting performance in ODI cricket shall score rapidly in the starting overs of their innings. A player rating framework was created by Johnston (1992) to rate the team as well as individual players in ODI cricket by utilizing dynamic programming model. Clarke and Norman (1999) examined the ideal batting techniques to be followed toward the later part of the innings in ODIs. Clarke and Norman (2003) contemplated the benefit of picking a ‘Night Watchman’ in test cricket utilizing dynamic programming model to accomplish the goal. Preston and Thomas (2000) have additionally connected this method with survival analysis to look at the batting strategies in ODI cricket.

The batting average is one of the prime measures to quantify batting performance. It is defined by the total runs scored by a batsman in n innings divided by the number of innings in which the batsman was dismissed. Wood (1945) redefined the batting average by considering the not-out scores of the batsman as completed innings. A further refined batting average was defined by Kimber and Hansford (1993) accommodating cases where a batsman remains not out in limited overs cricket. A similar effort, addressing the issue of not-out innings played by a batsman, was defined by Damodaran (2006) using an algebraic tool. Beaudoin and Swartz (2003) defined some new statistics to quantify the performance of batsmen and bowlers. The batting performance was based on runs scored, and the bowlers’ performance was based on runs allowed per match, respectively.

Borooah and Mangan (2010) devised a formula for calculating batting average by incorporating the thought of consistency into the calculation of averages and including the value of the player’s runs to the team’s total. An important component of the work concentrated on framing the optimal strategies. Most of the studies involving strategy building in cricket concentrate on building it either on batting techniques or on bowling techniques (Damodaran, 2006). Research associated with batting strategies has been carried out by authors like Clarke (1988), Clarke and Norman (1999), and Preston and Thomas (2000); and to bowling strategies, have been conducted by Rajadhyaksha and Arapostathis (2004).

To quantify bowling performance, Lemmer (2002) contributed to the combined bowling rate (CBR). The measure was a merger of the traditional bowling statistics, viz. bowling average, economy rate, and bowling strike rate. He used the harmonic mean to aggregate the three different measures for calculating CBR. Later, Lemmer (2004) also contributed another measure of batting performance (BP) specially designed for limited overs cricket. This measure was an exponentially weighted batting average that gave more weights to the most recent performances of the batsman in an exponential manner. The measure also used a standardized coefficient of consistency and a standardized strike rate of the batsmen. Later, several attempts were made by Lemmer (2005, 2006, 2007) to modify the proposed BP and CBR for a different format of the game of cricket by allocating some suitable weights but in an objective manner. Barr and Kantor (2004) proposed an alternative measure of batting performance by combining the batting average and strike rate of the batsmen through a multiplicative aggregation. Lewis (2005) defined a performance measure taking ball-by-ball data from cricket matches. This measure has the capacity to consider the match circumstance while quantifying the performance of players. Gerber and Sharp (2006) characterized different measures for batting, bowling, fielding and for quantifying the performance of all-rounders and wicket keepers. A parametric control chart was created by Bracewell and Ruggiero (2009) to screen individual batting exhibitions in cricket. In such a manner, Ducks ‘n’ Runs distribution was proposed using the concept of the mixture distribution. The beta distribution is utilized to demonstrate the zero batting scores (ducks), and a geometric distribution is utilized to portray the appropriation of non-zero batting scores (runs). A measure for wicket keepers in cricket is created by Lemmer (2011a) by combining the dismissal rate and batting average. The dismissal rate is characterized by the number of dismissals (measured in terms of the number of catches, stumpings, and run outs) divided by the number of matches played.

1.2.2 Team Selection Based on Players' Performance

Selecting an optimal and balanced cricket team prior to a given series or tournament is always a matter of concern for selectors as well as team management. In cricket, the selectors who have different expertise try to select a balanced cricket team based on a number of factors like a home or away ground, the experience of the players, pitch condition, etc. Sometimes selectors may subjectively decide to give a chance to a player in the team who is out of form. Many a time, the selectors also have to face criticism if the player could not perform well. Therefore, several authors have tried to develop a mechanism to select an optimum cricket team based on the on-field performance of the cricketers. A team selection method by the analytical hierarchical process was proposed by Kample, Rao, Kale, and Samant (2011) to choose a subgroup of players from an entire set of cricketers comprising of the batsman, bowler, all-rounder, and wicket keeper. A similar issue was also attended by Lemmer (2011b) through the integer programming approach to reach the solution, and Ahmed et al. (2013) used an evolutionary multi-objective optimization technique to choose the cricket team. The integer programming approach was also followed by Gerber and Sharp (2006) in order to select a squad of 15 players for limited overs cricket instead of a playing XI. The method was to select an ODI squad based on collecting the data from 32 South African cricketers. Taking forward a similar idea, an optimal Twenty20 South African cricket side was selected by Lourens (2009) based on performance statistics of a host of players who had participated in the domestic Pro20 cricket tournament. A fantasy league cricket team was selected by Bretteny (2010) under certain specified budgetary constraints using integer programming. This optimal team at every phase of the tournament measured the performance of cricketers up to the recent match played. Though most of the authors had used binary integer programming for optimum team selection, they had applied different techniques to measure the performance of cricketers. Some authors had used traditional statistics like batting average, strike rate, etc., for quantifying the performance of cricketers, while others had tried to combine such traditional measures into a refined measure to assess players' performance. Authors like Lourens (2009) and Bretteny (2010) have considered different refined measures proposed by the earlier authors in the process of selecting an optimal cricket team. Lemmer (2011b) focused on using the most appropriate measure for the selection of a cricket team after some matches had already been played. The need for applying optimization models was discussed by

Boon and Sierksma (2003) in selecting a balanced cricket team. Though the study was related to team selection in soccer and volleyball, it was a pertinent example of the application of transportation problem for this purpose. With special reference to cricket, Das (2014) applied a binary integer programming method to generate optimal sequences of teams in fantasy sports leagues. It has also taken into consideration the multiple aspects of team selection. However, none of the above-mentioned studies assumes the captain as a stagnant member of the team. In cricket, the captain of the team gets an obvious selection. But the expertise of the captain influences the selection of the other players of the team. Keeping this in mind, the problem of team selection with the captain being already included in the squad, Bhattacharjee and Saikia (2016) modeled the requirement of the team based on the expertise of the captain.

1.2.3 Home Advantage and Toss Effect

In international as well as domestic cricket, the match is played in a field/ground which generally belongs to one of the participating teams. Sometimes, especially in tournaments where more than two teams are involved some matches are played in grounds which are in no way related to any of the participating teams. Such grounds are termed as neutral grounds with reference to the participating teams. Other than neutral grounds, when a team plays in their home ground, they generally ask the curator of the cricket pitch to make some minor changes to suit the strength of their team or go against the weakness of their opponents. In addition to that, a team playing in their own ground gets the support of the home crowd and play under conditions with which they are more acquainted. Thus, in cricket like other team sports, a team is expected to have home advantage. Clarke (1988) felt that the study of home advantage in cricket was never seriously taken up by analysts in their studies. One such effort is that of Allsopp (2005), where home advantage in cricket was compared with other team sports. Some other researchers like Stefani and Clarke (1992) and Clarke and Norman (1995) have considered the effect of home ground advantage while modeling other cricketing issues. The application of binary logistic regression to quantify the home ground advantage was attempted by Crowe and Middledorp (1996). The study was framed in an Australian context considering the different visiting teams that went to play test cricket in Australia between 1977 and 1994.

Before a cricket match starts, the tossing of a coin takes place between the captains of the two competing teams. The captain who wins the toss takes a decision whether to bat or to field first. The captain of the other team has to follow the decision made by the captain who won the toss. The impact of winning the toss on match outcome was attempted by de Silva and Swartz's (1997). They applied a logistic regression model in international cricket matches. In their study, it was found that winning the toss does not provide any advantage on the match result compared to home advantage. Other works related to toss in ODI cricket by Clarke and Allsopp (2001) and Allsopp and Clarke (2004) also indicate no considerable impact on the match result. But, both these works indicated some advantage to the home team. However, taking data from domestic cricket in England (for the seasons 1996 and 1997) and on application of binary logistic regression found a significant positive effect of winning the toss on match outcome. A similar result based on data of the country cricket in England was also reached by Forrest and Dorsey's (2008). Bhaskar (2009) fitted a linear probability model to the outcome of the ODI matches based on matches played in home, neutral, and away venue. Nevertheless, there was no quantification of the advantage that the home teams get after winning the toss. Saikia and Bhattacharjee (2010) examined the effect of toss and home advantage in the context of international Twenty20 matches. Their study indicated that there was no remarkable impact of home ground on the outcome of Twenty20 cricket matches, but an impact of toss outcome was evident.

1.2.4 Target Resetting in Interrupted Limited Overs Cricket Matches

As mentioned before, cricket is being played in three formats, viz. 5-day test match (test cricket), 50-over one-day internationals (ODIs), and 20-over Twenty20 internationals. The ODI format was introduced internationally in the year 1971. In the early years of its introduction, ODIs were played with 40/45/55/60 overs in an Innings. Later, International Cricket Council (ICC) regularized 50 overs as the standard length of ODI innings. The first batting team (team A) sets a target for the second team (team B) with the help of ten wickets or in a maximum of 50 overs whichever ends first. In a maximum of 50 overs, Team B has to chase the target set by team A, with ten wickets in hand. However, unexpected situations like rain, bad light, floodlight failure, etc., sometimes enforce a shortened game. Thus, the length of an innings becomes reduced. In these shortened matches, since two teams have unequal batting and bowling resources, a revised target is set for the team B. In order to do so, the International Cricket Council (ICC) has adopted different methods of adjustment which are discussed in the following section.

Note that in what follows there are two prospects, either a team was chasing a revised target following a stoppage earlier in the match (in which case the target is calculated on the basis of the score needed by Team B in the revised number of overs) or else there is a terminal stoppage (when there is an interruption and no further play is possible). A team could already be chasing a revised target because of an earlier stoppage when a terminal interruption occurs necessitating a further recalculation.

Among all the methods of revised target resetting, the first and simplest method was the **Average Run Rate (ARR)**. It compares the runs per over of team A's complete innings against that of Team B for the number of overs they received.

The second method was the **Most Productive Over (MPO)** or **Highest Scoring Overs (HSO)** which compares the maximum runs scored by Team A in any set of overs (not necessarily consecutive) equal to the number of completed overs received by Team B against the score of Team B in those completed overs. So if Team B received 36.3 overs, their score after 36 overs is compared to the highest scoring 36 overs of team A's innings (so could be any 36 of the 50 overs).

ARR is a simple method and easy to compute, but it gives more emphasis only on the run rate and pays no attention to the match situation. In MPO and DMPO methods, the revised target is set by eliminating utmost economic overs of the first batting teams. Consequently, they often become more unfair to the first batting team by ignoring the bowling excellence of team B.

The next class of methods may be termed **Discounted Most Productive Over (DMPO)**; among these we have:

(i) **Adjusted highest scoring overs (AHSO)**

In this method, the runs scored in different overs by Team A are arranged in descending order. The runs scored in the first *n* number of overs after the said arrangement is then added together. The number *n* of overs is the complete overs played by Team B factored by an amount of 0.5 times of the overs they have lost, expressed in percentage. For example, Team B has played 32 overs when the rain started in a 50-over match. The runs scored by Team B in 32 overs are compared with the runs scored by Team A in their highest scoring 32 overs out of their innings of 50 overs factored down by 9% (which is 0.5 times the 18 overs left in the 50-over match).

(ii) **Highest scoring consecutive overs (HSCO)**

Team B while batting faced only *n* number of overs when the rain started. The maximum run scored by Team A in any *n* consecutive overs is compared with the runs scored by Team B in those *n* completed overs. For example, if Team B received 32.5 overs, then their score at the end of 32 overs is compared to the highest consecutive 32-over score of team A's innings (so that might be over 1–32 or 2–32, …, 18–50). If there are any further interruptions, then the target is calculated on the average runs per over derived from the first interruption.

Then there are various adjusted run rate methods:

(i) **Consecutive overs or run rate (CORR)**

In this strategy so as to win the match Team B through their total innings need to surpass the score accomplished by Team A in their last number of overs gotten by Team B and furthermore Team A's overall run rate. The last was connected with the goal that a team backing off after a quick beginning to their innings might be limited. Accordingly, if Team B were to get 30 overs in a 50-over match, they needed to surpass Team A's total runs for overs 20–50 and also team A's run rate for their 50 overs.

(ii) **Countback to score at the equivalent point of Team A (CB)**

Team B has played *n* number of overs after which the interruption took place. Now, the runs scored by Team A till the end of the n^{th} over is compared with that of Team B's runs till interruption.

(iii) **Countback to score at equivalent last completed over of Team A (CBC)**

The score of Team A is compared with the score of Team B at the end of the number of completed overs faced by Team B.

(iv) **Factored run rate (RRFAC)**

The total runs scored by Team A is increased by multiplying the actual run rate of Team A by a factor, which is the percentage of overs lost by Team B. The new target of Team A is then utilized to determine the required run rate of Team B. But in this method, the chance of a tie is almost nil; only possible if the calculation mentioned above results in an integer. For example, Team A scores 225 in 45 of their 50 overs. Then, the match was interrupted. This prohibited them to carry on until 40 overs remained. Team B's required run rate becomes 102.5% in this case, where each over lost is an increase of 0.5% (5 overs lost). So $225 \times 102.5/100 = 230.625$ or 231 was the target to win (and 230 is a loss).

(v) **Run rate wickets (RRW)**

In this method, a comparison is made between the run rate of team A's complete innings and the run rate of Team B for the number of overs they received. But in this case, the number of wickets that Team B can use is reduced.

(vi) **Runs per wicket on countback to score at equivalent point (RWCB)**

Comparison is made between team A's runs per wicket figure at the end of the exact number of overs received by team B, with the same statistic as that of team B.

In addition to all these, there is the **Parabola (or norm-normal performance) method (PARAB)**. Here, a parabolic curve is fitted with the score of Team A against the number of overs bowled. The curve is then utilized to estimate the target score for Team B. This purpose of the curve is to revise the target for Team B when the score of Team A is provided. This method was used in the ICC World Cup of 1996 played jointly in India, Sri Lanka, and Pakistan for computing revised target in interrupted matches.

Some other tournament-specific methods used for calculating revised targets are as follows:

East League Calculator Method
For a terminal interruption, there are tables which give a revised par score based on the resources left when the match had to be ended.

Lancashire League Adjusted Target
To calculate the adjusted target for the team batting second, i.e., Team B, the calculation is based on the run rate at which the team batting first, viz. Team A has scored. In case all the batsmen of Team A were dismissed before completing their full quota of allocated overs, then 75% of their run rate is considered and is multiplied with the number of overs lost. Thus, the target for Team B is run rate of Team A multiplied

by the overs available to Team B minus the runs adjustment (as explained in the previous line) plus one.

Some norms' table was designed for adjusting totals in relation to overs lost called as PARAB and WC96. But both the methods did not consider the number of wickets lost, and hence were not of much practical significance.

Apart from these, there are certain suggestions in scholarly literature like as follows:

VJD Method

Based on the number of remaining wickets at the point of stoppage and the number of overs for which the fielding restriction exist, Jayadevan (2002) proposed a method for computing a revised target in interrupted ODI matches. Depending on the resources left to the batting team at the time of interruption, the table developed by Jayadevan helps to calculate the revised target. Similarly, Preston and Thomas (2002) designed a method based on the probability of winning a match by a team before the stoppage took place. But its probability conservation approach often leads to impractical results.

Clark Curves Method

Clark (1988) applied dynamic programming to develop the Clark curves that can be utilized to solve the issue of rescheduling of targets. The curves are computed using the ratio of the final score to wickets remaining.

In this method, every innings is deemed to have three stoppages, and at every stoppage, the resources available with the team are different. The revised target is computed based on the resources available to the team. But, there is some indistinct way in computing the revised scores at the meeting point of two adjacent stoppages.

However, the most accepted method and the one currently in practice for target rescheduling at the international level is the **Duckworth/Lewis Method (D/L,** 1998). The method is currently called the **Duckworth/Lewis/Stern Method (D/L/S Method)**.

For interruption at any point of a limited-overs match, the Duckworth/Lewis tables are available. The table is about the amount of resource left with the team which is a function of the number of overs consumed and the number of wickets lost by the team at that instant when the interruption took place. But the tables were revised several times over the years based on several peculiar outcomes to which the method landed at different match situations.

Studies like Mchale and Asif (2013) compared several available target rescheduling rules and found the D/L more practical than any other method. Probably, this is the reason for which the method gained popularity for its robust applicability and fairness. The amount of resources available to a team at any instant of the match as per the D/L method is computed based on the number of overs remaining and the wickets that are yet to be dismissed. However, it does not consider the individual skill of the batsmen who are yet to be dismissed. Since individual skill of players varies, the number of wickets remaining shall be weighted using the individual players who are to be dismissed and not by counting the number of batsmen yet to be dismissed. This was the argument of Singh and Adhikari (2015) who recommended an alternative method to calculate the revised target in interrupted limited overs cricket matches. They further included the available bowling resources of the fielding team, which was never considered in the D/L method.

1.2.5 Technological Advances in Cricket

Cricket is a game of tactics. This is a game full of strategies and counter-strategies. But the game has also seen the application of modern technology. In case of different decisions where human limitations might influence the outcome of the game, the governing bodies have allowed the on-field umpires to take the help of technology to make decisions. There are several other situations where we shall see more applications of technology in the near future. For example, a 'no ball' call of the on-field umpire due to over-stepping by the bowler could be easily decided by the application of technology. This is something which might well be seen in future. A significant amount of literature has addressed the technological advances that are already applied in cricket, their advantages and scope of improvement. Some works have proposed the application of technology in improving the efficiency of the game. The work of Jain and Chakrawarti (2014) on the improvement necessary in the Hot Spot technology used for identifying the exact place (bat or pad or glove etc.) where the cricket ball has made its contact is one such example.

Some works have investigated the application of technology to understand how cricket equipment and gears have influenced the performance of professional cricketers. One such instance was the work of Morris (1976). He tested the variation in the performance of fielders in taking catches when the color of the ball and color of the background changed. Barton (1982) examined the character and degree of the swing of a cricket ball. A similar study considering the aerodynamic phenomena that influence the swing of a cricket ball was performed by Mehta (2005). Changing cricket bats with respect to the design and material used, remaining in the purview of the cricketing rules, John and Li (2002) studied their performances. In the study, it was detected that with less ball–bat contact time, the deflection in the bat diminishes and so more is the impact of the hit made on the ball by the bat. Slip-catching was the concern in the work of Koslow (1985) and Kingsbury, Scott, Bennett, Davids, and Langley (2000). Their study was performed on varying the luminance level and color of the ball. Umpiring decisions can be gauged applying the technological advances to match video. One such study is that of Chedzoy (1997). McAuliffe and Gibbs (1997) tried to apply appropriate technology to understand the nature of the cricket pitch. He attempted to predict the two important attributes of a cricket pitch, viz. pace and bounce by the application of technology. To understand how the eye movement of a batsman influences the shots taken by him, Land and McLeod (2000) performed a study which leads to the conclusion that the eye movement policy of the batsman contributes to his skill in the game. How the visual skills of a batsman influence the batting performance of cricketers was examined by Balashaheb, Mamman, and

Sandhu (2008). The surface hardness of the bat is supposed to influence the strength of a shot taken by a batsman using the bat; this was the topic of the work of Sayers, Koumbarakis, and Sobey (2005). Kolekar and Sengupta (2004) applied hidden Markov model approach to structure the analysis of cricket video sequences.

1.2.6 Players' Valuation in Cricket

Cricket is a game that generates huge numerical information and so there are several studies that concentrate on players' performance as well as team performance. But studies related to players' valuation were never carried out as there was no formal data available for players prior to the beginning of franchise-based Twenty20 cricket tournament like the Indian Premier League. However, for measuring the valuation of players, a need was felt to define some tournament-specific measures of performance. Many authors have addressed this need of the hour. A rundown of pointers of accomplishment in IPL was distinguished by Petersen, Pyne, Portus, Dawson (2008) investigating the performance of the different teams. Taking data from the inaugural Twenty20 World Cup, Lemmer (2008) examined the performance of batsmen and bowlers. Regarding the all-around performance of the players in the first season of IPL, van Staden (2009) authored the term ideal all-rounder, batting all-rounder and bowling all-rounder to recognize the exhibition of all-rounders. Taking that point forward, measure for cricketers' performance in Twenty20 cricket was created by him utilizing information from the inaugural season of IPL. A few cross-sectional models were compared by Lenten, Geerling, and Konya (2012) to consider the components that contribute toward the performance of cricketers in different formats of the game. Consolidating the batting, bowling, and fielding capacity of cricketers, Douglas and Tam (2010) analyzed the team performance of ICC Twenty20 World Cup played in 2009. The study led to the findings that for success in Twenty20 cricket, the team should pay emphasis on taking wickets and bowling dot balls while fielding and increase the number of 50+ partnerships and the number of hits to the boundary while batting. Saikia and Bhattacharjee (2011) applied stepwise multinomial logistic regression model to find out the significant factors that influence the performance of the all-rounders in IPL. A method was proposed by Bharathan, Sundarraj, Abhijeet, and Ramakrishnan (2015) that can be applied to evaluate the performance of players leading to team selection, to rate players, and to determine their salary caps. Some literature is identified which related to the valuation of cricketers in IPL and variables in charge of the valuation is additionally accessible. Parker, Burns, Natarajan, (2008) studied players' valuation of the inaugural season of IPL and found that the variation in the salary of the players in IPL can be explained by the experience of the players, batting and bowling strike rate, being an all-rounder and having an icon player status. Rastogi and Deodhar (2009) took up the same issue. In their study, they pointed out that recent batting and bowling performances are the significant features of influencing players' salaries. Considering performance of cricketers as a real option, Suleman and Saeed (2009) developed a performance index (SS Index) for cricketers applying the binomial option-pricing model and determined the present value of the players. The final bid prices of the players during first three seasons of IPL were examined by Depken and Rajasekhar (2010). As reported by them, the cricketers' salaries are influenced by player distinctiveness and that the marginal values have remained same during the first three seasons of IPL. Swartz (2011) expressed his view against the existing auction mechanism of IPL and suggested that it should be replaced by a system of player salaries. Singh, Gupta, and Gupta (2011) made a critical analysis of the IPL auction process and discussed the problems faced by bidders in a dynamic bidding process. Another work in the same lines as that of Lenten et al. (2012). They also proposed six alternative models to modify the bidding process of IPL. Saikia, Bhattacharjee, and Bhattacharjee (2013) based on performance-related statistics from the first three seasons of IPL identified cricketers who could and who could not justify their salary through their performance.

1.2.7 Umpiring Errors and Umpire Scheduling

An umpire in a cricket ground is responsible to make decisions about the outcome of the incidents that are taking place while the game is in progress. He is not a member of any one of the teams playing the match. The basic objective of the umpire is to ensure that the game is played as per the rules of the game, and the spirit of the game and sportsmanship is maintained. The umpire counts the number of deliveries and declares the end of an over, checks if each delivery is legally bowled, and takes decisions on appeals made by the fielding team against the batsman for a dismissal. In cricket, irrespective of the format of the game, there are two umpires who are present on the field. One stands near the stumps from where the bowler delivers (popularly called the bowling end or the non-striker's end) basically to watch the legitimacy of the ball delivered and the action followed by the batsman corresponding to the delivered ball. Most of the appeals for dismissal of the batsman are referred to him. The other umpire stands square of the batting end in the right side from where the bowler is delivering. This position in a cricket field is called the square leg, and accordingly, the umpire standing in that position is called the square leg umpire or simply leg umpire. The role of the leg umpire is to assist the main umpire in whatever way he can and give a decision on run out (at the batsman's end) and stumping, decide if the ball is of unplayable height, etc. At the end of every over, the umpires change their positions and their responsibilities. In addition to these two umpires, there is another umpire who is an off-field umpire. He is popularly called 'the third umpire.' This umpire becomes active whenever the on-field umpires are not sure about a decision. The third umpire takes the help of television replays and other technical devices to reach a decision referred to him by the on-field umpires. It is obvious that in a technical game like cricket, the umpires have an enormous role to play and so they must be well versed with all the cricket rules in different formats of the game and also the change in existing rules. In addition to their knowledge, they shall have good eyesight, able to take correct decisions, pleasant personality, and good communication skill. History has seen several incidents where wrong decisions from the umpires changed the course of the game in favor of one side and against the other. In some situations, a single wrong decision can change the outcome of the game (Manage, Mallawaarachchi, and Wijekularathna, 2010). With the availability of slow-motion television replays and other related technology, even the spectators are now informed about errors in umpiring decisions. In the words of Chedzoy (1997), ‘In cricket, where the result of decisions (by the umpire) can have such a significant effect on batsmen, on the progress or on the results of matches, the problem is likely to be more important than in any other competitive game.’ Thus, the quality of umpiring decision and the need for regular ranking of umpires based on their performance is definitely the need of the hour. Manage et al. (2010) used receiver operating characteristic (ROC) curves generally used to measure the accuracy of predictions by different methods to analyze the quality of decisions made by umpires. The authors showed with simulated data how the technique can be utilized to rank and rate umpires in cricket based on their ability to take correct decisions. They also suggested the application of the method for choosing the elite panel of umpires by cricket’s global governing body the International Cricket Council (ICC). Chedzoy (1997) analyzed the binary problem which an umpire faces. Each appeal for dismissal is referred to an umpire for the decision. Now, the umpire has to approve the appeal or reject it. Rejecting an appeal which was actually a dismissal (can be confirmed from TV replays) and accepting an appeal which was not actually out lead to a wrong decision by the umpire. Using this concept, Chedzoy (1997) simulated the impact of such wrong decisions on batsman, teams, and on the ultimate result of the game. His study was based on elaborate data taken from the 1975 test series between England and Australia.

Leg before wicket (LBW) is a way of dismissing a batsman in cricket that invites a lot of subjectivity to the game. Here the umpire has to make a binary decision of out/not out about the batsman, following an appeal from the fielding side, when the fielders feel that the legs came on the way of the ball approaching the stumps without hitting the bat. Several aspects of the ball need to be considered by the umpires like the flight of the ball, height, direction, position of the stumps, etc., before responding to the appeal. Considering all these issues, if the umpire is convinced that the ball has not hit the bat and would have hit the stumps if the leg of the batsman was not there, then he shall respond in affirmative to the appeal and declare the dismissal of the batsman. Because of the subjectivity involved, such decisions of the umpire are often encircled by controversy. Several authors analyzed the issue by applying quantitative techniques. Sumner and Mobley (1981) considered all test matches played between 1877 and 1980 and showed using a two-sample binomial test that away batsmen were out LBW more often in Australia, India, and Pakistan. Crowe and Middledorp (1996) used generalized linear models to look at LBW rates for test matches played in Australia in 1977–1994 classified by the visiting teams and found that batsmen of England, South Africa, and Sri Lanka are more vulnerable to LBW. Ringrose (2006) took test cricket data from 1978 to 2004 and then through the application of generalized linear models and mixed models attempted the impact of location, team, and the presence of neutral umpires on LBW dismissals. One significant finding of the study was that there was no significant impact on the presence of neutral umpires on LBW dismissals. This was against the popular belief that home umpires are biased toward home teams.

Ranking, rating, and scheduling umpires in a tournament can be an appealing area of research for those interested in cricket analytics as this is a lesser-explored research dimension. With the inclusion of Decision Review System (DRS) is where both batting and fielding teams get an opportunity to challenge the decision of the on-field umpire; the precisions of umpiring decisions in close situations are getting well documented. Several types of research shall come out from the availability of this input in the near future. Scheduling umpires in a tournament—in a way that the tournament can be run with a minimum number of umpires, evenly distributing the load of umpires and minimum travel of umpires between match venues can be an exciting problem for students interested in operations research.

1.3 Organization of the Book

This chapter concentrates on the overview of cricket and presents a brief literature review based on earlier studies. Chapter 2 discusses the evolution of the franchise cricket and the different mathematical models proposed for player selection in competitive bidding and the expected bid price of cricketers. Chapter 3, the existing performance measures of cricketers are discussed and the hybrid measures developed by cricket statisticians are reviewed. Some new performance measures based on the traditional performance statistics developed by the authors are also presented. Chapter 4 is devoted to the development of a performance measure for fielding in cricket, which is considered to be a challenging area in cricket, and no earlier attempts were successful. Chapter 5 of the book evaluates the present market valuation of the cricketers based on their on-field performances using binomial option-pricing model and then identifies the cricketers who were able to justify their salaries by their on-field performance in Indian Premier League (IPL). Chapter 6 deals with the most challenging part in sports, which is to examine the impact of age on the performance of the cricketers using a regression model with a random regression coefficient and Gini Index. Chapter 7, a pressure index is defined which quantifies the match situation under which a player performs in limited overs cricket. Accordingly, quantification of the performance of cricketers based on the pressure index is attempted. Chapter 8 of the book deals with the optimum or balanced national cricket team selection before the tournament based on the on-field performance of the cricketers in domestic tournaments. The binary integer programming method is used to select the balanced national cricket team for some national sides.

1.4 Conclusion

Huge data sets are generated every day from the world of sports. Most of these data are player-specific but for the team, sports data is generated both in terms of teams as well as for individual players and even season-specific in some cases. Many data scientists these days use different statistical and data mining tools to analyze sports-related data which leads to vital inputs for the players. Yet, the techniques of data mining are to find its proper place in the world of sports.

Cricket is a team game that generates data with every tick of the clock. The performances of every individual based on a combination of skills (i.e., batting, bowling, fielding) and strategies and counter-strategies distinguish cricket from other team sports. This book focuses on the application of data mining techniques in cricket. The entire text is an example of how data mining can be helpful for decision making in sports with special reference to cricket. More specifically, the quantitative features related to limited overs cricket, viz. one-day internationals and Twenty20, are the highlights of the book. The work includes performance quantification of cricketers (batsmen, bowlers, all-rounders, and wicket keepers), determining the market valuation of cricketers based on their on-field performances, the effect of age on the performance of the cricketers, etc. It also provides a detailed overview of the different aspects of the game where quantitative techniques are gainfully applied.

The attention of this book is to explore the different issues of cricket where data mining and statistical tools have a role to play. The issues of combining different abilities of cricketers, like consistency in batting, ability to score runs at a faster rate, wicket keeping, the capacity of providing breakthrough while bowling or economic bowling, etc., are dealt in details in this text. A recent concept of measuring the performance of cricketers given the match situation in limited overs cricket is improvised here. The book in one hand explains the application of different statistical, mathematical, and data mining tools in cricket, and on the other hand, it extracts the facts hidden behind cricket data. Different data mining and statistical tools are used here to evaluate, classify, and predict the cricketers' performance and their compensation packages in Twenty20 cricket. Apart from the age effect on cricketers' performance, team selection through optimization model is also discussed. Thus, the benefit of the book can be enjoyed by both the academic fraternity and the decision-makers in sports.

References

Ahmed F., Deb K, and Jindal, A. (2013) Multi-objective optimization and decision making approaches to cricket team selection. Applied Soft Computing, *13* (1), 402-414

Albert, J., & Koning, R. H. (2008). *Statistical thinking in sports*. London: Taylor & Francis Group; Chapman and Hall/CRC Press.

Albert, J., Bennett, J., & Cochran, J. J. (2005). *Anthology of statistics in sports*. Philadelphia: American Statistical Association and the Society for Industrial and Applied Mathematics.

Allsopp, P. (2005). Measuring team performance and modeling the home advantage effect in cricket (Ph. D. thesis). Swinburne University of Technology, Melbourne, Australia.

Allsopp, P., & Clarke, S. R. (2004). Rating teams and analyzing outcomes in one-day and test cricket. *Journal of Royal Statistical Society, 167,* 657–667.

Allssop, P., & Clark, S. R. (2001). Fair measures of performance: The World Cup of cricket. *Journal of the Operational Research Society, 52,* 471–479.

Balashaheb, T., Mamman, P., & Sandhu, J. S. (2008). The impact of visual skills training program on batting performance in cricketers. *Serbian Journal of Sports Sciences, 2,* 17–23.

Barr, G. D. I., & Kantor, B. S. (2004). A criterion for comparing and selecting batsmen in limited overs cricket. *Journal of the Operational Research Society, 55*(12), 1266–1274.

Barton, N. G. (1982). On the swing of a cricket ball in flight. In *Mathematical and physical sciences, Proceedings of the Royal Society of London (Series A),* London (pp. 109–131).

Beaudoin, D., & Swartz, T. (2003). The best batsmen and bowlers in one-day cricket. *South African Statistical Journal, 37,* 203–222.

Bharathan, S., Sundarraj, R. P., Abhijeet, & Ramakrishnan, S. (2015). A self-adapting intelligent optimized analytical model for team selection using player performance utility in cricket. In 9th annual MIT Sloan Sports Analytics Conference, Boston Convention and Exhibition Centre, Boston.

Bhaskar, V. (2009). Rational adversaries? Evidence from randomized trials in ODI cricket. *Economic Journal, 119*(534), 1–23.

Bhattacharjee D., Saikia, H. (2016) An objective approach of balanced cricket team selection using binary integer programming method. OPSEARCH *53* (2), 225-247

Boon, B. H., & Sierksma, G. (2003). Team formation: Matching quality supply and quality demand. *European Journal of Operation Research, 148,* 277–292.

Borooah, V. K., & Mangan, J. E. (2010). The "Bradman Class": An Exploration of some issues in the evaluation of batsmen for test matches, 1877-2006. *Journal of Quantitative Analysis in Sports, 6*(3). https://doi.org/10.2202/1559-0410.1201.

Bracewell, P. J., & Ruggiero, K. (2009). A parametric control chart for monitoring individual batting performance in cricket. *Journal of Quantitative Analysis in Sports, 5*(3), 1–19.

Bretteny, W. (2010). *Integer optimization for the selection of a fantasy league cricket team* (M.Sc. Dissertation), Nelson Mandela Metropolitan University, South Africa.

Chedzoy, O. B. (1997). The effect of umpiring errors in cricket. *The Statistician, 46*(4), 529–540.

Clarke, S. R. (1988). Dynamic programming in one-day cricket- optimal scoring rates. *Journal of the Operational Research Society, 50,* 536–545.

Clarke, S. R. (1994). Variability of scores and consistency in sport. In N. de Mestre (Ed.), *Proceedings of the Second Australian Conference on Mathematics and Computers in Sport* (pp. 69–81). Australia.

Clarke, S. R., & Allsopp, P. (2001). Fair measures of performance: The World Cup of cricket. *Journal of Operational Research Society, 52,* 471–479.

Clarke, S. R., & Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. *Statistician, 44,* 509–521.

Clarke, S. R., & Norman, J. M. (1999). To run or not? Some dynamic programming models in cricket. *Journal of the Operational Research Society, 50*(5), 536–545.

Clarke, S. R., & Norman, J. M. (2003). Dynamic programming in cricket: Choosing a night watchman. *Journal of the Operational Research Society, 54,* 838–845.

Croucher, J. S. (1979). The battle for the Ashes a statistical analysis of post war Australia–England cricket. In *Proceedings of the 4th National Conference of the Australian Society of Operations Research* (pp. 48–58), Sydney: Sydney University of Technology.

Crowe, S. M., & Middledorp, J. (1996). A comparison of leg before wicket rates between Australians and their visiting teams for test cricket series played in Australia, 1977–94. *The Statistician, 45*, 255–262.

Daily Mail. (2008). *Meet the man who invented Twenty20 cricket—The man missing out on millions*. www.dailymail.co.uk/sport/cricket/article-1025831/.

Damodaran, U. (2006). Stochastic dominance and analysis of ODI batting performance: The Indian cricket team, 1989–2005. *Journal of Sports Science and Medicine, 5,* 503–508.

Das, D. (2014). *Moneyballer: An integer optimization framework for fantasy cricket league selection and substitution*. http://debarghyadas.com/files/IPLpaper.pdf. Retrieved July 11, 2015.

de Silva, B. M., & Swartz, T. B. (1997). Winning the coin toss and the home team advantage in one-day international cricket matches. *New Zealand Journal of Statistics, 32,* 16–22.

Depken, C. A., & Rajasekhar, R. (2010, April). Open market valuation of player performance in cricket: Evidence from Indian Premier League. Retrieved from www.ssrn.com/abstract=1593196.

Douglas, J. M., & Tam, N. (2010). Analysis of team performances at the ICC World Twenty20 Cup 2009. *International Journal of Performance Analysis in Sport, 10*(1), 47–53.

Duckworth, F. C., & Lewis, A. J. (1998). A fair method of resetting the target in interrupted one-day cricket matches. *Journal of Operational Research Society, 49,* 220–227.

Elderton, W. P. (1927). *Frequency curves and correlation*. London: Layton.

Elderton, W. P., & Elderton, E. M. (1909). *Primer of statistics*. London: Black.

Fayyad, U., Shapiro, G. P., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *American Association for Artificial Intelligence* 37–54.

Forrest, D., & Dorsey, R. (2008). Effect of toss and weather on county cricket championship outcomes. *Journal of Sports Sciences, 26,* 3–13.

Gerber, H., & Sharp, G. D. (2006). Selecting a limited overs cricket squad using an integer programming model. *South African Journal for Research in Sport, Physical Education and Recreation, 28*(2), 81–90.

Hayden, M. (2011). *Standing my ground* (p. 346). India: Harper Collins Publishers.

Iyenger, N. S., & Sudarshan, P. (1982, December). A method of classifying regions from multivariate data. *Economic and Political Weekly* 2048–2052.

Jain, P. K., & Chakrawati, R. (2014). Improvement in hot spot technology in cricket technique for the best. *International Journal of Scientific and Engineering Research, 5*(10), 257–261.

Jayadevan, V. (2002). A new method for the computation of targets scores in interrupted limited over cricket matches. *Current Science, 83,* 577–586.

John, S., & Li, Z. B. (2002). Multi-directional vibration analysis of cricket bats. In S. Ujihashi & S. J. Haake (Eds.), *The engineering of sport 4* (pp. 96–103). Oxford: Blackwell Science.

Johnston, M. I. (1992). *Assessing team and player performance in one-day cricket using dynamic programming* (M.Sc. Thesis). Swinburne University, Australia.

Kample, A., Rao, R., Kale, A., & Samant, S. (2011). Selection of cricket players using analytical hierarchy process. *International Journal of Sports Science and Engineering, 5*(4), 207–212.

Kimber, A. C., & Hansford, A. R. (1993). A statistical analysis of batting in cricket. *Journal of Royal Statistical Society, 156*(3), 443–455.

Kingsbury, D., Scott, K., Bennett, S., Davids, K., & Langley, M. (2000). Effects of cricket ball colour and illuminance levels on catching behaviour in professional cricketers. *Ergonomics, 43*(10), 1681–1688.

Kolekar, M. H., & Sengupta, S. (2004). Hidden Markov model based structuring of cricket video sequences using motion and color features. In *Proceedings of Indian Conference on Computer Vision*, *Graphics, and Image Processing* (pp. 632–637), India.

Koslow, R. E. (1985). Peripheral reaction time and depth perception as related to ball colour. *Journal of Human Movement Studies, 11,* 125–143.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience, 3*(12), 1340–1345.

Leach, J. (2018) *Classification of cricket matches from 1697 to 1825*. https://www.jl.sl.btinterenet.co.uk/stampsite/cricket/histories/matches.html. Retrieved on July 16, 2018.

Lemmer, H. (2004). A measure for the batting performance of cricket players. *South African Journal for Research in Sport, Physical Education and Recreation, 26*(1), 55–64.

Lemmer, H. (2005). A method for the comparison of the bowling performances of bowlers in a match or series of matches. *South African Journal for Research in Sport, Physical Education and Recreation, 27*(1), 91–103.

Lemmer, H. (2006). A measure of the current bowling performance in cricket. *South African Journal for Research in Sport, Physical Education and Recreation, 28*(2), 91–103.

Lemmer, H. (2007). The allocation of weights in the calculation of batting and bowling performance measures. *South African Journal for Research in Sport, Physical Education and Recreation, 29*(2), 75–85.

Lemmer, H. H. (2002). The combined bowling rate as a measure of bowling performance in cricket. *South African Journal for Research in Sport, Physical Education and Recreation, 24*(2), 37–44.

Lemmer, H. H. (2008). An analysis of players performances in the first cricket Twenty20 World Cup series. *South African Journal for Research in Sports: Physical Education and Recreation, 30*(2), 71–77.

Lemmer, H. H. (2011a). Performance measure for wicket keepers in cricket. *South African Journal for Research in Sport, Physical Education and Recreation, 33*(3), 89–102.

Lemmer, H. H. (2011b). Team selection after a short cricket series. *European Journal of Sport Science*. https://doi.org/10.1080/17461391.2011.587895.

Lenten, L. J. A., Geerling, W., & Konya, L. (2012). *A Hedonic model of player wage determination from the Indian Premier League auction.* Retrieved 15 October, 2010, from https://www.econ.canterbury.ac.nz/research/pdf/Paper_Lenten.pdf.

Lewis, A. J. (2005). Towards fairer measures of player performance in one-day cricket. *Journal of the Operational Research Society, 56,* 804–815.

Lourens, M. (2009). *Integer optimization for the selection of a Twenty20 cricket team* (M.Sc. Dissertation). Nelson Mandela Metropolitan University, Mathematical Statistics, Faculty of Science, South Africa.

Manage, A. B. W., Mallawaarachchi, K., & Wijekularathna, K. (2010). Receiver operating characteristic (ROC) curves for measuring the quality of decisions in cricket. *Journal of Quantitative Analysis in Sports, 6*(2), Article 8.

Mani, E. (2009). A strong sport growing stronger: A perspective on the growth, development and future of international cricket. *Sport in Society, 12*(4&5), 681–693.

McAuliffe, K. W., & Gibbs, R. J. (1997). An investigation of pace and bounce of cricket pitches in New Zealand. *International Turfgrass Society Research Journal, 8,* 109–119.

McHale, I. G., & Asif, M. (2013). A modified Duckworth-Lewis method for adjusting targets in interrupted limited overs cricket. *European Journal of Operational Research, 225,* 353–362. https://doi.org/10.1016/j.ejor.2012.09.036.

Mehta, R. D. (2005). An overview of cricket ball swing. *Sports Engineering, 8,* 181–192.

Morris, D. G. S. (1976). Effects of ball and background colours have upon the catching performance of elementary school children. *The Research Quarterly, 47,* 409–416.

Parker, D., Burns, P., & Natarajan, H. (2008, October). Players valuations in the Indian Premier League. *Frontier Economics* 1–17.

Petersen, C., Pyne, D. B., Portus, M. J., & Dawson, B. (2008). Analysis of Twenty/20 cricket performance during the 2008 Indian Premier League. *International Journal of Performance Analysis in Sport, 8*(3), 63–69.

Pollard, R. (1977). Cricket and statistics. In S. P. Ladany & R. E. Machol (Eds.), *Optimal strategies in sports*. Amsterdam: North Holland.

Pollard, R., Benjamin, B., & Reep, C. (1977). Sport and the negative binomial distribution. In S. P. Ladany & R. E. Machol (Eds.), *Optimal strategies in sports*. Amsterdam: North Holland.

Preston I., Thomas J., (2000) Batting Strategy in Limited Overs Cricket. Journal of the Royal Statistical Society: Series D (The Statistician), *49* (1),95-106

Preston, I., & Thomas, J. (2002). Rain rules for limited overs cricket and probabilities of victory. *Journal of the Royal Statistical Society Series D (The Statistician), 51,* 189–202. https://doi.org/10.1111/1467-9884.00311.

Rajadhyaksha, G., & Arapostathis, A. (2004). Cricket strategy: Using a Bayesian network to recommend the best bowling option. In M. Hubbard, R. D. Mehta, & J. M. Pallis (Eds.), *The engineering of sport V* (pp. 425–431). International Sports Engineering Association.

Rastogi, S. K., & Deodhar, S. Y. (2009). Player Pricing and valuation of cricketing attributes: Exploring the IPL Twenty20 vision. *Vikalpa, 34*(2), 15–23.

Ray, A. (2007). *One-day cricket the Indian challenge*. New Delhi: Harper Collins Publishers.

Ringrose, T. J. (2006) Neutral umpires and leg before wicket decisions in test cricket. *Journal of the Royal Statistical Society, Series A, 169*(Part 4), 903–911.

Saikia, H., & Bhattacharjee, D. (2010). On the effect of home team advantage and winning the toss in the outcome of T20 international cricket matches. *Assam University Journal of Science and Technology, 6*(2), 88–93.

Saikia, H., & Bhattacharjee, D. (2011). On classification of all-rounders of the Indian Premier League (IPL): A Bayesian approach. *Vikalpa, 36*(4), 51–66.

Saikia, H., Bhattacharjee, D., & Bhattacharjee, A. (2013). Performance based market valuation of cricketers in IPL. *Sport, Business and Management: An International Journal, 3*(2), 127–146.

Saikia, H., Bhattacharjee, D., & Lemmer, H. H. (2012a). Predicting the performance of bowlers in IPL: An application of artificial neural network. *International Journal of Performance Analysis in Sport, 12,* 75–89.

Saikia, H., Bhattacharjee, D., & Lemmer, H. (2012b). A double weighted tool to measure the fielding performance in cricket. *International Journal of Sports Science and Coaching, 7*(4), 699–713.

Sankaranarayanan, V. V., Junaed Sattar, J., & Lakshmanan, V. S. (2014). Auto-play: A data mining approach to ODI cricket simulation and prediction. In M. Zaki, Z. Obradovic, P. N. Tan, A. Banerjee, & S. Parthasarathy (Eds.), *Proceedings of the 2014 SIAM International Conference on Data Mining* (pp. 1064–1072). http://dx.doi.org/10.1137/1.9781611973440.121.

Sayers, A. T., Koumbarakis, M., & Sobey, S. (2005). Surface hardness of cricket bat after 'knock-in'. *Sports Engineering, 8,* 233–240.

Schofield, J. A. (1988). Production functions in the sports industry: An empirical analysis of professional cricket. *Applied Economics, 20*(2), 177–193.

Schumaker, P., Solieman, O. K., & Chen, H. (2010a). *Sports data mining*. New York: Springer.

Schumaker, R. P., Solieman, O. K., & Chen, H. (2010b). Sports knowledge management and data mining. *Annual Review of Information Science and Technology, 44,* 115–157. https://doi.org/10.1002/aris.2010.1440440110.

Shah, P. M. (2012). *Evolution of cricket and comparison to baseball*. Unpublished Dissertation for Master of Science Program in Computer Science, San Diego State University, USA.

Singh, S., & Adhikari, A. (2015). A new net resource factor based alternative method to calculate revised target in interrupted one day cricket matches. *American Journal of Operations Research, 5*, 151–167. http://dx.doi.org/10.4236/ajor.2015.53012.

Singh, S., Gupta, S., & Gupta, V. (2011). Dynamic bidding strategy for players auction in IPL. *International Journal of Sports Science and Engineering, 5*(1), 003–016.

Solieman, O. K. (2006). *Data mining in sports: A research overview* (MIS Masters Project), Department of Management Information System, The University of Arizona, Tucson.

Stefani, R. T., & Clarke, S. R. (1992). Predictions and home advantage for Australian rules football. *Journal of Applied Statistics, 19,* 251–261.

Suleman, M. T., & Saeed, M. A. (2009). Option on human performance: A case study of Indian Premier League. Retrieved December 21, 2009, from http://papers.ssrn.com/sol3/…/SSRN_ID1474093_code1208558.pdf.

Sumner, J., & Mobley, M. (1981, July). Are cricket umpires biased? *New Scientist* 29–31.

Swartz, T. B. (2011). Drafts versus auctions in the Indian Premier League: Theory and methods. *South African Statistical Journal, 45*(2), 249–272.

Underdown, D. (2000). *Start of play: Cricket and culture in eighteenth-century England* (pp. 5–7). London: Allen Lane.

van Staden, P. J. (2009). Comparison of cricketers bowling and batting performances using graphical displays. *Current Science, 96,* 764–766.

William Elderton, George H. Wood, (1945) Cricket Scores and Geometrical Progression. Journal of the Royal Statistical Society *108* (1/2):12

Williamson, M. (2007), *The oldest international contest of them all*. www.espncricinfo.com/magazine/content/story/141170.html. Retrieved on July 15, 2018.

Wood, G. H. (1945a). What do you Mean by Consistency?. The Cricketer Annual. 22-28.

Chapter 2
Franchisee Cricket and Cricketers' Auction

2.1 Franchisee Cricket—The New Beginning

Cricket, in general, is termed as a gentlemen's game. Obviously, there are reasons why the game is called so. In the early days, the professionals and the amateurs were clearly differentiated. Professionals are those players who played for money and amateurs are the men who played for pleasure. A first-class cricket match between the *professionals* and the *amateurs* was played in England from 1806 to 1962 almost annually. But on January 31, 1963, the concept of amateurism was withdrawn by the MCC and all cricketers playing first-class cricket were considered to be professional players. In the Olympic Games, professional athletes were allowed to participate only in 1986. That way Cricket seems to be modern compared to the Olympics.

Cricket generally began in a territorial arrangement. Rather than that, franchisee cricket (with corporate marking) began with the Indian Premier League (IPL), pursued intently by a few others. The provincial organization is as yet famous, as confirmed in England. The most seasoned twenty-over cricket rivalry is the English Twenty20 challenge. Indeed, even as late as 2016, they have restricted a move from the customary region arrangement to franchisee design.

2.2 The Indian Premier League (IPL)

With the help of the International Cricket Council (ICC) the Board of Control for Cricket in India (BCCI), initiated the Indian Premier League (IPL) in 2008. The basic reason for starting it was to counter a similar cricket tournament stated earlier, in India with professional players by a television company, without the consent of BCCI. The tournament is a professional cricket competition with domestic as well as international players of twenty-over format. The format of the tournament and the way the teams were formed through an auction between the franchisees had some

H. Saikia et al., *Cricket Performance Management*, Indian Statistical Institute Series,
https://doi.org/10.1007/978-981-15-1354-1_2

similarity with the National Basketball Association of the USA and the English Premier League of England. Following its inauguration in 2008, IPL had significant commercial success. The IPL model was followed by several other countries playing cricket to design their own franchisee leagues like the, Bangladesh Premier League (BPL in Bangladesh), East Africa Premier League, Caribbean Premier League (earlier called the Caribbean Twenty20), Big Bash (in Australia), the Sri Lankan Premier League (SLPL), etc.

IPL could maintain its standard and sustained itself among such Twenty20 cricket tournaments that followed IPL's model and came into existence. Following the success of IPL in India, other sports too started their franchisee-based tournaments; this includes—Pro Kabaddi League, Indian Super League of Football, Champions Tennis League, Hockey India League, Premier Badminton League, Pro Volleyball League, etc.

The format of the IPL Twenty20 competition is twofold round robin, where each group (for a few seasons in gatherings) plays with each other in the group twice, one match at the home ground and the other in the home ground of their opponent called as the away match. The top four teams in the points table then reach the knock-out stage, called the playoffs. The other teams get eliminated for that season for the remaining part of that year IPL. Initially, the top four teams played in the semi-finals and the finals to get the ultimate winner. The number of teams that participated in the various seasons of IPL varied throughout the years running from eight to 11. This happened mainly due to poor governance. Since this is certainly not a central worry here, further subtleties of the diverse seasons can be found at IPL competition site, IPL Wikipedia page, or at the ESPN-Cricinfo site (see references).

Before the start of each season, the cricketers were auctioned. The franchisee is provided with an option to the team to retain players over several years if they are satisfied with the performance. In 2008, in IPL's inaugural season, 96 players were auctioned. In the auctions that followed in 2009 and 2010, fewer players were up for sale, only 17 and 11, respectively. In 2012, a total of 25 cricketers and 37 in 2013 was the number of players auctioned. However, in 2011, 2014, 2015, and 2017 auctions, a larger number of players viz. 127, 154, 67, and 66 cricketers, respectively, were sold. This presented an opportunity for researchers to analyze the auction patterns statistically.

We place our discussion of the IPL venture from the business point of view. The IPL was planned after the organizers made a serious investigation on the best professional sports leagues of the world. Lalit Modi, the Commissioner of IPL hired the services of International Management Group (IMG) to fine tune the structure of the league, which was initially modeled in the lines of the National Basketball Association and English Premier League (Kamath, 2008; Mitra, 2010). The establishment privileges of eight noteworthy Indian urban communities were sold by the IPL for a long time through an auction that pulled in a portion of the best industrialists and different superstars of the country and the complete move sum rounded up was $723.5 million. The 59 IPL matches were planned for a period spread over one and a half months. The broadcasting rights were sold for around $1000 million. This sum was to be paid over a time of ten years. In spite of the fact that for quite a while, cricket has been a worthwhile game in India, it has never recently included direct venture on this scale from the best business houses. The administrations of the domestic and overseas cricketers who registered to play the IPL were auctioned for almost $36 million every year for a three-year time span. The players were auctioned by methods for a sequential single item, and the terminology used in the standard English auction (Krishna, 2002). To make the auction procedure progressively appealing and energizing, IPL chose that new player auction would be held after every three years.

Following are the rules on an English Auction.

- The auction is conducted by a professional auctioneer. The auctioneer starts by announcing a starting price or reserve price for the item on sale which is called as a Suggested Opening Bid.
- Then the auctioneer invites the floor to participate in the bid with increasingly higher prices from those buyers with a possible interest in the item.
- At any point of time, the highest bidder shall win the bid unless there is any higher bid from a competing buyer.
- But if there is no competing bidder to challenge the highest bid, then the corresponding bidder wins the bid.

The English auction is a transparent type of option, as at the very outset the identity of the bidders participating in the process is disclosed to all others participating in the auction. This type of auction involves an iterative process where the price gets adjusted towards the upward direction which is in contrast to a Dutch auction where the price goes down as the bidding continues.

To clarify the colossal ventures from the business houses who were benefit driven and other individuals of high total assets, a short examination of the plan of action of IPL is significant. The wellsprings of income can be ordered comprehensively into two sections. The first is the focal incomes which incorporate the clearance of media rights (amounting to $1000 million more than 10 years), the closeout of title sponsorship of the tournament ($50 M more than five years), authorized stock, and so on. The significant piece of the focal incomes was allocated among the IPL, the establishments, and the prize cash support (Basu, 2008). The other option is the local earnings that originated from the franchises. A share amounting to 20% of what was raised from franchise rights went to their kitty. The franchisees fund-raised from moving promoting space in the arena for home matches, authorizing items like T-shirts, getting sponsorship for the jersey of the team, publicizing for match tickets, etc. Likewise, a franchise also earned an impressive amount from the tickets sold in home matches.

A major portion of the central revenue earned by the franchisee is used up to purchase the cricketers with the abnormally high bid price. The other expenses are controlled by local revenues, the main share of which comes from the sale of tickets of the home matches. Initially, the IPL governing bodies tried to augment the heads of central revenues by signing up several contracts, like that of, (i) live streaming of video on YouTube, (ii) screenings of live matches in theaters with UFO Moviez, and so on (Sheikh, 2012). This model helped the governance of the league to diminish the risk for a franchise. However, after the first four seasons of the tournament, there was a mismatch between the revenue earned by the franchises and the expenditure that they had to bear in running a team. The local revenue was good enough for the franchisee but the main source of their revenue that was from the central source was not as expected. The IPL authorities at some point of time canceled their contracts with YouTube and UFO Moviez and this stood out to be the main reason of less growth of central revenue than it was estimated (Sheikh, 2012). Also following the victory of India in 2011 World Cup cricket, which was played in India, there was a decline in the interest of corporate houses and the on-ground audience. Also because of the withdrawal of Kochi team from the fourth season of IPL, the number of matches also diminished. However, in spite of all these issues, the IPL's brand value in the fourth season was estimated at \$3.67 billion. This again reinforced the position of IPL as a very attractive investment option for business houses (Pande, 2011).

The culture of creating a fan base in India for each of the IPL teams, given the fact that such type of city/province wise fan base was a new concept in India, was a difficult task. But for retaining the interest of the people in the league, this was necessary. A simpler way to increase the revenue of the team is to make a strong team so that the team can win the trophy and earn price money of as high as 3 million dollars for the franchisee. But making a strong team means bidding for the highest-rated players with a very high price. Since the maximum amount of money in the kitty for purchase is fixed for each franchisee, so the auction turns into a problem of optimization. It is of rather building a balanced team—a balance between stars and utility players so that the team remains well-balanced and also a match-winning team. Thus, a franchise needs to create a team considering both revenue and cost.

2.3 Auction Rules for Players in IPL

It was for the first time that cricketers came under the auction hammer in 2008 for IPL. Eight franchisees were identified through a bidding process. The teams thus formed were named after eight Indian cities/provinces (called states). Each cricket team based on their regional superstars named one player as the icon player. This was done to attach regional sentiment to the teams. Two of the eight teams did not have any regional superstar and had to look for players outside their region/country as their icon player. The icon player captained the team. In perspective on the extraordinary enthusiasm for the endeavor among both the worldwide cricketing network and the Indian fans, the IPL authorities chose to sell 77 surely understood global cricket players for choice by the franchise. The purchased players shall stay with the franchise for the next three years. However, there was also a concept of mutual transfer. Each player was treated as a single item in the auction which was an English auction in nature. Till date, eleven editions of IPL are organized from 2008 to 2018. In January 2011, the second release of the player auction was organized. It was in this year that the number of franchisees was also increased from eight teams to 10 teams. But in the very next year, one team was suspended and the number of IPL franchisee came down to nine. In the second version of players' auction, 215 cricketers were placed for out of which 127 players were sold to the 10 franchises. These recreations created incredible fervor among the worldwide cricket audience. There was one more end in 2012 pursued by a substitution. In 2015, because of match-fixing charges, two teams were suspended for the next two seasons and supplanted by two other teams only for those two years.

Two aspects related to the IPL auctions are of interest to sports analytics. The first is about proper valuation of a cricketer. This is yet to be studied adequately. Only a few works on this issue are available. Saikia, Bhattacharjee and Bhattacharjee (2013) based on performance-related data of cricketers from the first three seasons of IPL tried to identify the cricketers who have justified their pay package by their performance applying the binomial pricing option. Applying multiple regressions, Rastogi and Deodhar (2009) tried to find out the factors that actually influenced the valuation of the players who participated in the IPL auction of 2008. The study found that along with the cricketing attributes, some other non-cricketing attributes like glamor, controversies, etc. too influence the price of the players. A similar work was also attempted by Parker, Burns, and Natarajan (2008). The valuation of players is found to be highly related to their past performances, age, nationality, and other parameters have been observed. Using the hedonic price equations, Karnik (2010) attempts to estimate the price of a cricketer in the IPL 2008 auction. The other issue is related to the assignment of players to the franchisees. The main concern in doing so is to develop the exercise in such a manner that the teams formed shall be to a greater extent evenly matched in their overall playing ability. Lenten et al. (2012) analyzed the price-related information from the auction of IPL 2008 using a different methodology. They found the evidence that Indian players were purchased at a higher price compared to cricketers of other countries. This may be because of the fact that the Indian players are available across the entire season but the foreign players might be unavailable because of their response to national duty. The findings of their study also indicated overpricing of the star players. Some other studies on the determinants of a player's wages include (Bennett & Flueck, 1983; Dobson & Goddard, 1998; Estenson, 1994; Hausman & Leonard, 1997; Jones & Walsh, 1988; Kahn, 1992; MacDonald & Reynolds, 1994). However, very few works on the allocation of players are noted. Here, the concern was assigning cricketers to teams such that the teams remain very competitive to each other. The auction of IPL was thus designed with some constraints on the franchisee in such a manner that natural balances of the teams are attained.

In a cricket match, 11 players take the field, but a team roster has at least 16 players. During the auction, a franchise is concerned with forming a core team. The IPL authorities passed a mandate that could allow each team to take at most eight foreign players in their roster. However, of all the foreign players in the squad, a maximum of four foreign players could participate in the playing 11. To encourage local players, the IPL authorities provided the franchise with a local area in and around the neighborhood of the city or region after which the franchise is named. From the said local area, the franchise is to include at least four players. In addition, to provide young cricketers exposure, four players under 22 years of age are to be included. Practically, each of the players accessible from the catchment zone had almost no experience of playing cricket at the international level. The inclusion of the best of the available overseas player in the alliance guaranteed a high standard of rivalry which suggested that none of the teams could bear to keep a job of a specialist role free for a local player.

The player auction was conducted by a well-known, independent, professional auctioneer. He is Richard Madley, one of the most renowned auctioneers stationed at New York. Each player was individually placed for bidding with player statistics displayed on the giant screen. The IPL authorities fixed a base price for each player, which was the reserve price for that player. Any franchise who wants to procure the services of a particular player had to start with a bid above his base price. The IPL authorities in consultation with a player set the base price of the player. So far, no dispute has been recorded in fixing the base price of the player between the player and IPL authorities. The base price for a player is based on the freely accessible data on the players (Borooah & Mangan, 2012) and along these lines, fills in as a decent proportion of the relative worth or the minimal income item (MRP; Fort, 2006) of the player, opposite different players. So, one can assume that a player having a base price higher is a player having more value corresponding to a player with a lower base price. The base prices of each of the player were made public before the auction so that the franchisees can frame their strategy before they join the auction. The auctioneer at the opening of the bid for that particular player also declares the base price of that player. The auctioneer declares the highest bid of a franchisee for a given player as the final bid at which the service of the player is procured by the concerned franchisee. The final bid is the fee that the franchise has to pay to the player for one season.

The accomplishment of IPL relies upon a highly challenging competition. Thus, the IPL needed to ensure the groups were generally of comparable quality. This was attained by putting an equivalent spending imperative on each team; it could spend a maximum sum of $5 million for getting the players. This kept the monetarily more grounded franchisees from obtaining all the best players. But there is a lower limit as well. Each franchisee has to spend at least $3.5 million for procuring the cricketers, failing which they have to pay as a penalty the amount less paid than $3.5 million IPL authority. The rules were such that the franchisees were forced to spend at least $3.5 million to the authorities of IPL. This constraint restricted a franchise from compiling a weaker team. During the first auction of IPL in 2008, the initial 77 players who expressed their interest to participate in the auction were divided into eight groups. The sets were labeled from A to H. From each set, the players were selected for the auction in a random fashion. The auction was done with the set A followed by set B and so on. Also, players are offered for auction one after the other. Thus, it is only after the completion of the bid of one player that another player is offered for bidding. Withdrawal of bid at any stage is not allowed. So, the franchisees have to decide before participating in the bid of a player. The highest bidder wins the player and the player shall play for the franchisee. The bidding process is public before all the franchisee. So, all the franchisee knows which player goes under which franchisee and at what price. The amount of money left to the kitty after bidding for a player is also known to the franchisee. To evaluate the current status of the franchisee and to make strategic changes there off, some time was provided. This was done at the end of auctioning all the players of a given set. In case there are no bidders for a player, the player shall go unsold. Each of the franchisees can bring in their representatives, who are generally cricket experts or veteran cricketers. However, there is no means of contacting anyone outside the auction room.

But the first action resulted in a completely unexpected result. The base price had no relation with the ultimate price at which the players were sold. Several well-known international stars were not sold in the first round of auction because of no bidders and had to be auctioned again. Most of such players were sold at their base price only. Some Indian cricketers like Ishant Sharma, who were new entrants to the Indian national team in 2008, got sold at a price which was six times more than his base price. Thus, the franchises were unable to correctly estimate the price of the cricketers (Fort, 2006). The process of valuation of the IPL authorities was different compared to the way in which the franchisees looked into it. In the work of (Borooah & Mangan, 2012), huge deviations in base prices and the final bid prices of cricketers in the IPL 2008 auction were pointed out. This happened mainly because in those days, Twenty20 was a lesser-known format and that most of the franchisees misjudged what the performance of the cricketers would be in the lesser-known format. Some ironic incidents that happen in IPL 2008 auctions were as—Shane Warne who not only created magic with his spin but also spun his team Rajasthan Royals (RR) to win the inaugural version of the IPL with his captaincy (where he managed a team without any star player) had a base price of $4,50,000 and was not bid by any franchisee in the initial round. RR later picked him up at his base price, and what a decision it was! Same happened to Glenn McGrath who was sold at his base price of $3,50,000 to Delhi Daredevils, as there was no other bidder. However, Ishant Sharma, just new to the Indian national team at that period, was sold at $9,50,000 with his base price of $1,50,000 much lower than Glenn McGrath's base price. Ishant Sharma ended up performing poorly for his franchisee and his worth was much less than the value he got. Rickey Pointing, the Australian captain of those days, was sold with a hike of just $65,000 to his base price $3,35,000 while Mahendra Singh Dhoni with a base price of $4,00,000 was ultimately purchased by Chennai Super Kings at a higher ultimate price of $15,00,000. Following the mismatch, Swartz (2011) proposed other methods of the auction that would be successful for IPL like—simultaneous ascending auctions (Cramton, 2006) and various other combinatorial auctions (Krishna, 2010).

Many franchises were unable to bid successfully for their targeted player and accordingly settled for a player with similar quality in the later part of the auction even at a higher price than expected. Borooah and Mangan (2012) note an interesting issue in connection to the IPL auction of 2011. Suppose a franchisee is interested in acquiring a particular fast bowler and another all-rounder. As the players are offered randomly for auction—let us assume, without any loss of generality that the all-rounder is offered for auction ahead of the first bowler. The franchisee interested in the all-rounder starts bidding in favor of him. Given that there are other bidders for the player, the price of the all-rounders keeps on increasing. Thus, the concerned franchisee is locked in a catch-22 situation, either they have to keep on bidding for the all-rounder and fail to acquire the bowler or to stop bidding for the all-rounder after a certain stage and keep the money to bid successfully for the bowler. The authors cited this as one of the main reasons for the large variation between the base price and the final bid price. The problem as explained above is termed as the exposure problem.

2.4 Mathematical Models for Players Selection in Auctions

One solution to the exposure problem in IPL auctions is to go for combinatorial auctions. The process of combinatorial auction allows bidders to place bids on packages (Klemperer, 2004; Krishna, 2002).

A combinatorial auction is a bidding process where items offered for bidding are bundled into different packages and the bidding is done for the entire package and not for an individual item. Initially, such auctions were very specific to estate auctions, but of late, such combinatorial auctions are used recently for truckload transportation, bus routes, industrial procurement, time slots allotted by airport authorities for takeoff and landing of flights, and in the allocation of radio spectrum for wireless communications (Chakraborty, Sen, and Bagchi, 2015).

But combinatorial auctions have some difficulties when compared to the existing single-item auctions. The first challenge is how to form the packages before it is submitted to the auctioneer. This may, for example, comprise of an opening batsman, an all-rounder, a wicket keeper, a fast bowler, and a spinner. Now, given that the best batsman of all the available batsmen falls in a package (Package A, say), it shall be compensated by a lower-ranked wicket keeper or a lower-ranked fast bowler, etc. so that the competitive balance between the packages is retained. This combinatorial auction problem can be modeled as a set packing problem. Several authors proposed different algorithms to solve this set packing problem in a combinatorial auction. Hsieh (2010) used a Lagrangian relaxation approach for the purpose.

Several problems coming out of combinatorial auctions, when applied to real-world situations, are well discussed in a book edited by Cramton, Shoham, and Steinberg (2006). The issues of computational complexity arising out of the set-packaging problem and their solution were initially discussed in Rassenti, Smith, and Bulfin (1982). This was the work, which first introduced the concept of such type of auction.

There are several fields of business where combinatorial auctions are successfully used. A detailed review of the application of such type of auction can be found in the work of Chakraborty, Sen, and Bagchi (2015). Briefly, the following applications are provided—(i) allotment of railroads (Leyton-Brown, Pearson, & Shoham, 2000), (ii) auction of adjacent pieces of real estate (Leyton-Brown et al., 2000), (iii) allocation of airport takeoff and landing slots (Rassenti & Smith, 1982), (iv) bundling of routes for transportation and logistics services (Yossi, 2004), (v) Federal Communications Commission spectrum auctions (Goeree, Holt, & Ledyard, 2006; Kwasnica, Ledyard, Porter, & DeMartini, 2005), and (vi) supply chain management (Walsh & Wellman, 2000) among others. In most of the aforesaid works, it was found that the use of combinatorial auctions led to an increase in bidder satisfaction and also augmented the earnings of the auctioneer.

However, the full benefit of the combinatorial auction is yet to be explored. The main reason why this type of auction is less popular is the problem of valuation. In a combinatorial auction, the bidders face a difficult situation where they are unable to determine the price to which they shall continue bidding for a package. This is called as the valuation problem. To avoid this problem, the bidders also need to be exposed in advance about the combination of each of the packages so that they can plan accordingly. Thus, the sealed bid formats, where the bidders shall be aware of the contents of the package only when the auction is in progress is not suitable for combinatorial auctions especially in connection to the player's auction. This problem can be moderated by conducting a multiround combinatorial auction. In such a type of auction, the bidders are provided with a chance at the start of each round to modify their bids on the basis of the prior information of the bids made by others.

A closely related issue of the valuation problem is the availability of data. In a single-item auction, the information about the most noteworthy offer on an item is sufficient for the bidder. But on account of the combinatorial auction, a bidder can, in any case, be one of the winners without putting the most astounding offer essentially as a result of the offers on different combinations. Along these lines, to enable bidders to offer genuinely, the auctioneer needs to give applicable data to the bidders on the present condition of the bid. For instance, the bidders need to realize the current temporary winning cost for each package which may change after each offer. So as to give this data, the barker needs to take care of the Winner Determination Problem (WDP) after each bid. This necessity of over and again illuminating the WDP is a computationally difficult issue (Sandholm, 2002). This shall also increase the preparing cost of the auctioneer.

2.5 Predicted Bid Price of Cricketers

This section provides a detailed review of literature on modeling of different aspects of cricket with special reference to auctions. Rastogi and Deodhar (2009) looked into players valuation in IPL. Some works addressed other issues of IPL than cricketers' auction. Petersen et al. (2008) quantified the performance of cricket teams and zeros down to the indicators of success in IPL. Ramanna (2009) noted the events that led to the development and launching of the first IPL season.

Swartz (2011) examines the 2008 player auction used in the IPL. He argued that the auction was less than satisfactory and supported the application of draft instead of auction, where the player's salaries are determined by draft order. He fitted a three-parameter lognormal distribution to the salaries of cricketers participating in the first season of IPL. Parameters of the distribution were set according to the financial constraints of the team. The paper is extensive in the sense that the application of draft procedure is explored in the context of the IPL auctions and in other sports which including basketball, highland dance, golf, tennis, car racing, and distance running.

Saikia and Bhattacharjee (2011) employed the stepwise multinomial logistic regression to identify the significant predictors of performance of all-rounders in IPL. They found that strike rate in ODI, strike rate in Twenty20, the economy rate in ODI, the economy rate in Twenty20, and bowling type (spin or fast) are significant in determining the class of an all-rounder. Thereafter, the naïve Bayesian classification model is used for forecasting the expected class of six incumbent all-rounders based on the significant predictors, who had played only in the fourth season of IPL. The prediction done before the fourth season of IPL was then compared with the actual situation at the end of the tournament.

Saikia, Bhattacharjee, and Bhattacharjee (2013) developed a measure that can quantify the batting, bowling, and wicket keeping the performance of a cricketer into one single index. They called the index as the performance index. Based on the values of the performance index of cricketers, from the first three seasons of IPL and using the binomial option pricing model, the neutral present values of the cricketers are determined. The distributional pattern of the present values of cricketers is identified and cricketers are classified based on the level of their present values. Similarly, the distributional pattern of the bid prices of cricketers is identified and the cricketers are classified on the level of their bid prices. The cross-tabulation of the classified bid prices and classified present values can be used to identify the cricketers who were able to justify their salaries, who were underpaid and who were paid more than their performance. The model can be used by franchisees to decide on which players should be considered and who should be dropped for a given price. The players can also use this model to understand what their market price should be and accordingly get themselves registered for IPL auction at a price they deserve.

In IPL, the bidding process is dynamic in nature. Once the bidding started, there was no break. The franchisee did not get an opportunity to sit back and rethink their bidding strategy, to take stock of the amount of money left and the kitty and plan differently. Every planning was to be done instantly. Thus, no fixed plan worked and the franchises had to respond and realign their strategy after each bid by taking into account whether the bid went in their favor or not. This phenomenon was not properly apprehended by some franchises. These franchisees found it difficult to revise their strategy when the auction was in progress.

An integer programming model for the resolution of this issue was suggested by Singh et al. (2011). He suggested this as an efficient and successful way to determine a strategy for bidding by the franchisees. The model can be developed in Excel-like a spreadsheet and enables bidders to take decisions in real time.

The papers named above explored different magnitudes of the league. However, it is surprising that only a few papers addressed the event of cricketers' auction directly. The players' auctions in IPL should have interested several researchers' where data on bidding, base prices, and performance of the cricketers is available. But such studies did not happen in the way we expected. Here, we tried to analyze four IPL auction oriented papers. Lenten et al. (2012) studied the 2008 IPL cricketers' auctions and indicated that most players are not worth the price at which they were sold. Accordingly, they proposed several alternate ways to explain the bidding behavior with the application of different sets of variables under the different proposed models. However, the paper enforces some dummies like X-Factor to summarily ignore the outliers, and also the number of variables is very high, while the authors did not adhere to the justification for inclusion or exclusion of different variables. The models lack predictive power and hence forecast the prices of players in the future auction cannot be done.

Parker et al. (2008) used OLS linear regression model to identify the significant variables that determine the bid price of a player. They suggest that experience of playing at the international level in both Twenty20 and one-day internationals is positively related to the valuation of a cricketer. Likewise, young and Indian players have a chance of being sold at a higher price. This model again lacks in its predictive power. The model is a post-facto one and it specifies dummies for outliers in connection to the final bid values of cricketers. The icon status of cricketers is also included in the model as dummy variables. But these are not to be modeled as variables as the salaries of the five icon players were decided through a contract with the icon players before the auction. Further, several 'artificial' variables are introduced in the model, like the top-10 all-rounders as a dummy variable. This has lead to overfitting of the model.

The approach of Karnik (2010) was from the other end vis-à-vis Parker et al. (2008). It has more variables on the basis of calculated intuitions that eventually emerge as a self-fulfilling prediction. Karnik (2010) presumes that a number of runs scored and the number of wickets taken is significant variables of the model. But a discussion on why in their work includes an explanation on runs or wickets in a particular format of cricket shall be considered is missing. Further, the work uses some constructed variables, which are never common in literature concerning cricket analytics. Variables like 'ratio of wickets taken by a player in one-day and T20 formats to the total wickets taken by all the 75 players, expressed as a percentage' is one such example. Such types of variables make the study sensitive to the sample data of those 75 cricketers only. This also adds doubt to the calculations and the predictability of the model.

Based on cricketing and non-cricketing qualities of several players, Rastogi and Deodhar (2009) break down the bidding behavior of the various franchisees. They, as in Karnik (2010), conjure the Hedonic Price Modeling yet begin with a bigger arrangement of factors and wiped out the non-significant factors in an iterative manner; in this way, giving a comparative yet better model than that of Karnik

(2010). In their study, some non-cricketing qualities, like that of age, nationality, and franchisee's own conduct are critical determinants of conclusive offered costs. For instance, Indian players claimed a higher premium over Pakistani players and Mumbai Indians offered higher bids than that of the other franchises. A portion of the cricketing properties, like batting strike rate in one-day internationals, number of 50 + scores, stumping, and wickets taken, significantly influence the bid price offered to the players.

The work is definitely a nice attempt, as it takes into consideration several variables in the model. But the predictive capacity of the model is likely to suffer as it includes several non-cricketing attributes of the player like appeal, controversy, icon player status, etc., which are also highly subjective in nature. The value of R-square increases because of the franchisee dummies. This is done to capture the franchisee specific eccentricities, as mentioned by the authors. But we doubt if such franchisee specific dummies have increased the actual predictive power of the model. As mentioned by Patterson (2000), Kennedy (2002), and Gujarati (1993), the model specification must have a reasonable combination of economic theory that passes the test of available empirical data, statistically significant regression coefficients with expected signs, a reasonably high (adjusted) R-square, and must maintain prudence with sufficient degrees of freedom. Depending on the light of the above discussion, the model of Rastogi and Deodhar (2009) needs some changes to arrive at a model that has higher predictive power.

In the work of Chakraboty, Sen, and Bagchi (2015), the variables considered are broadly classified into three parts—the on-field performance variables, some demographic characteristics of the player like—age and nationality, and other off-the-field attributes such as glamor quotient. The variables relating to cricketing performances again classified into four groups viz. batsmen, bowlers, wicket keepers, and all-rounders. The response variable is the final bid price of the player. While dealing with the independent variables, the major concern is to consider or ignore the performance-related statistics collected from the various formats of cricket—test matches, one-day internationals (ODIs), international Twenty20 matches, unlisted Twenty20 matches (like IPL), and first-class cricket. The variables considered from the aforesaid five formats were—runs scored, batting average, batting strike rate, number of centuries, number of half-centuries, player's age, number of catches taken, number of stumpings, number of wickets taken, bowling average, bowling economy rate, and bowling strike rate. The authors then applied a mixed approach to select variables from the long-list of cricketing, non-cricketing, and personal attributes across the different formats. Later, a mixture of logical arguments and three other papers on players' auction were referred to reach the desired model.

References

Basu, I. (2008, January 27). Does the IPL model make sense? *The Times of India*. Retrieved from http://articles.timesofindia.indiatimes.com/2008-01-27/special-report/27748861_1_ipl-indian-premier-league-prize-money.

Bennett, J. M., & Flueck, J. A. (1983). An evaluation of major league baseball offensive performance models. *The American Statistician, 37,* 76–82.

Borooah, V. K., & Mangan, J. (2012). Mistaking style for substance: Investor Exuberance in the 2008 Indian premier league auction. *Journal of Sports Economics, 13,* 266–287.

Chakraborty, S., Sen, A. K., & Bagchi, A. (2015). Combinatorial auctions for player selection in the Indian Premier League (IPL). *Journal of Sports Economics, 16*(1), 86–107.

Cramton, P. (2006). Simultaneous ascending auctions. In P. Cramton, Y. Shoham, & R. Steinberg (Eds.), *Combinatorial auctions* (pp. 99–114). Cambridge, MA: The MIT Press.

Dobson, S., & Goddard, J. (1998). Performance, revenue, and cross subsidization in the football league, 1927–1994. *The Economic History Review, 51*, 763–785.

ESPN-Cricinfo. http://www.espncricinfo.com/indian-premier-league-2014/engine/series/695871.html?view=records. As on January 22, 2015.

Estenson, P. (1994). Salary determination in major league baseball: A classroom exercise. *Managerial and Decision Economics, 15,* 537–541.

Fort, R. D. (2006). *Sports economics* (3rd ed.). NJ: Prentice Hall.

Goeree, J. O., Holt, A. C., & Ledyard, J. O. (2006). An experimental comparison of the FCC's combinatorial and non-combinatorial simultaneous multiple round auctions. Report for Wireless Telecommunications Bureau of the Federal Communications Commission (FCC). Retrieved from http://wireless.fcc.gov/auctions/data/papersAndStudies/fcc_final_report_071206.p.

Gujarati, D. N. (1993). *Basic econometrics* (4th ed., p. 516). New York: McGraw Hill.

Hausman, J. A., & Leonard, G. K. (1997). Superstars in the national basketball association: Economic value and policy. *Journal of Labor Economics, 15,* 586–624.

Hsieh, Fu-Shiung. (2010). Combinatorial reverse auction based on revelation of Lagrangian multipliers. *Decision Support Systems, 48*(2), 323–330.

IPL Tournament Website. http://www.iplt20.com/stats. As on as on January 22, 2015.

IPL Wikipedia. http://en.wikipedia.org/wiki/Indian_Premier_League. As on 22 January 2015.

Jones, J. C. H., & Walsh, W. D. (1988). Salary determination in the national hockey league: The effects of skills, franchise characteristics, and discrimination. *Industrial and Labor Relations Review, 41,* 592–604.

Kahn, L. M. (1992). The effects of race on professional football players' compensation. *Industrial and Labor Relations Review, 45,* 295–310.

Kamath, V. (2008, February 2008). IPL is an idea whose time has come. Business Line. Retrieved from http://www.thehindubusinessline.com/todays-paper/article1111722.ece.

Karnik, A. (2010). Valuing cricketers using hedonic price models. *Journal of Sports Economics, 11,* 456–469.

Kennedy, P. E. (2002). Sinning in the basement: what are the rules? The ten commandments of applied econometrics. *Journal of Economic Surveys, 16*(4), 569–589.

Klemperer, P. (2004). *Auctions theory and practice*. Princeton, NJ: Princeton University Press.

Krishna, V. (2002). *Auction theory*. London, England: Academic Press.

Krishna, V. J. (2010). *Auction theory* (2nd ed.). Burlington, MA: Academic Press/Elsevier.

Kwasnica, A. M., Ledyard, J. O., Porter, D., & De Martini, C. (2005). A new and improved design for multi-object iterative auctions. *Management Science, 51,* 419–434.

Lenten, L. J. A., Geerling, W., & Konya, L. (2012). A hedonic model of player wage determination from the Indian premier league auction: Further evidence. *Sport Management Review, 15,* 60–71.

Leyton-Brown, K., Pearson, M., & Shoham, Y. (2000). Towards a universal test suite for combinatorial auction algorithms. EC '00 Proceedings 2nd ACM conference Electronic Commerce, (ACM Press, New York), pp 66–76.

MacDonald, D. N., & Reynolds, M. O. (1994). Are baseball players paid their marginal products? *Managerial and Decision Economics, 15,* 443–457.
Mitra, S. (2010). The IPL: India's foray into world sports business. *Sport in Society, 13*(9), 1314–1333.
Pande, B. (2011, April 7). IPL-4: IPL's brand value falls by 11% to $3.67 billion. The Economic Times. Retrieved from http://articles.economictimes.indiatimes.com/2011-04-07/news/29392570_1_brand-finance-indian-premier-league-ipl.
Parker, D., Burns, P., & Natarajan, H., (2008). 'Player valuations in the Indian Premier League. *Frontier Economics*, 1–17 Oct 2008.
Patterson, K. (2000). *An introduction to applied econometrics* (p. 10). New York: St. Martin's Press.
Petersen, C., Pyne, D. B., Portus, M. J., & Dawson, B. (2008). Analysis of Twenty/20 Cricket performance during the 2008 Indian Premier League. *International Journal of Performance Analysis in Sport, 8*(3), 63–69.
Ramanna, V. S. (2009). *Indian Premier League* (SSRN Working Paper 1354534). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1354534&download=yes. As on January 22, 2015.
Rassenti, S. J., & Smith, V. L. (1982). A combinatorial auction mechanism for airport time slot allocation. *Bell Journal of Economics, 13,* 402–417.
Rastogi, S., & Deodhar, S. (2009). Player pricing and valuation of cricketing attributes: Exploring the IPL twenty-twenty vision (Working Paper). Ahmedabad, India: IIM Ahmedabad.
Saikia, H. & Bhattacharjee, D. (2011) On classification of all-rounders of the Indian Premier League (IPL): A Bayesian approach. *Vikalpa: The Journal for Decision Makers, 36* (4), 51–66.
Sandholm, T. (2002). Algorithm for optimal winner determination in combinatorial auctions. *Artificial Intelligence, 135,* 1–54.
Saikia, H., Bhattacharjee, D., & Bhattacharjee, D. (2013). Performance based market valuation of cricketers in IPL. *Sport, Business and Management: An International Journal, 3*(2), 127–146.
Sheikh, A. (2012, April 7). Five seasons on, IPL returns remain below expectations. LiveMint.com, The Wall Street Journal. Retrieved from http://www.livemint.com/2012/04/07003654/Five-seasons-on-IPL-returns-r.html.
Singh, S., Gupta, S., & Gupta, V. (2011). Dynamic Bidding Strategy for Players Auction in IPL. *International Journal of Sports Science and Engineering, 5*(1), 003–016.
Swartz, T. B. (2011). Drafts versus auctions in the Indian premier league. *South African Statistical Journal, 45,* 249–292.
Walsh, W. E., & Wellman, M. P. (2000). Combinatorial auctions for supply chain formation. ACM Conference on Electronic Commerce, (EC-00), 260–269.
Wiki TWENTY20. (2015). List of domestic Twenty20 cricket competitions. http://en.wikipedia.org/wiki/List_of_domestic_Twenty20_cricket_competitions. As on January 22, 2015.
Yossi, S. (2004). Combinatorial auctions in the procurement of transportation services. *Interfaces, 34,* 245–252.

Chapter 3
Quantifying Performance of Cricketers

3.1 Introduction

The performance statistics are often used in sports to quantify or measure the competence of the players. These performance statistics always help to indicate the level of achievement (Clarke, 2007) in terms of players, teams, games, etc. Most of the sports have performance-related statistics to quantify the level of achievement of the players. Likewise, as a team game, cricket has a number of performance-related statistics in comparison to the other team sports, as each match generates a massive amount of performance-related statistics. Relative to the other team games, the contribution of individual team members to the overall team performance is more crucial in cricket (Damodaran, 2006). Therefore, in every cricket match a scorecard is continuously maintained to collect the player specific statistical data. Now the time has come for all the teams to focus more on statistics than what they have seen in the field (Srinivas & Vivek, 2009). Since players' performances in a few matches can be quirk, statistics over a period of time cannot—and do not—fib. Earlier various common statistical concepts such as ratio and averages along with traditional performance statistics are used to measure the performance of the players in cricket. During the past few years, several authors have developed various performance measures to quantify the performance of the cricketers through combining the traditional performance statistics. The details of these performance statistics are explicitly discussed in subsequent sections of this chapter.

3.2 Performance Statistics for Batting and Bowling

Batting average: The batting average is the most well known and widely used performance measures in cricket. This statistic evaluates the batsmen's ability to score runs, and a higher value indicates a more desirable batsman. It is defined as the number of runs scored in all innings divided by the number of completed innings (i.e.,

H. Saikia et al., *Cricket Performance Management*, Indian Statistical Institute Series,
https://doi.org/10.1007/978-981-15-1354-1_3

the number of innings in which the batsman was out). The batting average of the i^{th} player is given by

$$\text{BA}_i = \frac{\text{Number of runs scored by the } i^{\text{th}} \text{ player}}{\text{Total number of complete innings by the } i^{\text{th}} \text{ player}} \tag{3.1}$$

Strike rate: Another important statistic in the limited overs cricket (i.e., ODI and Twenty20) is the strike rate of the batsmen. This statistic evaluates how quickly a batsman scores runs. This is a feature which is so crucial in limited over form of cricket. A player with a higher strike rate is generally preferred in limited overs cricket. Strike rate of a batsman is defined as the number of runs scored per k balls faced by the batsman. Generally, k is taken to be 100. Thus, strike rate of the i^{th} batsman is calculated by

$$\text{SRB}_i = \frac{\text{Number of runs scored by the } i^{\text{th}} \text{ player}}{\text{Total number of balls faced by the } i^{\text{th}} \text{ player}} \times 100 \tag{3.2}$$

Bowling average: It is the most widely used statistics in the early days of test cricket to measure the bowling ability of a player. It evaluates the average number of runs a bowler will concede per wicket taken. A lower value of bowling average indicates a more desirable bowler. The bowling average for a bowler is defined by the number of runs conceded by the bowler per wicket. The bowling average for the i^{th} player is calculated by

$$\text{BWA}_i = \frac{\text{Number of runs conceded by the } i^{\text{th}} \text{ bowler}}{\text{Number of wickets taken by the } i^{\text{th}} \text{ bowler}} \tag{3.3}$$

Economy rate: Economy rate is the statistic in cricket to measure the bowling performance of cricketers. This statistic evaluates the average number of runs a bowler will concede per over. The importance of this measure increases in limited overs cricket (i.e., ODI and Twenty20) where a bowler with a lower economy rate is often preferred to a bowler with a higher economy rate. A bowlers' economy rate is defined as the number of runs conceded per k balls, where k is often chosen to be 6. The economy rate of the i^{th} player is defined by

$$\text{ER}_i = \frac{\text{Number of runs conceded by the } i^{\text{th}} \text{ bowler}}{\text{Number of balls bowled by the } i^{\text{th}} \text{ bowler}} \times 6 \tag{3.4}$$

Bowling strike rate: One more important statistic to measure the bowling performance of bowlers in cricket is bowling strike rate. This statistic estimates the expected number of balls bowled per wicket taken by a bowler. A lower value of bowling strike rate is always desirable. The bowlers' strike rate was originally proposed by Sir Donald Bradman and defined by the number of balls bowled divided by the number of wickets taken. The bowlers strike rate for the i^{th} player is

$$\mathrm{BSR}_i = \frac{\text{Number of balls bowled by the } i^{\text{th}} \text{ player}}{\text{Number of wickets taken by the } i^{\text{th}} \text{ player}} \tag{3.5}$$

The above-mentioned performance statistics are commonly used to measure the batting as well as bowling performances of the cricketers. However, according to different authors or cricket analyst these performance statistics do not provide a fair reflection of a player's true ability of performance. Therefore, some alternative measures have been proposed through combining the traditional performance statistics by different authors in order to determine the more useful measures of player's performance. All these measures are briefly discussed in the next section.

3.3 Alternative Batting Performance Measures

Based on the above discussion and pertaining to the fact that existing simple statistical measures are not reasonable in detecting the true ability of players, different authors have developed measures to quantify the batting performance. Some of which are documented below.

3.3.1 Batting Average Defined by Wood (1945)

Wood (1945) defined the batting average of the i^{th} batsman as

$$B_i = \frac{R_i}{n}, \tag{3.6}$$

where 'n' is the number of innings batted by the batsman. Wood (1945) in his measure, defined in (3.6) considered a players' not-out scores as completed innings. If not-out innings could not occur, then B would be the mean score and would be a natural statistic regarding the set scores of a player in different innings. However, since about 10% (approx.) of all scores of the batsmen is not-out scores, computing batting average of a player using Eq. (3.6) is not compelling (Kimber and Hansford, 1993). Taking the denominator to be 'n' in Eq. (3.6) leads to a downward bias in the batting average. In addition, Eq. (3.6) introduces an upward bias in the batting average if the denominator excludes the innings in which a player remained not-out. This bias cannot seemingly be avoided.

3.3.2 Batting Average Defined by Kimber and Hansford (1993)

According to Kimber and Hansford (1993) measure, a batsman's scores were observed chronologically. Then consider a set of scores $x_1, x_2, \ldots, x_n$ for a batsman, together with indicators d_i $(i = 1, 2, \ldots n)$ where $d_i = 1$ if the batsman was out for x_i and $d_i = 0$ if the batsman was left not out with a score of x_i. They state that this situation is similar to a life test in reliability with the x_i as the component lifetimes and the d_i as censoring indicators. If the lifetimes are independent geometric random variables, each with probability mass function

$$p(x) = \theta(1-\theta)^x, \quad x = 0, 1, 2, \ldots \tag{3.7}$$

where $0 < \theta < 1$ is an unknown parameter, and if the censoring is non-informative, then it is easy to show that maximum likelihood estimate of the population mean lifetime is B. Thus, if the underlying distribution scores are geometric, B is optimal. Non-informative censoring here means that a score of x not out is representative of all scores of x or more, which is a reasonable assumption.

However, the geometric model cannot apply exactly in case of a cricketer's innings score because a cricketer's innings score does not increase by a fixed amount at each stage. In addition, since the statistical case for B rests on a geometric assumption, Kimber and Hansford (1993) checked on its empirical validity. They have seen that the geometric distribution gives a poor fit to the sets of scores, so that B is inconsistent. Yet, they have seen some evidence that the upper tail of a score distribution is roughly geometric.

Having seen that the poor fitting of a batsman's batting score to geometric distribution, Kimber and Hansford (1993) proposed an alternative batting average that does not depend on the geometric assumption. Let M and M^* be respectively the highest completed innings and the highest not-out score made by a batsman so that $H = \max(M, M^*)$. Now suppose that $M > M^*$ and let $\widehat{F}(x)$ denote the product limit (PL) estimate for $F(x)$, where $\widehat{F}(x)$ is defined for all integers x. Thus, the alternative batting average is denoted by T and it is defined as

$$T = \sum_{i=1}^{M} x\left\{\widehat{F}(x) - \widehat{F}(x+1)\right\} \tag{3.8}$$

This PL estimator (T) does not depend on parametric assumptions about the scores. It is easy to show that if a batsman has no not-out scores, then T and B coincide. Also, provided that $M > M^*$ it follows that T cannot exceed M. However, we cannot use T for Bradman's first-class career because his highest first-class not-out score is 452, and his highest completed innings score is 369 (i.e., $M < M^*$). So the assumption for PL estimator (i.e., $M > M^*$) is not valid.

However, since there is some evidence that the upper tail of a score distribution is roughly geometric, the following modified version of computing batting average formula (A) is defined for $M \leq M*$.

$$A = \frac{T + (RS/k)}{1 - R} \tag{3.9}$$

where k is the not-out scores at least as big as M, S is the sum of not-out scores and R is the probability mass function unassigned by the PL estimator. This essentially converts each not-out innings of M or more to an estimated completed score by adding on the overall average.

3.3.3 Batting Performance Measure Defined by Barr and Kantor (2004)

According to the proposed measure of Barr and Kantor (2004), a batsman's strike rate along with his batting average should be considered when establishing a batting performance measure. They argued that the batting strike rate is directly proportional to the probability of dismissal of a batsman. Barr and Kantor (2004) defined the probability of a batsman being dismissed as

$$P(\text{out}) = \frac{\text{Total number of times dismissed}}{\text{Total number of balls faced}} \tag{3.10}$$

It was observe that

$$\begin{aligned}\frac{\text{Batting strike rate}}{P(\text{out})} &= \frac{\text{Total number of runs scored/Total number of balls faced}}{\text{Total number of times dismissed/Total number of balls faced}} \\ &= \frac{\text{Total number of runs scored}}{\text{Total number of times dismissed}} = \text{Batting average}\end{aligned}$$

Now, if strike rate of the batsmen considered on the vertical axis and probability of getting out (i.e., P(out)) considered on the horizontal axis, then we may plot the characteristics of a batsman in a two-dimensional space. Let us consider that y represents the strike rate and x represents the probability of a batsman being dismissed on any given ball. It should be noted that because of uniqueness

$$\text{Batting average} = \frac{\text{Strike rate}}{P(\text{out})} = \frac{y}{x} \tag{3.11}$$

Consequently, the two-dimensional graphical depiction simultaneously captures the three crucial characteristics of a batsman's performance, viz. strike rate, the probability of getting out, and batting average (Barr & Kantor, 2004). As a result, the measure is a weighted product of batting average and strike rate. Let it be denoted as BK and was given by

$$\begin{aligned}\mathrm{BK} &= (\text{Strike rate})^{\alpha}(\text{Batting average})^{1-\alpha}\\ &= y^{\alpha}\left(\frac{y}{x}\right)^{1-\alpha}\\ &= \frac{y}{x^{1-\alpha}}\end{aligned}$$

where $0 \le \alpha \le 1$ is a parameter of the stability between strike rate and batting average. The stability parameter α varying from 0 to 1 reflects the significance of strike rate with the importance of batting average. If $\alpha = 0$, then it will not give importance on batting strike rate and, if $\alpha = 1$, then no importance on batting average. Thus, it would be better to put $\alpha = ½$ for assigning equal weightage to batting average and strike rate.

3.3.4 Batting Performance Measure Defined by Damodaran (2006)

Damodaran (2006) stated that the primary measure of a batsman's performance in cricket is the batting average, which is defined as the total number of runs scored by the player divided by the total number of innings in which the player got out. However, this measure suffers from the shortcoming that it does not provide answers to many questions that arise during the course of a game. Therefore, Damodaran (2006) makes an attempt for solving this shortcoming using a Bayesian approach.

Suppose a batsman 'i' remains not-out in his j^{th} innings and R_{ij} = runs scored by the i^{th} player in the j^{th} innings in which he remains not out. Let G_{rik} be a binary variable such that

$$G_{rik} = \begin{cases} 0, \text{ if } R_{ik} < R_{ij} \\ 1, \text{ if } R_{ik} > R_{ij} \end{cases}$$

where k is allowed to vary from $k = 1, 2, \ldots, (j - 1)$.
Let

$$n_{ij} = \sum_{k=1}^{j-1} G_{rik} \tag{3.12}$$

The value of (3.12) will be an integral value that is the number of innings in which the i^{th} batsman has accumulated more runs than his j^{th} innings. Also,
let

$$C_{ik} = \begin{cases} 0, & \text{if } R_{ik} < R_{ij} \\ R_{ik}, & \text{if } R_{ik} > R_{ij} \end{cases}$$

The estimate of the number of runs that the not-out batsman would have gone on to score is then given by

$$E_{ij} = \frac{\sum_i \sum_k C_{ik}}{n_{ij}} \tag{3.13}$$

It is the conditional average of the batsman at that point of time, given that he has already scored a certain number of runs. In every instance of a not out, the batsman's score in that innings j is replaced by the estimate E_{ij}. This approach has the advantage of handling deviations from the geometric distribution assumption.

3.3.5 Batting Average Proposed by Borooah and Mangan (2010)

In the early days of test cricket, the performance of a batsman is measured by his average score. However, a ranking based on simple averages suffers from two defects. First, it does not take into account the consistency of scores across innings. A batsman might have a higher career average but with low scores interspersed with high ones. Another might have a lower average but with much less variation in his scores. Second, it pays no attention to the contribution of a player's runs to the team total. For example, a century when the total score of a team's innings in a test match is 500 has less value compared to a half-century in an innings total of say 200. Therefore, Borooah and Mangan (2010) proposed new ways of computing batting averages using the Gini coefficient and they validate their measure using the data from the test cricket.

If N is the number of innings played by a batsman, M was completed innings, R_i be the number of runs scored by a batsman in innings i ($i = 1, 2, \ldots N$), and $\mu = \sum_{i=1}^{N} R_i/M$ represents his average score (i.e., consistency unadjusted score), then the Gini coefficient related with his scores is defined as

$$G = \frac{1}{2N^2\mu} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| R_i - R_j \right| \tag{3.14}$$

It is computed as half the mean of the difference in score between pairs of innings, divided by average score (μ). Thus, $G = 0.45$ means the difference in scores between two innings chosen at random will be 90% of the average score. If $\mu = 50$, this difference will be 45 runs.

Without reference to the team's performance, Borooah and Mangan (2010) specified that batting average cannot be a measure of performance of the players in an absolute sense. Thus, an important aspect of relative performance of the players is the contribution of individual batsmen to their team's total. Therefore, the authors had computed the total number of runs scored in all innings for each player as a percentage of the total number of runs scored by the team in those same innings.

Suppose a batsman plays $i = N$ innings ($i = 1, 2, \ldots, N$). Let $C = \sum_{i=1}^{N} R_i$ is his career's total no. of runs. Also, let his career average be $\mu = C/M$, where M ($\leq N$) is the total number of his completed innings. Then the adjusted average value for a batsman is defined by Borooah and Mangan (2010) as $\Omega = \mu + \theta$, where θ is the career average of the batsman which is adjusted either positively or negatively to reflect his contribution to the team total. However, the problem arises here is how to determine the value of θ for the batsmen. Let ν_i represents the value (per run) associated with the R_i runs scored in i^{th} innings. Thereafter the value-added adjustment (θ) for a batsman is defined as $\theta = \frac{\left(\sum_{i=1}^{N} \nu_i R_i\right)}{M}$. The value of ν_i is computed as $\nu_i = \frac{P_i - z}{z}$, where $P_i = R_i/T_i$ ($0 \leq P_i \leq 1$) and T_i represents the team's score in the i^{th} innings of a batsman's career and z represents a threshold contribution which is defined as $z = \frac{\sum_{i=1}^{N} P_i}{N}$.

3.3.6 Batting Performance Defined by Maini and Narayanan (2007)

Maini and Narayanan (2007) proposed another solution to the issue that is been based on the number of balls faced by the batsman. Let us define a variable d_{ij} such that

$$d_{ij} = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ batsman gets out in his } j^{\text{th}} \text{ innings} \\ \frac{n_{ij}}{\text{avg}(n_i),} & \text{if the } i^{\text{th}} \text{ batsman is not} - \text{out and } n_{ij} < \text{avg}(n_i) \\ 1, & \text{if the } i^{\text{th}} \text{ batsman is not} - \text{out and } n_{ij} > \text{avg}(n_i) \end{cases} \tag{3.15}$$

where n_{ij} is the number of balls played by the i^{th} batsman in the j^{th} innings. $\text{avg}(n_i)$ is the career average number of balls played by the i^{th} batsman in an innings. Under the restrictions defined above the average number of runs for the i^{th} batsman, one computed with total runs scored in the numerator and $\sum_j d_{ij}$ in the denominator.

3.3.7 A Batting Performance Measure (BP) Developed by Lemmer (2004)

The usual batting average assumes that this measure gives an equal weighting of each innings of a batsman in his career. However, Lemmer (2004) argued that this was not a realistic approach to the calculation of a batsmen's current batting average (Bretteny, 2010). Accordingly, he suggested that an exponentially weighted average should be used in which the most recent score has weight w, the second most recent score has weight $0.96w$, the third most recent score would have weight 0.96^2w, and so on. Lemmer (2004) had chosen the value of 0.96 in accordance with the weights used by Price waterhouse Coopers in their 2002 ratings of world cricketers.

Let the scores of a batsman are $X_1, X_2, \ldots, X_n$ where X_n indicates the batsman's most recent score, X_{n-1} represents the batsman's second most recent score, and so on. In this case, the exponentially weighted average (EWA) is defined as

$$\text{EWA} = wX_n + 0.96wX_{n-1} + 0.96^2wX_{n-2} + \cdots + 0.96^{n-1}wX_1 \tag{3.16}$$

where the sum of all the weights is equal to 1 (i.e., $\sum w = 1$)

Usually, the coefficient of variation (i.e., CV) is used to measure the consistency. The batting consistency of the batsmen was investigated by Barr and Van den Honert (1998) using geometric coefficient as the inverse of coefficient of variation (CV), and it was based on match scores from test cricket. However, subsequently Lemmer (2004) noticed that the CV was not an accurate measure of a batsman consistency as low not-out scores and score high above the average adversely influenced the standard deviation and thus CV (Bretteny, 2010) which would lead to inaccurate indications of a batsman's consistency. Then, Lemmer (2004) suggested that this error was corrected by using adjusted coefficient of variation (ACV). The measure ACV is calculated by traditional batting average as well as an adjusted standard deviation. The adjusted standard deviation is similar to the ordinary standard deviation except that scores above the average and not-out scores are not taken into account (Lemmer, 2004). Then, ACV is equal to the adjusted standard deviation divided by the average, and accordingly, the consistency coefficient is defined as CC= 1/ACV. This definition now has changed the direction of the measure that means the higher the value of CC, the more consistent is the batsman.

However, in order to incorporate a batsman's consistency into their batting performance measure the value of CC needed to be standardized. Then, Lemmer (2004) defined the standardized consistency coefficient as

$$\text{SCC} = \frac{\text{CC}}{\text{Avg(CC)}} \quad \text{where Avg(CC)} = \frac{1}{n}\sum_{i=1}^{n}\text{CC}_i \tag{3.17}$$

But Lemmer (2004) argued that in limited overs cricket, the batting strike rate was an important measure of a batsman's ability and should be included. In order to incorporate the strike rate into the batting performance measure, it too had to be standardized. Thus, the standardized strike rate (SSR) was defined as

$$\text{SSR} = \frac{\text{SR}}{\text{Avg(SR)}} \quad \text{where Avg(SR)} = \frac{1}{n}\sum_{i=1}^{n}\text{SR}_i \tag{3.18}$$

Then, Lemmer (2004) proposed that the batting performance measure (BP) is calculated as a product of the above three measures and it was defined as

$$\mathrm{BP} = \mathrm{EWA} \times \mathrm{SCC} \times \mathrm{SSR} \tag{3.19}$$

However, still the problem was that the standard deviation of standardized strike rate (SSR) differed to that of standardized consistency coefficient (SCC). Thus, in order to correct this Lemmer (2004) used the following technique to equate the standard deviations.

Let SSR_i is the standardized strike rate value of i^{th} player; then for all the values of i^{th} player, following transformation is made by Lemmer (2004).

$$\mathrm{SSR}_i^k = \left(\mathrm{SSR}_i^{k-1}\right)^{\frac{\mathrm{SD(SCC)}}{\mathrm{SD}\left(\mathrm{SSR}^{k-1}\right)}} \tag{3.20}$$

where k indicates the number of iterations performed, SD(SCC) represents the standard deviation of standardized consistency coefficient; and $\mathrm{SD}(\mathrm{SSR}^{k-1})$ represents the standard deviation of standardized strike rate of the i^{th} player. The process is repeated until the standard deviation of standardized strike rate (SSR) is equal to that of standardized consistency coefficient (SCC) to the required level of accuracy (usually up to 3 or 4 decimal places).

3.4 Pooled Measure for Bowling Performance

Based on the concluding discussion of Sect. 3.2 and pertaining to the fact that existing simple statistical measures of bowling performance are not reasonable in delectating the true ability of players, different authors have developed different measures to quantify bowling performance. One such very popular measure is documented below.

Lemmer (2002) proposed a bowling performance measure called the combined bowling rate (*CBR*) which comprises combining three traditional bowling statistics, viz. bowling average, economy rate and bowling strike rate. According to Lemmer (2002), the harmonic mean can be used to find the average of ratios, provided the numerator is considered as fixed and the denominator as variable. Therefore, to measure the bowling performance of bowlers Lemmer (2002) used the harmonic mean to combine the above-mentioned three traditional bowling statistics.

Let r be the total number of runs conceded by a bowler, w the total number of wickets taken by a bowler, and b the total number of balls bowled by a bowler in a series of matches. Then, the traditional bowling statistics can be defined as

$$\text{Bowling average} = \frac{r}{w}$$

$$\text{Economy rate} = \frac{r}{b/6}$$

$$\text{Bowling strike rate} = \frac{b}{w}$$

It is observed that the bowling average and economy rate have the same numerator. Thus, in order to combine three traditional bowling statistics using the harmonic mean, it is necessary to adjust the bowling strike rate to have the same the numerator as the bowling average and economy rate. For that purpose, Lemmer (2002) proposed the following adjustment to the bowling strike rate.

$$\text{Bowling strike rate} = \frac{b}{w} = \frac{b}{w} \times \frac{r}{r} = \frac{rb}{rw} = \frac{r}{rw/b} \tag{3.21}$$

This form of the bowling strike rate can now be used in the calculation of the harmonic mean to combine the three traditional bowling statistics. So, the combined bowling rate (CBR) defined by Lemmer (2002) was

$$\text{CBR} = \frac{3}{\frac{1}{\text{bowling average}} + \frac{1}{\text{economy rate}} + \frac{1}{\text{bowling strike rate}}} = \frac{3r}{w + b/6 + rw/b} \tag{3.22}$$

Thus, the values of CBR indicate the performances of the bowlers in a match or a series through the three prime skills of bowlers, viz. bowling average, economy rate, and bowling strike rate. Since low values of these performance statistics indicate good bowling, low values of CBR indicate good performances of the bowlers and *vice-versa*.

Later Lemmer (2005) improved the CBR to an adjusted measure called adjusted CBR which is denoted as CBR*. The adjusted CBR (i.e., CBR*) is more appropriate for quantifying bowling performance for small number of matches. The adjusted combined bowling rate (CBR*) for the i^{th} bowler is given by

$$\text{CBR}_i^* = \frac{3R_i'}{W_i^* + (B_i/6) + W_i^*\left(R_i'/B_i\right)} \tag{3.23}$$

where B_i is the number of balls bowled by the i^{th} bowler, W_i is the sum of weights of the wickets taken by the i^{th} bowler, and R_i' is the sum of adjusted runs (RA_{ij}) conceded by the i^{th} bowler in the j^{th} innings $\left(\text{i.e. } \sum_{j=1}^{n_i} \text{RA}_{ij}\right)$.

Now,

$$\text{RA}_{ij} = R_{ij}\left(\text{RPB}_{ij}/\text{RPBM}_j\right)^{0.5} \tag{3.24}$$

where R_{ij} = runs conceded by the i^{th} bowler in j^{th} match

$$\text{RPB}_{ij} = \frac{\text{Runs conceded by the } i^{\text{th}} \text{ bowler in the } j^{\text{th}} \text{ match}}{\text{Balls bowled by the } i^{\text{th}} \text{ bowler in the } j^{\text{th}} \text{ match}}$$

$$\text{RPBM}_j = \frac{\text{Total runs scored in the } j^{\text{th}} \text{ match}}{\text{Total number of balls bowled in the } j^{\text{th}} \text{ match}}$$

The two issues regarding this measure are noteworthy. The factor $\left(\text{RPB}_{ij}/\text{RPBM}_j\right)$ considers the match situation in which the i^{th} bowler delivered, and the factor W_i^* refuses to give the equal importance to all the wickets taken by the bowler but weights them differently based on their batting position (cf. Appendix 3.1). Since the adjusted combined bowling rate has a negative dimension, viz. lower the value better is the bowler, so to bring parity with the batting performance, CBR* is inversed and standardized by the average value of inverse CBR* across all the bowlers. If we denoted this measure by *S,* then

$$S = \frac{1/\text{CBR}_i^*}{\underset{i}{\text{Avg}}\left(1/\text{CBR}_i^*\right)} \tag{3.25}$$

3.5 Finding a Way Out to Wicket Keepers' Performance Measure

Probably, the busiest person in the cricket field is the wicket keeper. While a side is fielding, the wicket keeper has to remain alert and active during each and every ball bowled. At any point of time in the match, the batsman may offer the wicket keeper a chance or the bowler may ball a wide and the team would expect a full stretched dive from the keeper to restrict bye runs. These days—we find keepers encouraging the bowlers and the fielders probably to keep their focus. In modern cricket, the wicket keepers are expected to be a good batsman too. Several teams have wicket keepers who open the batting for the team. Sometimes, they even captain their side. Yes, you must be thinking of Sangakara of Sri Lanka. There was a time when Sangakara was the captain of Sri Lankan side, kept wickets for his team and also opened the innings while batting. How many more responsibility a single cricketer shall take! While keeping wickets, any keeper needs to stop every ball not hit by the batsman (to restrict bye runs), take catches, do stumping, implement run outs with or without the help of other fielders, provide guidance to the fielders about the direction in which the ball is moving after been hit by the batsman, take the return throws from the fielders and occasionally provide directions to the bowler.

Thus, defining a performance measurement of wicket keeper in cricket seems to be a complex exercise. While we deal with mathematical modeling, we try to convert a real-life situation into an equation or equations. Initially, we try with a simple equation(s). Such a simple equation(s) might be far from reality and may not take the load of the complexities that a real-life situation encounters. Accordingly, models are improved and bit by bit they got closer to a real-life situation. However, in the process, the model may become complicated and complex. For a model to have general acceptance, on the one hand, it shall be replicating the real-life situation for which it is modeled, and at the same time, it shall not be too complex in nature.

Very few works attempted quantification of the performance of wicket keepers. But the three major works that have done it successfully are that of Narayanan (2008), Lemmer (2011), and Hemachandran (2009). These works shall be read by anyone who wants to work on the quantification of wicket keepers. The method of performance measure of wicket keepers discussed in this section is inspired by all the aforesaid three works but does not match any one of them exactly. The measure defined here can be considered as an extension of their work.

The performance measure of the i^{th} wicket keeper is given by,

$$S_i = w_1 S_{i1} + w_2 S_{i2} \tag{3.26}$$

where S_{i1} is the performance score of the i^{th} wicket keeper for his batting ability and S_{i2} is the performance score for of the i^{th} wicket keeper for his wicket keeping skills, with w_1 and w_2 being the weights generally derived as variance stabilizers so that any one of the performance score does not get a chance to dominating the other for having more variance. The matter is discussed in detail in the subsequent sections.

3.5.1 Batting Performance of Wicket Keepers ($\mathrm{S}_{\mathrm{i}1}$)

Lemmer (2008) derived a technique that can convert the runs scored in a match to the adjusted runs based on the match condition and opposition's bowling strength. This technique is been used to convert the runs scored by a wicket keeper while batting into adjusted runs. The adjusted runs scored by the i^{th} player in the j^{th} match is denoted by T_{ij} and is defined by,

$$T_{ij} = R_{ij}\left(\mathrm{SR}_{ij}/\mathrm{MSR}_j\right)^{0.5} \tag{3.27}$$

where R_{ij} is the runs scored by the i^{th} batsman in the j^{th} match.

$$\mathrm{SR}_{ij} = \text{Strike Rate of the } i^{\text{th}} \text{ batsman in the } j^{\text{th}} \text{ match} = \frac{R_{ij}}{B_{ij}} \times 100 \tag{3.28}$$

where B_{ij} is the number of balls faced by the i^{th} batsman in the j^{th} match and strike rate of the j^{th} match (MSR_j) for all the batsmen of both teams

$$\mathrm{MSR}_j = \frac{\text{Total no. of runs scored in the match}}{\text{Total no. of balls bowled in the entire match}} \times 100 = \frac{R_j}{B_j} \times 100 \tag{3.29}$$

These adjusted runs are then used to define the batting performance measure S_{i1}. It often happens that in a match, batsman may remain not out because the innings might have got terminated because all the overs got exhausted or the team batting second might have got the runs necessary for victory or a batsman might have retired hurt or run short of partners. Thus, if we consider runs scored by a given batsman in different innings, some of which shall be complete innings and remaining few of them may be his not out scores. This is clearly a case of right truncated data. If a batsman getting out is considered as the event of interest and adjusted runs scored by the batsman in an innings replaces the time lapsed in survival analysis, then one can use the Kaplan–Meier estimate of the mean survival time (here mean adjusted runs) as the batting performance measure. So,

$$\mathrm{BP}_i = \text{Kaplan} - \text{Meier estimate of the mean adjusted runs scored by the } i^{\text{th}} \text{ batsman}$$

The mean survival time is estimated using the adjusted runs scored by a batsman in his complete (when the batsman gets out) and the incomplete innings (when the batsman remains not out).

The batting performance score (BP_i) of the i^{th} wicket keeper thus obtained is then standardized by the average value of BP across all the wicket keepers under consideration,

$$S_{i1} = \frac{\mathrm{BP}_i}{\underset{i}{\mathrm{Avg}}(\mathrm{BP}_i)} \tag{3.30}$$

3.5.2 *Keeping Skills of the Wicket Keeper* ($\mathbf{S_{i2}}$)

For measuring the performance of keeping skills of the wicket keeper two factors are considered. They are (i) dismissal rate and (ii) bye runs conceded (rate). According to Narayanan (2008), the dismissal rate of a wicket keeper is defined as the number of dismissals (stumping and catches) per match. Here, the term ‘match’ refers only to those matches where the player under consideration kept wickets for his team. We also counted the run outs in dismissals if it is indicated in the scorecard that the wicket keeper has implemented it.[1] Again, all the batsmen dismissed by the wicket keeper may not be of equal importance. In order to introduce the weight of the batsman dismissed by the wicket keeper, instead of only counting the number of dismissals, we have included the weights of the batsman using the concept of Lemmer (2005). In the said work, Lemmer weighted batsman based on their batting order. The weights of the batsman given their batting position are provided in Appendix 3.1.

[1] In case the score card indicates that a run out is implemented by the wicket keeper with the help of a fielder or fielders, then the dismissal is given a count of 0.5. But if the score card indicates that the run out is implemented by the wicket keeper alone, then the dismissal is given a count 1.

$$\text{Dismissal Rate } (D_i') = \frac{\text{Total wicket weights of batsmen dismissed by the } i^{\text{th}} \text{ keeper}}{\text{No. of matches in which } i^{\text{th}} \text{ player kept wickets}} \tag{3.31}$$

The rate in which bye runs were conceded is defined as

$$\text{Byes Rate } (B_i') = \frac{\text{Total bye runs conceded by the } i^{\text{th}} \text{ keeper}}{\text{No. of matches in which } i^{\text{th}} \text{ player kept wickets}} \tag{3.32}$$

It should be that while 'dismissal rate' has a positive dimension (i.e., positively associated with the skill of the players), but 'bye runs conceded' has a negative dimension (i.e., lesser the byes rate better is the wicket keeper). Thus, in place of using B_i' to bring parity between the two rates, $(1/B_i')$ is considered. However, to pool these two measures viz. D_i' and $(1/B_i')$ into a single measure, it is essential to standardize them and hence standardization is defined as

$$D_i = \frac{D_i'}{\underset{i}{\text{Avg}}(D')} \text{ and } B_i = \frac{1/B_i'}{\underset{i}{\text{Avg}}(1/B_i')}$$

Thus, $D_i \times B_i$ can be considered as a performance measurement of wicket keeper. However to ensure that dismissal rate and bye rate are comparable, an adjustment of B_i in terms of scale is necessary by raising a real number α to the exponent of B_i so that standard deviation of D_i and that of B_i^α is exactly the same. The value of α can be estimated using an iterative method (see Lemmer (2004) for details). So, $D_i \times B_i^\alpha$ can be considered as a performance measure of the wicket keeper, but this measure gives equal importance to both the factors, viz. dismissal rate and bye rate. Since a dismissal leads to loss of resources of the opponent team, so this shall get reasonably more importance compared to the bye runs conceded. Thus, the factor $D_i \times B_i^\alpha$ is reformulated as a weighted measure giving relative importance of the aforementioned two factors. This leads to the definition of S_{i2} as, somewhat following Lemmer (2011)

$$\begin{aligned} \text{WK}_i &= D_i^\beta \times (B_i^\alpha)^{1-\beta}, \quad 0 < \beta < 1 \\ &= \left(\frac{D_i}{\text{Avg}(D_i)}\right)^\beta \times \left(\left[\frac{1/B_i'}{\text{Avg}(1/B_i')}\right]^\alpha\right)^{1-\beta}, \quad 0 < \beta < 1 \end{aligned} \tag{3.33}$$

The value of β regulates the relative importance of the factors and acts as a stability between the dismissal measure and bye rate. The number of bye runs conceded by a wicket keeper also depends on the quality of bowling which is relatively less important than dismissals. Making the dismissal rate eight (8) times more important than saving bye runs, Narayanan (2008) allocated 5 points to bye runs conceded and 40 to dismissal rate. However, such type of allocation is utterly subjective. Even the author also did not find enough scientific basis for such allocation. In the absence of adequate literature and difference in expert opinion, it is difficult to converge an objective value of β. However, Pareto ordering for multi-objective evaluation is useful to find out that the value of β which has maximum compatibility. As conceding of bye runs is less important than the dismissal rate, the values of β are allowed to vary from 0.5 through to 0.9 with an increment of 0.05. A value of $\beta = 0.5$ gives equal importance to dismissal rate and bye rate and $\beta = 0.9$ makes dismissal rate nine times more important than the bye rate. The different values of β lead to alterations in the corresponding values of WK_i and so the ranks of the wicket keepers under consideration change extraordinarily. Thus, with Pareto ordering, we identify the set of ranks which has the maximum compatibility with all other rank sets. An explicit discussion of this method can be seen in Chakrabarty and Bhattacharjee (2012).

Let, β takes the values $\beta_1, \beta_2, \ldots, \beta_m$

$R_i^{\beta_j}$ = Rank of the i^{th} wicket keeper for a given value of $\beta = \beta_j$ (say)
$d_i^{\beta_j,\beta_p}$ = Square of difference between ranks of the i^{th} wicket keeper for $\beta = \beta_j$ and $\beta = \beta_p = \left(R_i^{\beta_j} - R_i^{\beta_p}\right)^2$
D_{β_j} = Overall distance of ranks when $\beta = \beta_j$ with other values of $\beta = \sum_{\substack{p=1 \\ p \neq j}}^{m} \left(R_i^{\beta_j} - R_i^{\beta_p}\right)^2$

Thus, the compatibility score corresponding to $\beta = \beta_j$ is given by $\overline{D}_{\beta_j}$ as defined in (3.34). The measure is an average distance of ranks of $\beta = \beta_j$ with other β values. Lesser the compatibility score of a given index more is the compatibility of that index with a set of similar other indices.

$$\overline{D}_{\beta_j} = \frac{D_{\beta_j}}{m - 1} \tag{3.34}$$

That value of β ($= \beta_j$ say) for which $\overline{D}_{\beta_j}$ is lowest has the maximum compatibility associated with the other values of β. Consequently, the most compatible value of β is estimated and is replaced in (8) for subsequent analysis. The most compatible value of β is dependent on the data values and shall not remain fixed. The wicket keeping performance score (WK_i) thus obtained is then standardized by the average value of WK across all keepers, i.e.,

$$S_{i2} = \frac{\text{WK}_i}{\underset{i}{\text{Avg}}(\text{WK}_i)} \tag{3.35}$$

3.5.3 Computation of the Weights for Combining S_{i1} *and* S_{i2}

The weights w_1 and w_2 as defined earlier represent the weights of the cricketers due to batting and wicket keeping skills deduced from the entire data set. These are variance stabilizing functions restricting the undue dominance of S_{i1} or S_{i2} in the composite index defined in (3.26) due to more variance

$$w_1 = \frac{C}{\sqrt{\operatorname*{Var}_i(S_{i1})}} \quad \text{and} \quad w_2 = \frac{C}{\sqrt{\operatorname*{Var}_i(S_{i2})}} \tag{3.36}$$

where $w_1 + w_2 = 1$ and C is a normalizing constant that follows:

$$C = \left[\frac{1}{\sqrt{\operatorname*{Var}_i(S_{i1})}} + \frac{1}{\sqrt{\operatorname*{Var}_i(S_{i2})}} \right]^{-1}$$

The choice of the weights in this manner would ensure that the large variation in any one of the factors would not unduly dominate the contribution of the rest of the factors (Iyenger and Sudarshan, 1982). Now, for a given wicket keeper (often referred as the i^{th} player) the values of S_{i1} and S_{i2} are computed using (3.30), (3.35) and the weights by (3.36). The values are then replaced in (3.26) to get the performance score of the i^{th} wicket keeper (i.e., S_i). The wicket keeper with the highest value of S_i is the best wicket keeper among the competing wicket keeper. The application of this method can be seen in Appendix 3.2.

3.6 Combined Performance Measures Based on All Cricketing Abilities

Considering the fact that all bowlers also bat, all-rounders are reasonably good batsman as well as bowlers and wicket keepers with considerably better batting skills are preferred, and the need was felt to combine all the different skills of a player into a single measure. This shall enable a comparison between cricketers with different expertise. To be more precise, such measures can help to compare a fast bowler with an all-rounder or a batsman with a wicket keeper. Conceiving this concept in mind, several authors forwarded different measures. Some such measures are as follows.

3.6.1 SS Index Developed by Suleman and Saeed (2009)

Suleman and Saeed (2009) developed a performance index called SS Index to quantify the performance of various abilities of the cricketers in Twenty20 cricket. It is a weighted combination of different factors that are related to the performances of the cricketers. All these factors have weighted according to their importance in Twenty20 cricket. The SS Index can be used to measure the performance of batsmen, bowler, all-rounder (i.e., batting and bowling) and wicket keeper in Twenty20 cricket. The various factors that are considered to measure the performance of batsmen, bowler, all-rounder, and wicket keeper are as follows:

For batsmen: Number of innings played, number of matches in which the player was not out, number of runs scored, batting average, strike rate, the fame of the player.
For bowler: Number of innings played, number of wickets taken, bowling average, bowling strike rate, economy rate, the fame of the player.
For all-rounder: All the factors that are considered for batsmen and bowler are considered for an all-rounder.
For wicket keeper: The number of innings played, the number of matches in which the player was not out, the number of runs scored, batting average, batting strike rate, the fame of the player, number of catches taken, number of stumping done.

3.6.1.1 Points Assigning Criteria for the Batsman

- 10 points are assigned if a batsman has played 25 innings or more. If batsmen played less than 25 innings, weight is calculated by

$$\frac{\text{Number of innings played}}{25} \times 10 \tag{3.37}$$

- 10 points are assigned if a batsman is not-out in 5 times or more. If he is not out for less than 5 times, then the weight is calculated proportionately.
- 10 points are assigned if a batsman has scored 750 runs or more. If he scores less than 750 runs, then the weight is calculated proportionately.
- 25 points are assigned if a batsman has a batting average 50 or more. If his batting average is less than 50, then the weight is calculated proportionately.
- 30 points are assigned if a batsman has strike rate 150 or more. If his strike rate is less than 150, then the weight is calculated proportionately.
- 15 or fewer points are assigned according to the fame of the batsmen.

3.6.1.2 Points Assigning Criteria for the Bowler

- 10 points are assigned if a bowler has played 25 innings or more. If a bowler played less than 25 innings, then the weight is calculated proportionately, i.e.,

$$\frac{\text{Number of innings played}}{25} \times 10 \tag{3.38}$$

- 25 points are assigned if a bowler takes 20 wickets or more. If a bowler takes less than 20 wickets, then the weight is calculated proportionately.
- 20 points are assigned if a bowler has an economy rate 6 or less. If his economy rate is more than 6, then the weight is calculated proportionately.

- 15 points are assigned if a bowler has bowling average 25 or less. If his bowling average is more than 25, then the weight is calculated proportionately.
- 15 points will be assigned if a bowler has bowling strike rate 20 or less. If his bowling strike rate is more than 20, then weight will be calculated proportionately.
- 15 or fewer points will be assigned according to the fame of the bowler.

3.6.1.3 Points Assigning Criteria for All-Rounder

The points to the all-rounder will be assigned similarly like the batsman and the bowler, but 50% weight will be given to the batting side and 50% to the bowling side.

3.6.1.4 Points Assigning Criteria for Wicket Keeper

- 35 points for wicket keeping, out of which 20 points are assigned if a wicket keeper has taken 15 catches or more and 15 points are assigned if a wicket keeper has done 5 stumpings or more. If a wicket keeper has taken less than 20 catches and has done less than 5 stumpings, then the weight is calculated proportionately.
- 10 points are assigned if a wicket keeper has played 25 innings or more. If a wicket keeper has played less than 25 innings, then the weight is calculated proportionately.
- 5 points are assigned if a wicket keeper remains not out in 5 times or more. If a wicket keeper remains not out in less than 5 times, then the weight is calculated proportionately.
- 10 points are assigned if a wicket keeper has scored 750 runs or more. If he scored less than 750 runs, then the weight is calculated proportionately.
- 10 points are assigned if a wicket keeper has a batting average of 50 or more. If a wicket keeper has a batting average less than 50, then the weight is calculated proportionately.
- 15 points are assigned if a wicket keeper has strike rate 150 or more. If a wicket keeper has strike rate less than 150, then the weight is calculated proportionately.
- 15 or fewer points are assigned according to the fame of the wicket keeper (Table 3.1).

The measure has a lot of subjectivity, and there is no proper reasoning behind such subjective values.

Table 3.1 Subjective weights for the SS index

Innings	Not out	Runs scored	Batting average	Strike rate	Fame		Total
10	10	10	25	30	15		100
Innings	Wickets taken	Economy rate	Bowling average	Bowling strike rate	Fame		Total
10	25	20	15	15	15		100
Innings	Not out	Runs scored	Batting average	Strike rate	Fame	Wicket keeping	Total
10	5	10	10	15	15	35	100

3.6.2 Performance Indices Developed by Gerber and Sharp (2006)

Gerber and Sharp (2006) proposed that measures are developed to give an indication of each player's batting, bowling, fielding, all-rounding, and wicket keeping abilities. The relevant measure would then be applied to each player in his area of expertise. The method proposed by Gerber and Sharp (2006) to measure the batting ability of a player (i.e., the batting index) was defined as

$$\mathrm{BT}_i^{\mathrm{GS}} = \left(\frac{\text{Batting average of } i^{\text{th}} \text{ batsman}}{\text{Sum of all the batting averages of all the batsmen}}\right) \times \text{Number of specialist batsmen} \quad (3.39)$$

which compared the batting average of the given batsman to that of all the other batsmen.

For bowlers, Gerber and Sharp (2006) developed a performance index which is defined as

$$\mathrm{BL}_i^{\mathrm{GS}} = \left(\frac{v_i}{\text{Sum of all } v_i}\right) \times \text{Number of specialist bowlers} \quad (3.40)$$

where $v_i = \left[k - \left(\frac{\text{Economy rate of specialist } i^{\text{th}} \text{ bowler}}{\text{Sum of all the economy rates of all the specialist bowler}}\right)\right]$ and where the constant k is chosen to be the smallest positive integer such that

$$k > \left(\frac{\text{Economy rate of specialist } i^{\text{th}} \text{ bowler}}{\text{Sum of all the economy rates of all the specialist bowler}}\right)$$

According to Gerber and Sharp (2006), introducing this constant ensured that more desirable bowlers (i.e., bowlers with low economy rates) had higher values of v_i.

Gerber and Sharp (2006), measured the fielding ability of i^{th} player using a fielding index and it is defined as

$$\text{FLD}_i^{\text{GS}} = \left(\frac{\text{Dismissal rate by the } i^{\text{th}} \text{ player}}{\text{Sum of the dismissal rates of all the fielders}}\right) \times \text{Number of specialist fielders} \tag{3.41}$$

where the dismissal rate of player is defined as the average number of fielding dismissals (i.e., run outs, catches) made by the player per match.

All-rounders can be defined as players who perform well in different roles in the team. For example, a player who can both bat and bowl is considered as an all-rounder. Gerber and Sharp (2006) suggested that all-rounders can be divided into four categories, viz. cricketers who can both bat and bowl, cricketers who can both bat and field, cricketers who can both bowl and field, cricketers who can bat, bowl, and field. However, the category of an all-rounder that had been of interest in their study was cricketers who can both bat and bowl. The all-rounder performance ability for each was defined as

$$\text{ALR}_i^{\text{GS}} = \left(\frac{\text{Sum of } i^{\text{th}} \text{ player's index values}}{\text{Sum of that category of index values}}\right) \times \text{Number of players in all-rounder category} \tag{3.42}$$

The wicket keeping ability by Gerber and Sharp (2006) was defined as

$$\text{FLD}_i^{\text{GS}} = \left(\frac{\text{Dismissal rate by the } i^{\text{th}} \text{ wicket keeper}}{\text{Sum of the dismissal rates of all the wicket keepers}}\right) \times \text{Number of specialist wicket keeper} \tag{3.43}$$

3.6.3 A Performance Measure Defined by Beaudoin and Swartz (2003)

Beaudoin and Swartz (2003) define a new statistic, the runs per match for a cricketer, which is applicable to measure the performance of both batsmen and bowlers in ODI. It is defined as

$$\text{RM} = 100 \times \left(\frac{\text{Total number of runs}}{\text{Total resources used}}\right) \tag{3.44}$$

where the totals are taken over all the cricketers appearances and total resources should be a combination of the number of overs completed and the number of wickets taken during a cricketer's tenure as either a batsman or bowler. According to Beaudoin and Swartz (2003), this definition probably is the best way to measure a player's performance as a weak batsman (bowler) can not have a high (low) value of RM. Duckworth and Lewis (1998) stated that an average of 233.1 runs was scored with standard error 3.0 in ODI matches from mid-1997 through early 2002. One can use this piece of evidence as a standard for determining the excellence of batsmen and bowlers. Obviously, a good bowler has an RM value below 233 and a good batsman has an RM value much higher than 233.

Consider a cricketer (either a batsman or bowler) for whom x_i runs and r_i resources are recorded in the i^{th} appearance ($i = 1, 2, \ldots, n$). Now let

$$T = \frac{\mathrm{RM}}{100} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} r_i} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i}{\frac{1}{n}\sum_{i=1}^{n} r_i} \tag{3.45}$$

From the multivariate central limit theorem, the approximate result would be

$$\begin{pmatrix} \frac{1}{n}\sum\limits_{i=1}^{n} x_i \\ \frac{1}{n}\sum\limits_{i=1}^{n} r_i \end{pmatrix} \sim \text{Normal}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \ \frac{1}{n}\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right] \tag{3.46}$$

Using a Taylor series expansion and an application of the Delta theorem

$$T = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i}{\frac{1}{n}\sum_{i=1}^{n} r_i} \approx \frac{\mu_1}{\mu_2} + \begin{pmatrix} \frac{1}{\mu_2} & -\frac{\mu_1}{\mu_2^2} \end{pmatrix}\begin{pmatrix} t_1 - \mu_1 \\ t_2 - \mu_2 \end{pmatrix} \tag{3.47}$$

$$\mathrm{Var}\left(T = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i}{\frac{1}{n}\sum_{i=1}^{n} r_i}\right) = \frac{1}{n}\begin{pmatrix} \frac{1}{\mu_2} & -\frac{\mu_1}{\mu_2^2} \end{pmatrix}\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\begin{pmatrix} \frac{1}{\mu_2} \\ -\frac{\mu_1}{\mu_2^2} \end{pmatrix} \tag{3.48}$$

In order to get Var(T), Beaudin and Swartz (2003) estimate each of those quantities with consistent estimators.

$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the consistent estimator of μ_1;
$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} r_i$ is the consistent estimator of μ_2;
$\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{n-1}{n}\sigma_1^2$ is the consistent estimator of σ_1^2;
$\hat{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n} (r_i - \bar{r})^2 = \frac{n-1}{n}\sigma_2^2$ is the consistent estimator of σ_2^2;
$\hat{\sigma}_{12} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(r_i - \bar{r}) = \frac{n-1}{n}\sigma_{12}$ is the consistent estimator of σ_{12};

where the fact is that $Cov\ (x_i, r_j) = 0$ if $i \neq j$ (i.e., runs scored in the i^{th} appearance are independent of resources used in the j^{th} appearance).

Finally, $T = \frac{\mathrm{RM}}{100}$ so

$$\mathrm{Var}(T) = \frac{1000}{n}\begin{pmatrix} \frac{1}{\bar{r}} & -\frac{\bar{x}}{\bar{r}^2} \end{pmatrix}\begin{pmatrix} \frac{n-1}{n}S_x^2 & \frac{n-1}{n}S_{xr} \\ \frac{n-1}{n}S_{xr} & \frac{n-1}{n}S_r^2 \end{pmatrix}\begin{pmatrix} \frac{1}{\bar{r}} \\ -\frac{\bar{x}}{\bar{r}^2} \end{pmatrix} \tag{3.49}$$

where S_x^2 is the sample variance of the runs scored, S_r^2 is the sample variance of the resources used, and S_{xr} is the sample covariance between these two variables.

3.6.4 A Performance Measure Developed by Lewis (2005)

Traditional measures of performance for batting and bowling in cricket are not appropriate (Lewis, 2005). For example, over a series of matches comparing a batting average of 63 with a bowling average of 23 for a given player cannot be undertaken objectively. Moreover, these traditional performance measures cannot be aggregated to measuring a player's performance in batting and bowling for a match or over a series of matches. Thus, Lewis (2005) proposed a performance measure for cricketers using ball-by-ball information of the match based on well-established Duckworth/Lewis method. The Duckworth/Lewis model for one-day cricket is

$$Z(u, w) = Z_0 F(w)\left\{1 - \exp\left[-\frac{bu}{F(w)}\right]\right\} \tag{3.50}$$

where $Z(u, w)$ represents the average further runs obtained in the u remaining overs when w wickets have been lost and Z_0 and b are positive constants. $F(w)$ is a positive decreasing step function with $F(0) = 1$. This function is interpreted as the proportion of runs that are scored with w wickets lost compared with that no wicket lost. Mathematically, it can be defined as

$$F(w) = \frac{Z(u, w)}{Z(u, 0)} \tag{3.51}$$

In cricket, since there are six balls in every over so $i = 6u$, $(0 \le i \le 300)$ will represent the i^{th} remaining ball of the innings. Now at any stage of an innings, if there are i balls remaining and w wickets have been lost then the expected runs (i.e., r_i) from next ball will be either

$$r_i = Z(i, w) - Z(i-1, w) \tag{3.52}$$

$$\text{or} \quad r_i = Z(i, w) - Z(i-1, w+1) \tag{3.53}$$

depending on whether the batsman survives on that ball or loses his wicket, respectively. If the batsman scores s_i runs from that ball, then using (3.52) and (3.53), his net contribution c_i for that ball is either

$$c_i = s_i - \{Z(i, w) - Z(i-1, w)\} \tag{3.54}$$

$$c_i = s_i - \{Z(i, w) - Z(i-1, w+1)\} \tag{3.55}$$

depending on whether or not the batsman survives that ball. According to Eq. (3.50), in early of a team's innings, a single run will make a positive contribution but toward the end of an innings with wickets in hand, and a single run will result in a negative contribution as more than a run a ball is then expected. However, there are several forms of extras in cricket. The bye and leg bye count are neither for the batsman nor against the bowler yet such runs scored, are aggregated with other runs to form the tota; score of the team. Similarly, the wide ball and no ball are debited against the bowler, added to the batting side score, but not credited to the batsman facing at the time. Thus, there needs to be slight adjustment to contributions of batsman and bowlers for 'extras,' in order to correct the total net contributions of completed innings is indeed zero.

3.6.4.1 Adjusting for Extras

Let e_i $(1 \leq i \leq 300)$ be the number of extras conceded by the i^{th} ball remaining. Again, let g_i be the merged number of byes and leg byes and h_i be the merged number of wides and no balls. Now, $e_i = g_i + h_i$ and the total of the teams is defined as $S = \sum s_i + \sum e_i = \sum s_i + \sum g_i + \sum h_i$, and the summations are pertinent over all 300 balls of the innings. For a normal ODI, the total score of the team batting first is considered as $Z(50{,}0) = 235$. Now any batting team has exceeded expectations from the D/L method, if $S > 235$ and conversely if $S < 235$. If there were $\sum e_i$ extras given in an innings and if the sum of the batsmen's contributions exceeds $Z(50{,}0) - \sum e_i$ (i.e., $235 - \sum e_i$), then the batsmen have performed above average. In this instance, the contributions of batsmen adjusted for the effects of extras for the i^{th} ball remaining are achieved by

$$c_i = (s_i - r_i) \times \left(\frac{235 - \sum e_i}{235} \right) \tag{3.56}$$

As the total average runs (i.e., $\sum r_i$) in ODI is 235, it is certainly understood that $\sum c_i = S - 235$. Except the extra bowled by bowler consider as forfeit, the net contribution for the bowler is the negative of that for the batsman facing the same ball. The extras are not including against the bowler in the innings represented by $\sum g_i$. Therefore, the net contribution of a bowler on any ball which is adjusted for the effect of extras is defined by

$$\bar{c}_i = r_i \times \left(\frac{235 - \sum g_i}{235 - (s_i + h_i)} \right) \tag{3.57}$$

It is simply observed that the total of the net contributions for the bowlers over the 300 balls of the innings is $\sum \bar{c}_i = 235 - S$. Consequently, $\sum c_i + \sum \bar{c}_i = 0$ so that after taking different efficacies of extras between batsman and bowlers, the zero-sum nature of the game is captured. However, within 300 balls if the team batting second has won the match (i.e., if an innings is not completed) then $\sum c_i + \sum \bar{c}_i \neq 0$. This difference is ignored by the D/L method because the effect is likely to be negligible.

3.6.4.2 Average Runs for Resources Consumed and Resources Contributed

The D/L methodology changes the $Z(u, w)$ in Eq. (3.50) into the proportion of the average total in 50 overs as

$$P(u, w) = \frac{Z(u, w)}{Z(50, 0)} = \frac{Z(u, w)}{235} \tag{3.58}$$

This function is interpreted as the percentage of resources that a team has remaining for u overs it has left when w wickets have been lost. Again, the consumption of resources p_i for the i^{th} ball (where $i = 6u$) remaining is represented by either

$$p_i = P(i, w) - P(i-1, w) \tag{3.59}$$

$$\text{or } p_i = P(i, w) - P(i-1, w+1) \tag{3.60}$$

depending on whether or not the batsman survives on that ball respectively. As mentioned earlier, resources are neither consumed by the batsman nor contributed by the bowler yet runs made by the batsman off the no ball are credited to the batsman, but these forms of extras are debited against the bowler.

3.6.4.3 Batting and Bowling Performance by Quantifying Resource Utilization

Suppose for any ball i, a batsman scores s_i runs and consumes p_i resources, then his average run contribution per unit of resource consumed to the team's total can be evaluated by $\sum s_i / \sum p_i$, where the sums are over the number of balls that the batsman faced. Again, the bowler's performances are evaluated using the total number of runs, no balls, and wides that are conceded. A bowler's average run contribution per unit of resource contributed is measured by $\sum (s_i + h_i) / \sum p_i$, where the sums are over the number of balls the bowler has bowled. Thus, the higher resource average indicates better batting performance for the batsman and for the bowler, lower resource average indicates the better performance.

Though this method capable of considering different match situations to measure the performance of bowler and batsman, it is computationally tough to implement. Even if we overlook the computational aspect, one has to collect ball-by-ball information of the match. It is a tedious task to do and performance of the cricketers cannot be measured in the absence of such information. Someone has to watch each and every ball of a match or all the matches of an entire series to collect the ball-by-ball information. These days, a few Websites are providing ball-by-ball information of the match in descriptive forms (e.g., *espncricinfo.com*, *cricwaves.com*.). So, one has to read the information and note down in quantitative forms for necessary calculation. Also, this ball-by-ball information of a match provided in *HTML* sites is not adequate all the time to quantify the performance of a player. Thus, it would be helpful if the scorecard of a match can be used for quantifying the performance of the players. This has lead to the development of a new performance measure for cricket in this chapter.

3.6.5 A New Measure to Quantify the Performance of Cricketers

Though there are several measures for quantifying or accumulating the batting or bowling or wicket keeping the performance of a match, the need for a combined measure is felt in different situations. Through such combined measure, cricketers with different expertise shall be compared to each other. Namely, a wicket keeper can be compared with an all-rounder or the performance of a spinner can be compared with the performance of an opening batsman. Such a measure can be helpful in the grading of cricketers. It is a common practice to determine the salary of cricketers by the cricket boards. In every cricket match, an award named '*man of the match*' is given to a player for his remarkable performance on that match (Lewis, 2005). The organizers, as well as decision makers, require quantifiable on-field information to assess the performance of that player. As each game of cricket generates a huge amount of performance-related statistics, the on-field performance of a player can be obtained through a scorecard of a match. A match scorecard only provides information of traditional performance statistics. The batting average and strike rate are usually used to understand the performance of batsmen. Conversely, bowling average, economy rate, and bowling strike rate are used to measure the performance of bowlers. However, the aforesaid statistics have severe drawbacks in judging the true skills of a players' performance (Lewis, 2005). Moreover, the different traditional performance statistics are in different units of measurement; so to evaluate a player's all-around performance, it is very difficult to merge them. All these drawbacks to measure the performance of the cricketers are well discussed by Lewis (2005).

Barr and Kantor (2004) and Lemmer (2008) developed performance measures for batsmen by combining the traditional measures of performances. But both the measures cannot be used to measure the performance of all-rounders. Thus, Lewis (2005) proposed an alternative measure of player performance to overcome this limitation on the basis of the well-established Duckworth/Lewis method as explained in the previous section which can be used not only to measure the performances of batsmen and bowlers, but it can also be used to measure the performances of all-rounders. But it has its own limitations as discussed. Thus, the current measure is proposed. This performance measure can be used to quantify the performance of batsman, bowler, fielder, all-rounder, and wicket keeper simultaneously.

3.6.5.1 The Performance Measure

The performance measure of the i^{th} player is given by

$$S_i = S_{i1} + S_{i2} + \delta_i \tag{3.61}$$

where

$$\delta_i = \begin{cases} S_{i3}^{a_i} + S_{i4}^{1-a_i} - 1, & \text{if } i^{\text{th}} \text{ player is either a bowler or wicket keeper} \\ 0, & \text{if } i^{\text{th}} \text{ player is neither a bowler nor wicket keeeper} \end{cases}$$

where a_i is an indicator variable with,

$$a_i = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ player is a bowler} \\ 0, & \text{if } i^{\text{th}} \text{ player is a wicket keeper} \end{cases}$$

with

S_{i1} = Batting performance score;
S_{i2} = Fielding performance score;
S_{i3} = Bowling performance score;
S_{i4} = Wicket keeping performance score.

The factors considered under each of the abilities (viz. batting, fielding, bowling, and wicket keeping) and the computation of performance indicators for each ability are explained in detail in the subsequent sections.

3.6.5.2 Factors Considered for Batting Performance Measure

A number of factors are considered to measure the batting performance of cricketers. These factors are batting average, strike rate, and average percentage of contribution to the team total. The average percentage of contribution to the team total is defined by

$$\text{APC}_i = \underset{j}{\text{Average}}\left(\frac{\text{Runs scored by the } i^{\text{th}} \text{ player in the } j^{\text{th}} \text{ innings}}{\text{Team's total score in the } j^{\text{th}} \text{ innings}} \times 100\right)$$

At first, the values of these factors for each player are normalized and then weights are calculated based on their relative importance. All the normalized scores of these factors are multiplied by their corresponding weights and then added to get S_{i1}.

3.6.5.3 Factors Considered for Fielding Performance Measure

Like batting, a number of catches taken by the player as a fielder and number of run outs caused by the player in a match or series of matches are considered to quantify the fielding performance. As we know that only these two factors are available in the scorecard of the match, only these two factors are considered to quantify the fielding performance of cricketers. The values of these two factors for each player are normalized, and then, the weights are calculated based on their relative importance. Finally, all the normalized scores for these factors are multiplied by their corresponding weights and then added to get S_{i2}.

3.6.5.4 Factors Considered for Bowling Performance Measure

For measuring the bowling performance, the factors are bowling average, economy rate, and strike rate of the bowler. Thereafter, the values of these factors are normalized for each player and then weights are calculated based on their relative importance. As mentioned earlier, the normalized scores for these factors are multiplied by their corresponding weights and then added to get S_{i3}.

3.6.5.5 Factors Considered for Performance Measure of Wicket Keeping

For measuring wicket keeping performance, the factors are the number of catches taken per match, the number of stumping per match, and the number of bye runs conceded per match. The phrase 'per match' signifies the number of matches when the player kept the wickets for his team. It is because some of the teams have more than one player in playing eleven who are capable of wicket keeping. Likewise, the values of these factors for each player are normalized and then weights of are calculated. All the normalized scores for the factors are multiplied by their corresponding weights and then added to get S_{i4}.

3.6.5.6 Normalization

Considering the abilities of batting, fielding, bowling and wicket keeping; the performance measure defined in Eq. (3.61) is a linear combination of traditional performance statistics. According to Lewis (2005), the traditional performance statistics of batting and bowling abilities cannot be combined as they are based on incompatible scales. To overcome this limitation, normalization is essential corresponding to each of the factors mentioned earlier. This normalization helps to remove the unit of measurement and variability effect of all the traditional performance statistics. Also on the basis of normalization, under the abilities of batting, fielding, bowling, and wicket keeping, the traditional performance statistics come within a range of 0–1.

Since normalization makes the statistics unit free, they can be combined through addition.

Let X_{ijk} be the value of the j^{th} factor (i.e., batting average, strike rate) for the i^{th} player in the k^{th} ability (i.e., batting, fielding, bowling, and wicket keeping). Among the factors considered for performance measurement, some are having positive dimension (e.g., batting average, batting strike rate, and number of stumping). That means they are directly related to the abilities of the player. While some of the factors (e.g., economy rate, number of bye runs conceded) have negative dimension, they are negatively related to the abilities of the player. Now, if the factor represents a positive dimension associated with the abilities of the player, then it is normalized as

$$Y_{ijk} = \frac{X_{ijk} - \min\left(X_{ijk}\right)}{\max\left(X_{ijk}\right) - \min\left(X_{ijk}\right)} \tag{3.62}$$

and if the factor represents a negative dimension related to the abilities of the player, then it is normalized as

$$Y_{ijk} = \frac{\max(X_{ijk}) - X_{ijk}}{\max(X_{ijk}) - \min(X_{ijk})} \tag{3.63}$$

3.6.5.7 Determination of Weights

We know that a simple average always provides equal importance to the factors or variables are involved in the computation of the average. On the other hand, a composite index is a weighted measure where the relative importance of the factors or variables is considered. According to Iyenger and Sudarshan (1982), the weights are varied inversely proportional to the variation in the respective variables. This method has been adopted in this study to determine the weights of the factors that are associated with the different abilities of the cricketers.

Let Y_{ijk} be the normalized value of the i^{th} players for the j^{th} factors of the k^{th} abilities where $i\,(= 1, 2, \ldots, n)$ represents players; $j\,(= 1, 2, 3, 4)$ for the four different factors considering under each of the k^{th} ability (e.g., under batting ability different factors are number of innings, strike rate, batting average, average percentage of contribution to the team total); k (= 1, 2, 3 4) represents the different abilities, viz. batting (1), fielding (2), bowling (3), and wicket keeping (4). Now, if w_{jk} represents the weight of the j^{th} factor under the k^{th} abilities, then it is calculated as

$$w_{jk} = \frac{C_k}{\sqrt{\underset{i}{\operatorname{Var}}\left(Y_{ijk}\right)}}, \quad j = 1, 2, 3, 4 \text{ and } k = 1, 3, 4 \tag{3.64}$$

where $\sum_{j=1}^{4} w_{jk} = 1$ for $k = 1$, 3, and 4 and C_k is a normalizing constant that follows

$$C_k = \left[\sum_{j=1}^{4} \frac{1}{\sqrt{\underset{i}{Var}(Y_{ijk})}} \right]^{-1}$$

Since to quantify fielding performance of the cricketers, the defined performance measure has considered only two factors under fielding ability (viz. number of catches taken and run outs), so if w_{jk} represents the weight of the j^{th} factor under the fielding ability, then it is calculated as

$$w_{jk} = \frac{C_k}{\sqrt{\underset{i}{\mathrm{Var}}\left(Y_{ijk}\right)}}, \quad j = 1, 2 \text{ and } k = 2 \tag{3.65}$$

where $\sum_{j=1}^{2} w_{jk} = 1$ for $k = 2$ and C_k is a normalizing constant that follows

$$C_k = \left[\sum_{j=1}^{2} \frac{1}{\sqrt{\underset{i}{\mathrm{Var}}(Y_{ijk})}} \right]^{-1}$$

The choice of the weights in this way ensures that large variation in any one of the factor would not excessively dominate the contribution of the rest of the factors (Iyenger & Sudarshan, 1982).

3.6.5.8 Computation of Performance Score

The performance scores for batting ($k = 1$), fielding ($k = 2$), bowling ($k = 3$) and wicket keeping ($k = 4$) of the i^{th} player are calculated by

$$S_{ik} = \sum_{j=1}^{n_k} w_{jk} Y_{ijk} \tag{3.66}$$

where $n_k = 3$ for $k = 1$, 3, and 4 and n_2=2

Now, on obtaining the values of S_{i1}, S_{i2} S_{i3} and S_{i4} the performance score S_i of the i^{th} player is computed using Eq. (3.61). The performance score of all the players is computed and then converted into the corresponding performance index (P_i). The performance index of the i^{th} player is denoted by P_i and is given by

$$P_i = \frac{S_i}{\underset{i}{\max}(S_i)} \tag{3.67}$$

The performance index value P_{i} for a player is always lying between 0 and 1 (i.e., $0 < P_i \leq 1$). Higher value of the performance index indicates better performances of the players. The detail application of this new measure to quantify the overall performance of the cricketers can be seen in Appendix 3.3.

3.7 Performance Prediction of Batsmen and Bowlers Using Neural Network

In recent times, the application of data mining techniques in the field of sports has evolved. These data mining tools and techniques help us to explore and analyze the sports data in order to predict the tournament result, performance modeling, classification, etc. The different data mining techniques that are commonly used in various sports are decision tree, artificial neural network, clustering, genetic algorithm, etc. Among these tools, the artificial neural network is one of such technique that is frequently used in practice. The different concepts and approaches of artificial neural networks in different sports were discussed by Perl (2001). The importance of pattern recognition in sports was discussed by Perl and Weber (2004) using a neural network. Some other applications of the artificial neural network in different sports are modeling swimming performance of swimmers by Silva et al. (2007) and Young and Weckman (2008) used a neural network to predict the performance of football players by translating players' rating values to National Football League (NFL) combined values. Using the Duckworth/Lewis (DL) method in the game of cricket, Bailey and Clarke (2006) used a neural network to predict the outcome in one-day international cricket matches while the game is in progress. Likewise, Choudhury, Bhargava, Reena, and Kain (2007) used an artificial neural network for predicting the outcome of cricket tournaments. Moreover, with artificial neural network, Saikia, Bhattacharjee, and Lemmer (2012) predicted the performance of bowlers using data from Indian Premier League (IPL).

The most important characteristic of an artificial neural network (ANN) is its ability to learn. The process of ANN is designed to incorporate key features of neurons in the human brain and to process the data in a manner analogous to the human brain. Both the human brain and neural network depend heavily on being able to acquire knowledge from events that had happened in the past and to apply this knowledge in future circumstances. This logic is the basic foundation of artificial neural network. Getting knowledge through the learning rule is a procedure for modifying the weights (or it may be called parameters) and biases of a network. The purpose of the learning rule is to train the networks to perform some task. There are many types of neural network learning rules; however, all of them fall into one of the three broad categories, viz. supervised learning, unsupervised learning and reinforcement (or graded) learning. Interested readers can see the work of Abraham (2005) for an extended discussion of different learning rules of the neural network.

Now let us consider the relation $y = f(x)$ which describes the functional form of the basic neuron model, where $f(x) = w_1x_1 + w_2x_2 + \cdots + w_nx_n + \in$. More conveniently, it can be written as

$$y = \sum_{i=1}^{n} w_i x_i + \in \tag{3.68}$$

In Eq. (3.68), containing a set of n inputs x_{i}, each input is multiplied with an associated weight w_{i} before it is applied to the process and $\in$ is a bias term. Though the above functional form looks to be linear, it may or may not be true. However, if a linear relationship between the response variable (y) and predictors (i.e., x_i's) is appropriate, then the result of the neural network should closely approximate to a linear regression model. In fact, one can argue that linear regression is a special case of the neural network. The classical linear regression model can acquire knowledge based on data through the least squares method and store that knowledge in the form of regression coefficients (Norusis, 2007). In this sense, linear regression is a type of neural network. However, classical linear regression assumes that the functional relationship between response and predictor variables is linear. Moreover, it imposes some set of assumptions before learning the data. Conversely, an artificial neural network does not require any such set of assumptions and linear relationship between the response variable and predictors before learning the data. Usually, the form of relationship between the response variable and predictors in a neural network is determined during the learning process. Therefore, it would always be better, if we consider that we have some inputs $x_1, x_2, \ldots, x_{\mathrm{n}}$ and the desired output is represented by the response variable 'y' instead of thinking about the functional form of the model structure. In the literature, there are so many types of artificial neural networks, viz. multilayer perceptron (MLP) and, radial basis function (RBF).

In cricket, to predict the performance of batsmen and bowlers using an artificial neural network, one has to choose a response and predictor variables. Here, response variable may be the performance of the batsmen or bowlers, and of course, the predictors will be different. For batsmen, the predictors or input variables that are supposed to influence the batting performance are age, number of international matches played, batting average in ODI, strike rate in ODI, Twenty20 matches played, batting average in Twenty20, strike rate in Twenty20, experience in international cricket, and batting hand (left/right). Similarly, for bowlers, the predictors are age, number of international matches played, bowling average in ODI, economy rate in ODI, bowling strike rate in ODI, bowling average in Twenty20, economy rate in Twenty20, Bowling strike rate in Twenty20, experience in international cricket, and bowling hand (left/right). Note that in an artificial neural network, the response variable may be continuous (i.e., scale) or categorical (i.e., nominal or ordinal). Unlike the regression model, an artificial neural network can handle more than one dependent variable. Nowadays, different statistical software is available in the market, viz. Statistical Package for Social Science (SPSS), R-Language, and MATLAB. All these software can gainfully be utilized to perform artificial neural network analysis based on real-life sports data.

3.8 A Bayesian Approach to Classify the All-Rounders

To become a cricketer, the batting and bowling are undoubtedly the prime skills in the game of cricket at all times. Every balanced cricket team has specialist batsmen and bowlers. Yet, players with reasonably good at both bat and ball are always vital assets to a team. Such players are recognized as all-rounders. They are capable to make a place in the team either his batting or bowling (Bailey, 1989). Players like Andrew Flintoff, Yuvraj Singh, Ravindra Jadeja, and Jack Kallis are good enough to be selected both as a batsman and as a bowler occupying only one place in the team. The all-around performance of those players played an important role for the team. The respective captain of a team could utilize them merely as a batsman or a bowler whenever he liked during a match. Such players are often referred to as utility players. They have a chance to perform with the ball in case they fail to bring achievements to their team with their willow and vice versa. In actual fact, an all-rounder is better at batting than bowling or vice versa. Very few are equally good at both and hardly any has been outstanding at both. A genuine all-rounder is a strength to the team, but in the absence of such all-rounders, the responsibility of all-rounders is shared by batting all-rounder(s) or/and bowling all-rounder(s).

To classify the all-rounders for any given tournament, one can use a naïve Bayesian classification model. However, before applying naïve Bayesian classifier we need some essential information from the training sample. This training sample consists of a sample of all-rounders who had been played in any previous well competitive tournament. Since all-rounders are equally good in both bat and ball, two significant factors, viz. strike rate and economy rate, can be considered to classify the all-rounders using training sample. The strike rate is a directly measurable quantity (i.e., greater the strike rate more effective the batsman), and on the other hand, the economy rate is an inverse quantity (i.e., smaller the economy rate better as the bowler). Ideally, an all-rounder is supposed to have a high strike rate and a low economy rate. Now considering median (divides a data set into exactly two equal parts) as a threshold value for both the factors, one can define the following four classes using a scatter plot by strike rate in the x-axis and economy rate in the y-axis.

1. An all-rounder with a strike rate above the median and economy rate below median is a *performer*, which is a genuine all-rounder.
2. An all-rounder with a strike rate above the median and economy rate above the median is a *batting all-rounder*.
3. An all-rounder with a strike rate below median and economy rate below median is a *bowling all-rounder*.
4. An all-rounder with a strike rate below the median and economy rate above the median is recognized as *under performer*, which is the worst case.

Several independent variables that are supposed to influence the above-mentioned classification are considered, viz. age, strike rate in ODI, economy rate in ODI, international cricket experience, batting hand, bowling type (spin/fast), batting average in Twenty20, and country. Some of the variables are nominal, some are may be discrete, and others are continuous. Thereafter, stepwise logistic regression could be used to identify the significant variables for the performance of the all-rounders. The dependent variable for stepwise logistic regression will be different classes of said all-rounders.

Now, based on the significant variables identified through stepwise logistic regression (Saikia & Bhattacharjee, 2011), one can use naïve Bayesian classification model to predict the expected class of an incumbent all-rounder who will be played in an upcoming tournament or playing the ongoing tournament.

On the basis of conventional Bayes theorem, a well-known probabilistic classifier is the naïve Bayesian classifier. This classifier can predict the class membership probabilities in such a way that the probability of a given subject belongs to a particular class. According to Zhang (2004), whether it is continuous or categorical, it can manage an arbitrary number of independent variables. It assumes that the effect of a variable value on a given class is independent of the values of the other variables. This assumption is called as class conditional independence (Flach & Lachiche, 2004).

Although in reality, the assumption of class conditional independence is not always accurate, still naïve Bayes classifier gives surprisingly good results. The reason is that its classification accuracy does not depend on the dependencies that may exist among the attributes (Domingos & Pazzani, 1997). It ignores the interaction between variables within individuals of the same class. Another important advantage of the naïve Bayes classifier is that it requires a small size of training data to estimate the parameters necessary for classification. The step-by-step procedure of how this naïve Bayesian classifier works is given below.

A sample vector $\boldsymbol{X}$ is classified to a class C_i if and only if

$$P(C_i|\boldsymbol{X}) > P\left(C_j|\boldsymbol{X}\right)\text{for } 1 \le j \le m \text{ and } j \ne i \tag{3.69}$$

From Bayes theorem, we can get an expression for P(C_i|$\boldsymbol{X}$)

$$P(C_i|\mathbf{X}) = \frac{P(\mathbf{X}|C_i)P(C_i)}{\sum P(\mathbf{X}|C_i)P(C_i)} \tag{3.70}$$

The class prior probability would be estimated by

$$P(C_i) = \frac{s_i}{s} \tag{3.71}$$

where s_i is the number of training samples of class C_i, and s is the total number of training samples.

The attributes are assumed to be independent of each other. Then, one can write

$$P(\mathbf{X}|C_i)P(C_i) = \prod_{k=1}^{n} P(X_k|C_i)P(C_i) \tag{3.72}$$

Now, the probabilities $P(X_k|C_i)$ ($k = 1, 2,\ldots, n$) can be estimated from the training sample in the following manner:

(a) If the k^{th} attribute A_k is categorical, then

$$P(X_k|C_i) = \frac{s_{ik}}{s_i} \tag{3.73}$$

where s_{ik} in the number of training sample points of class C_i for which the k^{th} attribute attains the value X_k.

(b) However, if the k^{th} attribute A_k, is continuous, then $P(X_k|C_i) \sim N(\mu_{c_i}, \sigma_{c_i})$, so that

$$p(x_k|c_i) = \frac{1}{\sqrt{2\pi}\sigma_{c_i}} \mathrm{e}^{-\frac{\left(x_k-\mu_{c_i}\right)^2}{2\sigma_{c_i}^2}} \tag{3.74}$$

where μ_{c_i}, σ_{c_i} are the mean and standard deviation of the values of the attribute A_k belonging to the i^{th} class. However, before setting up the distribution it is obligatory to test such a claim. In case of non-normality, the appropriate distribution be identified and its parameters be estimated from the data provided by the training sample and then to proceed in the same manner as in this case.

3.9 Identifying the Balanced Bowler in Cricket

Bowling performance in cricket is not a one-dimensional activity. It involves ability to dismiss batsmen of opponent team and restricting the scoring of runs by the opponent. Traditionally, bowling performance is measured with three derived statistics, viz. economy rate, bowling strike rate, and bowling average, for a specific period of time in terms of season, career, etc. The ordered and tabulated values of the said bowling statistics do not provide a clear picture about the performance level of the bowlers. In order to that, a graphical representation is always useful accompanying a table of bowling facts.

If B represents the number of balls bowled by a bowler, W number of wickets taken by a bowler, and R the runs conceded by the bowler in a match or a series of matches, then

$$\text{BSR} = \frac{B}{W}, \quad \text{ER} = \frac{R}{B} \times 6 \quad \text{and} \quad \text{BA} = \frac{R}{W}$$

The three measures have different purposes. The BSR indicates the average number of balls bowled by a bowler to take a wicket, and BA measures the average number of runs conceded by the bowler between wickets. Both these measures quantify the wicket-taking ability of bowlers but concerning two different criteria, viz. the number of balls delivered and the runs conceded. However, both the measures do not define clearly how easier scoring of runs is against the bowler. Thus, comes ER, it provides the average number of runs conceded by the bowler per over. This statistic is very important and specific to limited overs cricket, where bowlers are expected to concede fewer runs along with their ability to take wickets.

Using the above three statistics, quite a lot of performance measures have been developed to quantify the performance of bowlers by Beaudoin and Swartz (2003), Gerber and Sharp (2006), Lemmer (2002, 2005, 2016), etc. Lemmer (2006) developed a measure of current bowling performance of players in combining with their career performances. Receiver operating characteristic (ROC) curves were used by Manage, Mallawaarachchi, and Wijekularathna (2010) for measuring quality of decisions in cricket. Indika and Wickramasinghe (2014) analyzed the performance of bowlers using the data from 2013 Champions Trophy. An explicit measure was developed by Lemmer (2016) which can be used to quantify the wicket-taking ability of the bowlers. However, a few studies are observed which have specifically focused graphical representation for comparing the performance of bowlers. A contour plot was depicted by Kimber (1993) for comparing the performance of bowlers in cricket. Though the plot was simple as well as powerful, it does not seem to be widely used because of its complexity in the construction and interpretation (van Staden, 2009). Therefore, the use of contour plot was extended by van Staden (2009) for comparison of batting, bowling, and all-round performance of cricketers in Indian Premier League (IPL) 2008. In this section, we are going to introduce trilinear plot to identify the balanced bowler in limited overs cricket. The trilinear plot is not applied in sports analytic earlier, and so, this is the unique contribution of the work to the existing literature.

3.9.1 Data Collection

A better bowler in limited overs cricket is expected to possess the ability of taking wickets and restricting of runs simultaneously. Thus, a balance between the two traits quantified by the three bowling statistics is necessary. However, all the three measures (i.e., BA, ER, and BSR) are having negative dimension; i.e., smaller the value better is the bowler. Usually, in practice all the bowlers do not have the three statistics at the same level. To understand the progress of a bowler along the three measures, we need to bring the three measures to a comparative scale by normalization. Accordingly, we select a sample of bowlers and their performance statistics. The bowlers need to satisfy the following two criteria.

- Bowlers who have dismissed at least 10 batsmen in IPL in 2018;
- Bowlers who have bowled at least 120 balls in IPL 2018.

Based on the criteria mentioned above, 31 bowlers are selected in the sample. Performance statistic of the bowlers in IPL 2018 is provided in Appendix 3.4.

3.9.2 Normalization for Balanced Score of a Bowler

Next, BSR, ER, and BA are normalized using the formula provided below. The normalization brings the values BSR, ER and BA lie within 0 and 1 and hence is in comparative scale and also makes them unit free and so the normalized values can be added to each other.

$$\text{Norm\_BSR}_i = \frac{\text{BSR}_i - \min_i(\text{BSR})}{\text{Range(BSR)}} \tag{3.75}$$

where BSR_i is the bowling strike rate of the i^{th} bowler in the sample and min (BSR) is the bowler with minimum value of bowling strike rate across all the bowlers in the sample. Similarly, Norm_ER and Norm_BA are defined. Like BSR, ER, and BA their normalized values too have negative dimensions. For each of the bowler, we now add the values of Norm_BSR_i, Norm_ER_i, and Norm_BA_i and calculate the percentage contribution of each of Norm_BSR_i, Norm_ER_i, and Norm_BA_i to the total. Ideally, a balanced bowler is expected to have a trio of (33.33, 33.33, and 33.33%).

3.9.3 Application of Trilinear Plot to Identify a Balanced Bowler

To represent the trio, that gives equivalent percentage contribution of each of Norm_BSR_i, Norm_ER_i, and Norm_BA_i to their total, we use the trilinear plot. The trilinear plot is used for the special case where all rows of three variables are positive, can be added together, and sums up to a fixed quantity. All the three features are attended successfully by the percentage contribution data of Norm_BSR_i, Nrom_ER_i, and Norm_BA_i.

The plot frame of the graph is an equilateral triangle. Each perpendicular line forms the vertex to the opposite arm (called as altitude) of the triangle represents a variable. Scales are distributed along the altitudes with 0% at the base and 100% at the vertex (c.f. Figure 3.1). To prevent the scale and labels from interfering with the data points and with one another, the scales are sometimes projected to the sides of the triangle. The plot coordinates for point (x_1, x_2, x_3) are determined by drawing a line from each vertex to the opposite side of the triangle such that the line is perpendicular to the opposite side. The intersection of these three lines defines the plot point. This graphical technique can represent three variables in two dimensions. For details on the plot, one may refer to Harris (1999).

The plot below shows the performance of the 31 bowlers of our sample. The plot is produced in *R* with the help of the packages *ggplot* and *ggtern*. The bowler who is closest to the centroid of the triangle is the most balanced bowler with almost equal contribution from the normalized values of all the three-bowling performance statistics. The Euclidean distance of a bowler from the centroid can provide us with

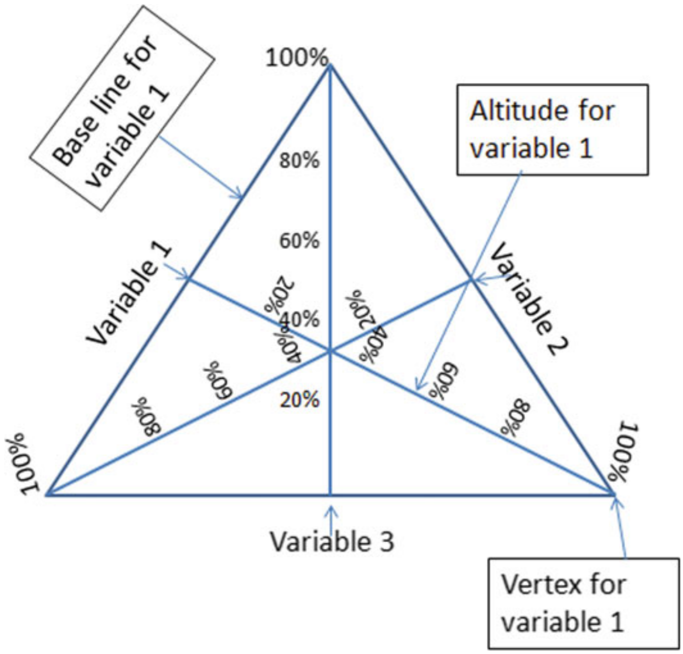

Fig. 3.1 A trilinear plot showing the three axes

the numerical value of his balance. The lesser the value, more balanced is the bowler. A balanced bowler does not mean that he is a good bowler instead; it reflects that he is equally qualified in restricting runs as well as picking up of wickets. Looking at the graph (Fig. 3.2) and the Euclidean distance of the bowlers from the centroid of the graph (c.f. Appendix 3.4), we find that Unadkat, Chawla, Gowtam, Siraj, and Shakib are the most balanced bowlers of IPL 2018.

Having balance is alone not sufficient. A selector shall look at balance (characterized by the balance in restricting runs and taking wickets) as well as the overall performance of the bowler. Thus, for measuring the performance of the bowlers we resort to CBR$^{\#}$ developed by Lemmer (2012), described in detail in Sect. 3.4. The CBR$^{\#}$ is basically the harmonic mean of the BSR, ER, and BA of a bowler adjusted for match situation and importance of the batsmen dismissed by the bowler. This is too a measure of negative dimension.

When we plot the (Euclidean distance, CBR$^{\#}$) of bowlers in a scatter diagram, we get Fig. 3.3. In the figure, the bowler with low values of Euclidean distance and CBR$^{\#}$ shall adjust themselves in the lower left corner of the graph (Q_1). As lower the Euclidean distance more is the balance, and lower the CBR$^{\#}$, better is the performance so in order to bring stability to a team some of the bowlers from the lower left corner shall be considered while selection. Figure 3.3 suggests that Kuldeep, Gopal, Mishra, Kunal, Chahal, Chahar, and Sunil Narine are some of the bowlers who are balanced and have a better performance score.

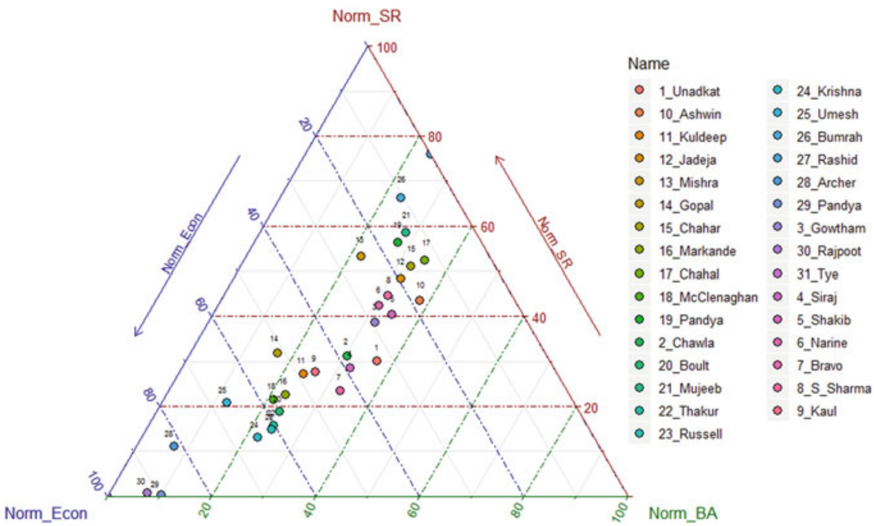

Fig. 3.2 Trilinear plot representing the balance of bowlers in IPL 2018

It is not correct to infer that the bowlers outside Q_1 are not good enough, especially those in Q_4 might even be good bowlers, but they may be good either at picking wickets or in restricting of runs but not exceptionally well in both. Bowlers in Q_1 are expected to be all season bowlers and can be relied upon during most of the time in a match. However, when taking of wickets or restricting of runs becomes the priority (any one, not both), the team must resort to bowlers like Mujib or Bumrah (for restricting runs) or to Umesh or Tye (for wickets).

3.10 Impact of Withdrawal Batting Powerplay on Bowlers

Modifications in the rules of a game at the international level are done by the governing bodies of the concerned sports with a definite goal in mind. The rules governing different sports are often open to modifications at the international level. Common reasons for such amendment of rules in sports include improving players' performance, attracting more spectators, simplifying the game, attending to commercial pressure, adapting sports to children's needs and interests, and preventing injuries (Aris, Argudo, & Alonso, 2011). Rules are also modified if teams/players are found to take undue advantage of existing ones. The International Cricket Council (ICC) often changes the rules in cricket at the international level. In the words of Silva, Manage, and Swartz (2015) opines that—"Rules in cricket, at the international level are changed more frequently compared to the other major sports like football, baseball, etc."

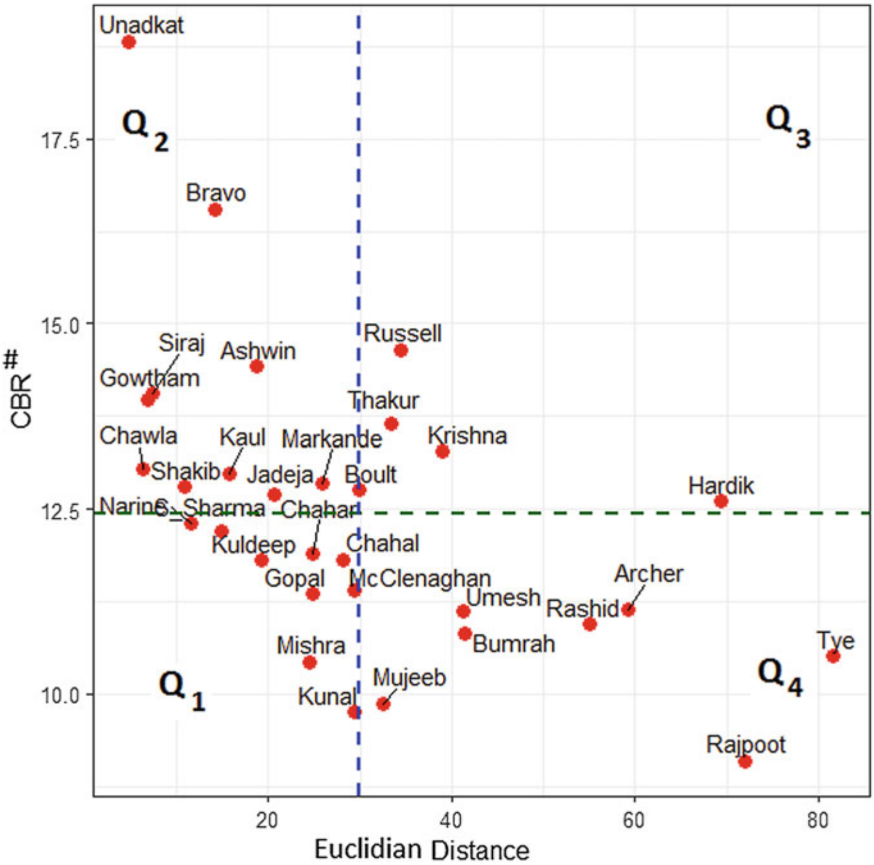

Fig. 3.3 Showing Euclidean distance versus CBR$^{\#}$ of the bowlers in IPL 2018

In test cricket, each team bats for two innings and the team with more aggregate runs scored in their two innings wins the test match. In test matches, there is no restriction on the number of overs played in an innings. Thus, taking wickets is more vital than conceding runs by bowlers in test cricket. Test cricket has a fixed viewer base, which comprises quality cricket enthusiast. One-day International Cricket (ODI) and Twenty20 cricket are limited overs cricket. While ODIs are generally restricted to 50-overs per innings, Twenty20, comprises of 20-overs per innings. Introduction of Twenty20 cricket in the international format in 2005 expanded the spectator base and made the game more exciting and entertaining compared to the ODIs. This suddenly made ODIs sluggish and ineffectual in terms of popularity. Crowd reduced both in the ground and at home in front of television. However, when Twenty20 constitutes a shallow edition of cricket, ODIs constitutes the essence of the game. Popularity of one should not come at the expense of the other. Thus, the sustenance and restoration of the popularity of 50-over match were imperative. Thus, for the continued existence of ODI, onus came on the ICC to bring necessary modifications.

Accordingly, ICC responded with the introduction of 'powerplay' in ODI matches played from July 2005 (Aashish, 2011). Before that in ODIs, each innings used to start with a fielding restriction for the first 15 overs. During those overs, the fielding team could place only two players outside the 30-yard circle, guarding the boundary. This was done to encourage high scoring shots, i.e., fours and sixes by the batsmen. Following the end of fielding restrictions, i.e., at the end of 15 overs, the field was spread out (maximum of five fielders guarding the boundary) making lofted shots from batsman carrying a risk of getting out, caught at the boundary. At this stage (16th–40th over), batsmen try to keep the scoreboard moving scoring ones and twos, taking advantage of the spread out the field and hardly taking any risk. Generally, this part of the innings is dull to watch as chances of dramatic things to happen during the middle overs are lesser. Then starts the 'death overs' (41th–50th over), the last ten overs of ODI innings. This is the stage when the ball gets old, and the bowlers are tired and the last opportunity for the batsmen to score runs. The batsmen played risk-free cricket during the middle overs to store their best for the last when scoring runs matters and preserving wickets are not. During this stage, the ODI innings becomes exciting again. To introduce some excitement in the middle overs, ICC from July 2005 introduced three sets of fielding restrictions and termed it as the powerplay. The first power-play is comprised of the first ten overs of an innings, which are mandatory. This is followed by two further powerplay blocks, of five overs each, to be decided by the fielding team's captain. Also, both these powerplays shall end before the start of the 41st over.

The ICC believed that the strategy would work and restore life to the middle over, but it did not seem to work in the way it was contemplated. Captains of the bowling teams started taking the two blocks of powerplays consecutively between the 11th to the 20th over unless the batsmen are very ruthless on the bowlers. This strategy was planned, as bowlers in the remaining part of the innings can get an opportunity to make up for the losses during the power-play. This shifted the middle overs from 16th to 40th to 21st to 40th, when batsmen play risk-free cricket making the innings dull and non-happening at that stage.

Therefore, in October 2008, ICC introduced the batting powerplay. This comprises of a five-over block a fielding restriction to be decided by the batting team. This was in addition to the mandatory first ten overs of powerplay and a five-over block of powerplay to be decided by the fielding team's captain to be called as the bowling powerplay. As the batting powerplay is to be decided by the batting team, it shall be chosen when the innings are poised toward the batting side, and the batsmen are well set and ready to score quick runs taking advantage of the fielding restriction. ICC conceptualized that most of the fielding teams kept overs of their best bowlers for the death overs. Now with batting powerplay coming in teams needs to employ their best bowlers in the batting powerplay overs, so they need to resort to a lesser bowler during the death overs (Chopra, 2011). This shall enhance runs scoring and also a brain teaser for both the captains as to 'when to take batting powerplay' and 'how to use the bowling resources during the batting powerplay.' Later in October

2012, ICC scrapped the bowling powerplay, keeping the mandatory powerplay of first ten overs and the batting powerplay (a block of five overs) at the desecration of the batting team but before the start of the 41st over.

The batting powerplay, however, did not work the way it was expected. While it was believed that this powerplay should be a nightmare for the bowlers, it also increased the possibility of wicket fall. Several articles like Aashish (2011), and Chopra (2011) with relevant analysis, supported that batting powerplay was not altogether a batsmen's paradise, but bowlers too had their share of the pie.

However, a committee formed by ICC for analyzing the existing ODI rules headed by Indian bowler of yesteryears Anil Kumble in May 2015 recommended the withdrawal of the batting powerplay along with some other changes. Consequently, ICC scrapped the batting powerplay from the ODIs played after July 5, 2015. The reason cited was to provide a bit of 'breathing room' to bowlers in 50-over cricket. The above discussion sets the background of the investigation 'Has the withdrawal of the batting powerplay benefitted the bowlers?'.

3.10.1 Methodology

The batting powerplay was withdrawn with an aim of creating a situation which better suits the bowlers. It is thus expected that the bowlers must have improved the performance post-withdrawal of the batting powerplay. In order to that the performance of the bowlers is quantified using the measure adjusted combined bowling rate ($CBR^{\#}$) developed by Lemmer (2012) (c.f. Sect. 3.4). The bowlers who have the experience of bowling in ODI matches both in the pre- and post-withdrawal of batting powerplay shall be considered. Moreover, only those bowlers shall be considered who bowled during the batting powerplay. After experimentation, the following criteria for the selection of the bowlers for analysis are decided.

(a) Bowlers who had bowled at least one over each in batting power-play during the period when batting power-play was in existence in 10 or more ODIs shall be considered.
(b) Bowlers who have bowled at least in 10 ODIs after batting power play were withdrawn by the ICC in 2015.

A bowler who satisfies both (a) and (b) is selected for analysis. While measuring the performance of a bowler (included in analysis), only those matches where the bowler has bowled at least one over in the batting powerplay shall be taken. This is done so that the influence of batting powerplay in the bowling performance could be included. However, after the withdrawal of the batting powerplay all the innings, in which the bowler has bowled at least one over, are considered for performance measurement.

3.10.2 Data and Computation

At the very selection stage of the bowlers for analysis, the researchers need to know if a bowler under consideration has bowled in the batting powerplay of a match in which he played (apparently when the batting powerplay was in existence). This information (which bowlers bowled during the powerplay) is not available in the scorecard of the matches. So, some reliable source which stores the commentary of ODI matches along with information about the powerplays is required. The Web site www.espncricinfo.com keeps such commentary in their archives and is the source of data.

Regarding the data types, if the bowler has bowled in the batting powerplay, innings-wise number of balls bowled by the bowler, innings-wise runs conceded by the bowler, and innings-wise number of wickets dismissed by the bowler and the batting positions of those wickets. Also, the total runs scored in the match, and the total number of balls bowled in the game are necessary to compute the different measures of bowling performance.

Now for each of the bowler selected for analysis, the values of bowling strike rate (BSR), economy rate (*ER*), bowling average (BA) and $\mathrm{CBR}^{\#}$ are computed based on the aggregate values of B, R, R', and W (c.f. Sect. 3.4 using Lemmer (2012)) from all matches in which the bowlers bowled during batting powerplay. The same exercise is repeated for the selected bowlers for the matches played after the withdrawal of batting powerplay. Thus, for each bowler, we shall get values of BSR, ER, BA, and $\mathrm{CBR}^{\#}$ in pairs one while batting powerplay was in existence and the other after its withdrawal. This calls for the application of Wilcoxon matched-pair signed-rank test separately for BSR, ER, BA, and $\mathrm{CBR}^{\#}$ values to test if the withdrawal of batting powerplay has improved the performance of bowlers. The tests shall be done for all the bowlers taken together as well as separately for spinners and pace bowlers included in the sample. Normal approximation of the statistic can be applied for sample sizes more than 25 and is used here only for all bowlers with a sample size of 27. For the other two samples of spinners and pace bowlers the Wilcoxon matched pair signed rank test without normal approximation are done. It may be recalled here that, unlike others in case of this test, if the value of the test statistic is less than the critical value than the null hypothesis is rejected. It shall be noted here that the alternative hypotheses of the tests shall be one-tailed hypothesizing that the average values of BSR, ER, BA and $\mathrm{CBR}^{\#}$ across all the bowlers have decreased significantly after the withdrawal of the batting powerplay. As $\mathrm{CBR}^{\#}$ is successful in providing match specific bowling performance of a bowler, and combines all the other rates so it is more comprehensive compared to the other measures. The CBR is an appropriate measuring bowling performance for a single match or fewer number of matches.

Out of 27 bowlers, there are 8 spinners and 19 pace bowlers. Details of the bowlers along with their performance statistics can be seen in Appendix 3.5. The Wilcoxon matched-pair signed-rank test is used to compare the performance related statistics of bowlers (BSR, ER, BA, and $\mathrm{CBR}^{\#}$) when the batting powerplay was in existence

(October 2008–July 2015) and after its withdrawal (July 2015–July 2018). The results of the tests are provided in Table 3.2.

3.10.3 Results and Discussion

Table 3.2 shows that the median values of BSR, ER, BA, and CBR$^{\#}$ have shown an increase after the withdrawal of the batting powerplay when the performance of all the bowlers is taken together. This means that the performance of bowlers, in aggregate, in terms of their wicket-taking ability and ability to restrict runs has gone down significantly after the withdrawal. When only spinners are considered, the values of BSR, ER, BA, and CBR$^{\#}$ on an average have remained same (statistically). In case of pace bowlers too, all the three factors viz. BSR, BA, and CBR$^{\#}$, are statistically the same on an average both when the batting powerplay was in existence and after the withdrawal of the batting powerplay. Thus, in no way (wicket-taking ability or in terms of restricting of runs) bowlers of any expertise (spin or pace) is benefitted by the withdrawal of batting powerplay from one-day matches, rather in certain cases, it has become worse, viz. economy rate of fast bowlers and BSR, ER, BA, and CBR$^{\#}$ for all the bowlers taken together.

As discussed, change in the rules in any sports is a regular phenomenon. Rules are changed for myriad number of reasons. Each occasion rationale for alteration in rule varies and the modification is carried out with a particular objective. Thus, a critical assessment of the impact of a particular alteration in rule is crucial to assess its success in achieving the intended goal. In this backdrop, when the impact of the withdrawal of batting powerplay in ODI in 2015 is examined, it is found that the objective was not achieved in the direction it was expected. Batting powerplay was withdrawn primarily to offer bowlers with an opportunity to improve their performance. The purpose was to create an ambience in which bowlers can tackle and withstand the onslaught of batsmen in a more proficient manner. However, in reality there did not take place any substantial change in the performance of the bowlers. It is perhaps due to the fact that the idea of the withdrawal of the powerplay was ill conceived and it was not tested through a reflective process in domestic or other lower levels.

In 2008, batting powerplay was introduced with some definitive objectives. One was that there would be uncertainty as regards the period in which powerplay can be introduced, as it depended on the convenience of the batting side of the team. It was also believed that it would push the scoreboard up with batsmen in an advantage to play big shots at lower risks. Most significantly, with the powerplay on, the best of the bowlers have to get deployed in the match. This was obvious to result in the exhaustion of best of bowlers in the team before the final '10 overs' begins. This, in turn, not only puts the batting team in advantage but also makes the game exciting per se with huge score to be chased likely.

However, in reality, it did not happen the way it was expected. Batting power play was used mostly at a time when a batting partnership is steady and runs are been scored mostly by two well-set batsmen in the wicket in a consistent manner.

Table 3.2 The Result of the Wilcoxon matched-pair signed-rank test to compare the performance statistics of bowlers during and after the withdrawal of the batting powerplay

		Bowling strike rate		Economy rate		Bowling average		CBR#	
		During	After	During	After	During	After	During	After
All bowlers ($n = 27$)	Median	35.28	38.19	5.21	5.28	31.28	35.31	12.22	12.62
	Variance	469.74	540.75	0.45	0.33	358.69	462.99	3.463	4.212
	Test-statistic	−1.7778		−2.3064		−1.7298		−2.1863	
	p-value	0.0377 (<0.05)		0.0105 (<0.05)		0.0418 (<0.05)		0.0144 (<0.05)	
	Conclusion	Median BSR increased significantly after the withdrawal of batting powerplay		Median ER increased significantly after the withdrawal of batting powerplay		Median BA increased significantly after the withdrawal of batting powerplay		Median CBR# increased significantly after the withdrawal of batting powerplay	
Spinners (n =8)	Median	40.71	47.5	4.59	5.04	31.83	36.18	10.33	11.26
	Variance	1426.3	1246.8	0.59	0.26	1101.1	849.9	3.576	2.674
	Test-statistic	5		9		6		6	
	Critical value (5%)	5		5		5		5	
	Conclusion	No significant difference in median BSR during and after the withdrawal of batting powerplay		No significant difference in MEDIAN ER during and after the withdrawal of batting powerplay		No significant difference in median BA during and after the withdrawal of batting powerplay		No significant difference in median CBR# during and after the withdrawal of batting powerplay	
Pacers ($n = 19$)	Median	33.63	37.79	5.55	5.56	29.18	35.31	12.32	12.7
	Variance	60.24	190.29	0.24	0.25	67.84	325.57	2.553	4.354
	Test-statistic	74		46		67		57	

(continued)

Table 3.2 (continued)

		Bowling strike rate		Economy rate		Bowling average		CBR#	
		During	After	During	After	During	After	During	After
	Critical value (5%)	53		53		53		53	
	Conclusion	No significant difference in median BSR during and after the withdrawal of batting powerplay		Median ER increased significantly after the withdrawal of batting powerplay		No significant difference in median BA during and after the withdrawal of batting powerplay		No significant difference in median CBR# during and after the withdrawal of batting powerplay	

With the taking of batting powerplay, the batsmen have to change gears and go after the bowling. They have to resort to risky high scoring shots than they can afford against the best of opposition bowlers. But unlike death overs, loosing wickets here matters, as a significant portion of the innings remains even after the end of batting powerplay. This lack of clarity in approach leads to the downfall of the batsmen during batting powerplay. Losing of wickets during batting powerplay, in an attempt to accelerate the run rate tells on the batting side, which in turn, helped the bowlers to improve their performance figures. Thus, the batting powerplay did not always end up being batsmen's delight, but bowlers had their own stories of success during the batting powerplay overs and the overs that followed the batting powerplay. With the sudden withdrawal of the batting powerplay, the bowlers lost the advantage, which is reflected in this analysis.

The sport's governing body initiates every rule modification process with a definite goal in mind. It is necessary to revisit after a fair amount of time and clarify to what extent the goals underlying the modification are attained. This work attends the same issue for cricket, i.e., analyzing the change made in the powerplay rules in ODI in 2015. The international cricket governing body ICC made the change with an aim to improve the performance of the bowlers. Nevertheless, the exercise finds that the change did not go in the favor of the bowlers.

Modification in rules of any sports at the highest level should be investigated through a reflective process at domestic levels or lower levels before their final introduction. This is what is expected from a well-managed sport. Nothing of this type happened before the withdrawal of the batting powerplay, and so the change ended up in a reverse outcome. In days to come, several other changes may be seen in the powerplay rules in cricket. However, taking a cue from this exercise, changes made by a sport's governing body in the rules of the game shall be initially implemented in lower level (or domestic level), before executing the changes at the international level. The impact of the changed rules at the lower level shall be analyzed using appropriate analytical tools, so that the outcome of the change on the game can be properly identified. If the impact of the change at the lower level is in keeping with the purpose for which the change is proposed, only then such rule change shall be implemented at the international level. This shall make the process of change more time consuming, but it is expected to provide better outcome than changing rules arbitrarily based on some trivial conceptualization.

Aris, Argudo, and Alonso (2011) in a review work on changes in rules of different sports observed that eighty percent of the studies do not report the outcome of the modification of rules in various sports. Thus, it is very important to analyze the matches played post-modification to deduce conclusion about the success of a particular alteration in rule. This is the motivation that has guided to undertake the present work. The study concludes that the withdrawal of the batting powerplay has failed to fulfill the purpose. 'Breathing space' for bowlers, which was a phrase coined by the ICC spokesperson after the withdrawal of batting powerplay, was not attained in reality. This is definitely not an example of good governance.

3.11 Conclusion

Here, we have discussed the various performance measures that are available in cricket to quantify the performance of cricketers. In a balanced cricket team, every player in the playing eleven is expected to perform well whether he is a batsman or bowler or all-rounder or wicket keeper. As we know that cricket is a game of uncertainty, so sometimes players are unable to give his optimal performance for the team. Thus, it is essential to evaluate the performance of each player according to his ability to identify the efficiency of a cricket player. Therefore, a new performance measure is developed that can be used to quantify the batting, fielding, bowling, all-rounding, and wicket keeping performance of cricketers and combines them into a single index based on data from the scorecard of matches. The proposed measure of performance can also be used to quantify the performance of a whole cricket team in terms of batting, bowling, fielding and wicket keeping. It can also be used for comparing the performance of a cricket team in home, away and neutral ground. This performance evaluation would be helpful for the captain, coach, selectors, and team management of the concerned teams to choose an optimum squad of players.

In addition, a graphical display called trilinear plot is implemented to identify the balanced and the better bowler in cricket based on bowling average, economy rate and bowling strike rate. Using this simple and powerful graph, the selectors can identify the bowlers who can be relied upon at taking wickets or in restricting of runs or exceptionally well in both. This graph is applicable for any format of cricket and also can be used to identify the balanced batsmen, all-rounders, etc. Moreover, International Cricket Council (ICC) governing cricket at the international level has made several changes in the powerplay rules since its introduction. The withdrawal of 'batting powerplay' by ICC in 2015 was done to provide an advantage to the bowlers. Therefore, using the different measures of bowling performances whether the performance of the bowlers is improved or not after the withdrawal of the batting powerplay is also discussed. The findings suggest that the bowlers' performance has deteriorated significantly on an aggregate after the change. Thus, the modification in the rules of a game without proper planning and experimentation indicates the need for better governance.

Appendix 3.1

Table 3.3 Wicket weights of the batting position by Lemmer (2005)

Batting position	Wicket Weights
	Lemmer (2005)
1	1.30
2	1.35
3	1.40
4	1.45
5	1.38
6	1.18
7	0.98
8	0.79
9	0.59
10	0.39
11	0.19

Appendix 3.2

For application of wicket keeping performance, we have visited subsequent pages of the website www.espncricinfo.com and collect the data related to wicket keepers' performance of the Indian Premier League (IPL) for the year 2018. Since Indian as well as overseas players participated in the IPL, so for such an exercise we have to set some selection criteria. These are very simplified selection criteria but merely essential. The following restrictions are thus imposed for considering a wicket keeper in IPL 2018 season.[2] Any wicket keeper who has

(i) Kept wickets in at least three innings in IPL 2018;
(ii) Dismissed at least three batsmen in IPL 2018;
(iii) Have batted for at least three innings in IPL 2018;
(iv) Faced at least 60 balls as batsman in IPL 2018;
(v) Having a batting average of at least 10 in IPL 2018.

The conditions (i)–(v) are satisfied by the wicket keepers are J. C. Buttler (RR), K. D. Karthik (KKR), P. A. Patel (RCB), M. S. Dhoni (CSK), K. L. Rahul (KXIP), W. P. Saha (SRH), Q. de Cock (RCB), S. Goswami (SRH), R. R. Pant (DD), and I. Kishan (MI).

[2]IPL 2018 Teams: Chennai Super Kings (CSK), Rajasthan Royals (RR), Kolkata Knight Riders (KKR), Royal Challengers Bangalore (RCB), Delhi Daredevils (DD), Kings XI Punjab (KXIP), Mumbai Indians (MI), Sunrisers Hyderabad (SRH)

Table 3.4 Performance of wicket keepers in IPL 2018

Name	Match	Total wick. wts	Bye runs conc.	Avg. adj runs	S_{i1}	S_{i2}	S_i	Rank
J. Buttler	13	15.75	11	48.241	1.380	1.125	1.185	2
M. S. Dhoni	16	16.04	10	53.576	1.532	1.140	1.233	1
Q. de Kock	8	13.56	6	22.824	0.652	1.077	0.976	6
K. L. Rahul	14	14.97	8	48.891	1.398	1.102	1.172	3
D. Karthik	16	19.47	11	36.867	1.054	1.036	1.041	5
S. Goswami	6	8.99	1	16.973	0.485	1.095	0.951	7
R. R. Pant	14	8.78	6	52.054	1.489	0.945	1.074	4
I. Kishan	14	13.58	13	25.053	0.716	0.805	0.784	9
P. A. Patel	6	3.71	0	29.771	0.851	0.851	0.851	8
W. P. Saha	11	9.55	1	15.294	0.437	0.819	0.728	10
Most compatible β value = 0.7				Weights	0.237	0.763		

In the Table 3.4, we can see that M. S. Dhoni has acquired the first position followed by J. Buttler in IPL 2018. The worst wicket keeping performance in IPL 2018 as per the ranking was given by W. P. Saha. The study however could not consider the non-quantitative attributes of the wicket keepers like his cricketing sense or interaction with other fielders or the guidance that the wicket keeper passes on to the bowlers. These issues could not be quantified and hence cannot be made a part of the performance measurement of a wicket keeper.

Appendix 3.3

To demonstrate the new measure to quantify the overall performance of the cricketers, we have selected 40 players from first four seasons of IPL, viz. IPL 2008, IPL 2009, IPL 2010, and IPL 2011. These 40 players were selected based on the five different criterions. The criterions are as follows:

(i) The player was in the playing eleven in at least 5 matches in each of the IPLs.
(ii) Either the player has bowled at least 10 overs in each of the IPLs or faced at least 100 balls in each of the IPLs as batsman or both.
(iii) Kept wickets in at least three innings in each of the IPLs.
(iv) Dismissed at least three batsmen in each of the IPLs.

Table 3.5 Different factors with its respective weights in IPLs

	Factors	IPL1 weights	IPL2 weights	IPL3 weights	IPL4 weights
Batting	Batting average	0.298	0.332	0.325	0.345
	Batting strike rate	0.387	0.421	0.361	0.397
	Average percentage of contribution	0.315	0.247	0.316	0.258
Fielding	No. of catches taken	0.536	0.472	0.511	0.473
	No. of run outs	0.464	0.528	0.489	0.527
Bowling	Bowling average	0.327	0.302	0.342	0.348
	Economy rate	0.374	0.379	0.349	0.331
	Bowler's strike rate	0.299	0.319	0.309	0.321
Wicket keeping	No. of catches	0.375	0.314	0.279	0.328
	No. of stumping	0.266	0.328	0.382	0.315
	No. of bye runs conceded	0.359	0.358	0.339	0.357

In Table 3.6, it has been revealed that SK Raina was the best performer in first four seasons of IPL followed by Y. Singh and Y. K. Pathan, respectively. However, if we look at closely individual performance of the players in IPL then in IPL I and III, the best performer was S. K. Raina and, in IPL II and IPL IV the best performer was A. Gilchrist followed by J. H. Kallis.

Table 3.6 Performance scores of the cricketers in IPLs

S. no.	Players name	Performance score in				Average Perf_Score	Rank
		IPL4	IPL3	IPL2	IPL1		
1	S. K. Raina	0.701	1.000	0.908	1.000	0.902	1
2	Y. Singh	0.915	0.602	0.856	0.906	0.820	2
3	Y. K. Pathan	0.908	0.775	0.729	0.839	0.813	3
4	K. D. Karthik	0.430	0.966	0.829	0.815	0.760	4
5	R. G. Sharma	0.522	0.695	0.866	0.894	0.744	5
6	A. Gilchrist	0.578	0.725	1.000	0.626	0.732	6
7	I. K. Pathan	0.717	0.722	0.755	0.729	0.731	7
8	J. H. Kallis	1.000	0.882	0.564	0.470	0.729	8
9	M. S. Dhoni	0.872	0.657	0.691	0.625	0.711	9
10	V. Sehwag	0.597	0.891	0.422	0.936	0.711	10
11	J. A. Morkel	0.703	0.670	0.695	0.727	0.699	11
12	D. Bravo	0.677	0.390	0.849	0.738	0.663	12
13	P. Chawla	0.620	0.590	0.630	0.803	0.661	13
14	A. Mishra	0.681	0.636	0.641	0.535	0.623	14
15	R. Uthappa	0.674	0.718	0.560	0.529	0.620	15
16	T. Dilshan	0.364	0.519	0.672	0.901	0.614	16
17	R. P. Singh	0.660	0.535	0.609	0.606	0.603	17
18	V. Kohli	0.751	0.457	0.613	0.561	0.595	18
19	Z. Khan	0.589	0.677	0.573	0.513	0.588	19
20	R. Bhatia	0.701	0.416	0.581	0.608	0.577	20
21	S. Warne	0.541	0.335	0.689	0.704	0.567	21
22	V. Kumar	0.733	0.431	0.591	0.509	0.566	22
23	A. B. de Villiars	0.691	0.454	0.675	0.368	0.547	23
24	A. Agarkar	0.556	0.447	0.544	0.565	0.528	24
25	P. Ojha	0.605	0.575	0.467	0.454	0.525	25
26	S. Tendulkar	0.496	0.510	0.602	0.442	0.512	26
27	K. Sangakara	0.481	0.512	0.529	0.513	0.508	27
28	S. Dhawan	0.942	0.334	0.190	0.549	0.504	28
29	P. Kumar	0.504	0.272	0.637	0.585	0.499	29
30	S. Trivedi	0.471	0.551	0.470	0.483	0.494	30
31	M. Jayawardene	0.452	0.601	0.494	0.424	0.493	31
32	W. Saha	0.390	0.301	0.449	0.764	0.476	32
33	L. Balaji	0.618	0.393	0.504	0.352	0.467	33
34	I. Sharma	0.548	0.277	0.562	0.429	0.454	34
35	M. Muralidharan	0.314	0.441	0.645	0.393	0.448	35

(continued)

Table 3.6 (continued)

S. no.	Players name	Performance score in				Average Perf_Score	Rank
		IPL4	IPL3	IPL2	IPL1		
36	G. Gambhir	0.428	0.438	0.311	0.452	0.407	36
37	R. Dravid	0.386	0.446	0.323	0.408	0.391	37
38	S. Badrinath	0.365	0.385	0.312	0.485	0.387	38
39	V. Rao	0.321	0.120	0.714	0.370	0.381	39
40	S. Sreesanth	0.435	0.173	0.428	0.384	0.355	40

Appendix 3.4: Raw and Processed Performance Statistics and Euclidean Distance from Centroid of the Selected Bowlers

Name of bowler	Balls	Runs	Wickets	Econ. rate	Strike rate	Bowling average	Euclidean distance	CBR#
Unadkat	302	486	11	9.66	27.45	44.18	4.72	18.82
Chawla	294	412	14	8.41	21.00	29.43	6.36	13.04
Gowtham	222	294	10	7.95	22.20	29.40	6.85	13.98
Siraj	246	367	11	8.95	22.36	33.36	7.46	14.06
Shakib	342	456	14	8.00	24.43	32.57	10.95	12.80
Narine	366	467	17	7.66	21.53	27.47	11.53	12.30
Bravo	321	533	14	9.96	22.93	38.07	14.21	16.55
S_Sharma	264	333	12	7.57	22.00	27.75	14.83	12.20
Kaul	396	547	21	8.29	18.86	26.05	15.79	12.97
Ashwin	304	410	10	8.09	30.40	41.00	18.78	14.43
Kuldeep	308	418	17	8.14	18.12	24.59	19.18	11.81
Jadeja	246	303	11	7.39	22.36	27.55	20.58	12.71
Mishra	222	264	12	7.14	18.50	22.00	24.53	10.43
Gopal	186	236	11	7.61	16.91	21.45	24.77	11.36
Chahar	229	278	10	7.28	22.90	27.80	24.77	11.90
Markande	264	368	15	8.36	17.60	24.53	25.87	12.85
Chahal	300	363	12	7.26	25.00	30.25	28.13	11.80
McClenaghan	240	332	14	8.30	17.14	23.71	29.39	11.41
Kunal	241	284	12	7.07	20.08	23.67	29.46	9.77
Boult	317	466	18	8.82	17.61	25.89	29.82	12.76
Mujeeb	224	262	11	7.02	20.36	23.82	32.54	9.86
Thakur	280	431	16	9.24	17.50	26.94	33.38	13.64
Russell	227	355	13	9.38	17.46	27.31	34.46	14.65

(continued)

(continued)

Name of bowler	Balls	Runs	Wickets	Econ. rate	Strike rate	Bowling average	Euclidean distance	CBR#
Krishna	168	260	10	9.29	16.80	26.00	38.91	13.28
Umesh	319	418	20	7.86	15.95	20.90	41.28	11.11
Bumrah	324	372	17	6.89	19.06	21.88	41.44	10.82
Rashid	408	458	21	6.74	19.43	21.81	55.00	10.95
Archer	211	282	14	8.02	15.07	20.14	59.17	11.14
Hardik	256	381	18	8.93	14.22	21.17	69.31	12.60
Rajpoot	157	224	11	8.56	14.27	20.36	71.95	9.09
Tye	312	424	22	8.15	14.18	19.27	81.65	10.50

Source www.espncricinfo.com

Appendix 3.5: Bowling Strike Rate, Economy Rate and Bowling Average of Selected Players During and After the Withdrawal of Powerplay

Name of the bowler	Type	Innings		Bowling strike rate		Economy rate		Bowling average		CBR#	
		During*	After	During*	After	During*	After	During*	After	During*	After
S. Benn	Spin	12	13	141.6	138	5.15	4.96	121.6	114.2	13.61	13.16
C. Anderson	Spin	13	10	23.64	30.6	6.29	5.11	24.8	26.1	12.69	10.94
R. Ashwin	Spin	47	11	39.75	53.81	4.88	5.79	32.38	52	10.72	13.67
T. Chatara	Pace	16	14	47.26	39	5.31	5.43	41.89	35.31	12.81	12.70
J. Fulkner	Pace	27	24	29.66	34.58	5.79	5.52	28.64	31.86	12.32	11.87
Mohd. Hafeez	Spin	27	19	48.08	65.76	3.95	4.36	42.71	47.84	9.11	10.46
Mohd. Irfan	Pace	23	15	35.31	43	4.96	5.09	29.18	36.5	10.95	12.88
R. A. Jadeja	Spin	53	15	42.07	76.36	4.67	5.27	32.75	37.09	10.54	12.72
K. Jarvis	Pace	13	10	50.57	42.5	6.15	5.28	51.85	37.4	15.84	12.61
N. Kulasekhara	Pace	48	15	48.1	34.06	5.21	5.66	41.76	32.12	12.26	12.62
S. Lakmal	Pace	14	40	32.17	37.79	4.95	5.18	26.54	32.67	10.42	11.32
L. Malinga	Pace	76	20	31.59	54.17	5.46	6	28.81	54.17	12.23	14.60
M. Morkel	Pace	51	17	28.78	42.42	4.99	5.27	23.95	37.26	10.64	11.60
M. Mortaza	Pace	13	44	41.96	36.75	4.94	5.12	34.53	31.41	11.61	11.48
C. Mpofu	Pace	19	13	47.66	38.19	5.63	5.87	44.71	37.38	14.25	14.26
W. Parnell	Pace	24	18	29.93	27.93	5.77	5.79	28.8	26.93	12.71	11.51
N. L. T. C. Perera	Pace	18	28	30.19	33.07	5.64	5.95	28.42	32.8	12.61	13.46
R. Hussain	Pace	21	21	35.28	33.17	5.55	5.73	32.66	31.69	12.67	13.14
S. A. Hasan	Spin	52	29	41.67	41.19	4.5	5.14	31.28	35.27	10.12	11.57
T. Southee	Pace	36	35	33.63	45.63	5.61	5.56	31.5	42.32	12.37	13.11

(continued)

(continued)

Name of the bowler	Type	Innings		Bowling strike rate		Economy rate		Bowling average		CBR#	
		During*	After	During*	After	During*	After	During*	After	During*	After
M. Starc	Pace	22	29	27.14	28.98	4.73	5.14	21.38	24.82	9.60	10.66
D. Steyn	Pace	50	12	33.89	36.78	4.69	5.89	26.52	36.11	10.21	13.05
S. Narine	Spin	35	13	37.98	38	3.83	4.16	24.26	26.37	8.09	8.74
I. Tahir	Spin	12	46	27.33	35.46	4.45	4.94	20.29	29.17	8.87	10.63
J. Taylor	Pace	13	10	32.36	88.5	6.27	7.17	33.86	105.8	13.30	19.96
W. Riaz	Pace	26	19	26.51	58.88	5.73	5.97	25.33	58.56	12.01	14.89
C. Woaks	Pace	14	41	42.12	33	6.26	5.12	43.89	28.15	15.07	10.73

References

Aashish. (2011). *Game changer rules like power play, UDRS, Free hit in ODI Cricket!*. http://blogcricketlovers.blogspot.com/2011/02/game-changer-rules-in-odi-cricket.html#.W-FnctUzbIU. Retrieved on September 18, 2018.

Abraham, A. (2005). Handbook of measuring system design. In P. H. Sydenham & R. Thorn (Eds.) (pp. 901–908). Wiley Ltd.

Arias, J. L., Argudo, F. M., & Alonso, J. I. (2011). Review of rule modification in sport. *Journal of Sports Science and Medicine, 1,* 1–8.

Bailey, T. (1989). *The greatest since my time*. London: Hodder and Stoughton.

Bailey, M., & Clarke, S. R. (2006). Predicting the match outcome in one day international cricket matches, while the game is in progress. *Journal of Sports Science and Medicine, 5*(4), 480–487.

Barr, G. D. I., & Kantor, B. S. (2004). A criterion for comparing and selecting batsmen in limited overs cricket. *Journal of the Operational Research Society, 55*(12), 1266–1274.

Barr, G. D. I., & Van Den Honert, R. (1998). Evaluating batsmen's scores in test cricket. *South African Statistical Journal, 32*(2), 169–183.

Beaudin, D., & Swartz, T. (2003). The best batsmen and bowlers in one day cricket. *South African Statistical Journal, 37,* 203–222.

Borooah, V. K., & Mangan, J. E. (2010). The "bradman class": An exploration of some issues in the evaluation of batsmen for test matches, 1877–2006. *Journal of Quantitative Analysis in Sports, 6*(3). https://doi.org/10.2202/1559-0410.1201.

Bretteny, W. (2010). *Integer optimization for the selection of a fantasy league cricket team* (M.Sc. Dissertation), Nelson Mandela Metropolitan University, South Africa.

Chakrabarty, N., & Bhattacharjee, D. (2012). Aggregation and weightage issues concerning composite index development: Experience with digital divide in Asian countries. *Asia Pacific Journal of Library and Information Science, 2,* 24–37.

Chopra, A. (2011) *Timing the powerplays right*. http://www.espncricinfo.com/story/_/id/22460850/powerplays-foxing-teams. Retrieved on September 12, 2018.

Choudhury, D. R., Bhargava, P., Reena, & Kain, S. (2007). Use of artificial neural networks for predicting the outcome of cricket tournaments. *International Journal of Sports Science and Engineering, 1*(2), 87–96.

Clarke, S. R. (2007). *Studying Variability in Statistics via Performance Measures in Sport.* 56th session, Kolkata: Indian Statistical Institute.

Damodaran, U. (2006). Stochastic dominance and analysis of ODI batting performance: The Indian cricket team, 1989–2005. *Journal of Sports Science and Medicine, 5,* 503–508.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss function. *Machine Learning, 29,* 103–130.

Duckworth, F. C., & Lewis, A. J. (1998). A fair method of resetting the target in interrupted one-day cricket matches. *Journal of Operational Research Society, 49,* 220–227.

Flack, P. A., & Lachiche, N. (2004). Naïve Bayesian classification of structured data. In *Machine learning* (pp. 1–37). Boson: Kluwer Academic Publishers.

Gerber, H., & Sharp, G. D. (2006). Selecting a limited overs cricket squad using an integer programming model. *South African Journal for Research in Sport, Physical Education and Recreation, 28*(2), 81–90.

Harris, R. L. (1999). *Information graphics a comprehensive illustrated reference* (pp. 423–424). New Yoer: Oxford University Press.

Hemachandran, B. (2009). *Cricket's best wicket keepers*. Retrieved 18 October 2011, from http://www.cricforum.com/viewtopic.php?t=15492.

Hermanus Hofmeyr Lemmer, (2011) Predicting the names of the best teams after the knock-out phase of a cricket series. European Journal of Sport Science 14 (sup1):S191-S196.

Indika, R., & Wickramasinghe, P. (2014). Bowlers' performance in 2013 champions trophy. *Annals of Applied Sports Science, 2*(1), 1–10.

Iyenger, N. S., & Sudarshan, P. (1982, December). A method of classifying regions from multivariate data. *Economic and Political Weekly* 2048–2052.

Kimber, A. (1993). A graphical display for comparing bowlers in cricket. *Teaching Statistics: An International Journal for Teachers, 15*(3), 84–86.
Kimber, A. C., & Hansford, A. R. (1993). A statistical analysis of batting in cricket. *Journal of Royal Statistical Society, 156*(3), 443–455.
Lemmer, H. (2004). A measure for the batting performance of cricket players. *South African Journal for Research in Sport, Physical Education and Recreation, 26*(1), 55–64.
Lemmer, H. (2005). A method for the comparison of the bowling performances of bowlers in a match or series of matches. *South African Journal for Research in Sport, Physical Education and Recreation, 27*(1), 91–103.
Lemmer, H. H. (2002). The combined bowling rate as a measure of bowling performance in cricket. *South African Journal for Research in Sport, Physical Education and Recreation, 24*(2), 37–44.
Lemmer, H. H. (2008). An Analysis of Players Performances in the First Cricket Twenty20 World Cup Series. South African Journal for Research in Sports: Physical Education and Recreation. 30(2): 71-77.
Lemmer, H. H. (2011). Performance measure for wicket keepers in cricket. *South African Journal for Research in Sport, Physical Education and Recreation, 33*(3), 89–102.
Lemmer, H. H. (2016). Wicket taking ability of bowlers. *International Journal of Sports Science and Coaching, 11*(2), 231–236.
Lewis, A. J. (2005). Towards fairer measures of player performance in one-day cricket. *Journal of the Operational Research Society, 56,* 804–815.
Maini, S., & Narayanan, S. (2007). The flaw in batting averages. *The Actuaries* 30–31.
Manage, A. B. W., Mallawaarachchi, K., & Wijekularathna, K. (2010). Receiver operating characteristics (ROC) curves for measuring the quality of decisions in cricket. *Journal of Quantitative Analysis in Sports, 6*(2), 1–13.
Narayanan, A. (2008). Test wicketkeepers—An analysis. The ESPNcricinfo Blog. http://blogs.cricinfo.com/itfigures/archives/wicketkeepers/
Norusis, M. (2007). *The SPSS statistical procedure companion*. India: Prentice Hall.
Perl, J. (2001). Artificial neural networks in sports: New concepts and approaches. *International Journal of Performance Analysis in Sports, 1*(1), 106–121.
Perl, J., & Weber, K. (2004). A neural network approach to pattern learning in sports. *International Journal of Computer Science in Sports, 3*(1), 1–4.
Saikia, H., & Bhattacharjee, D. (2011). On classification of all-rounders of the Indian Premier League (IPL): A Bayesian approach. *Vikalpa A Journal of Decision Makers, 36*(4), 25–40.
Saikia, H., Bhattacharjee, D., & Lemmer, H. H. (2012). Predicting the performance of bowlers in IPL: An application of artificial neural network. *International Journal of Performance Analysis in Sports, 12*(1), 75–89.
Silva, A. J., Costa, A. M., Oliveira, P. M., Reis, V. M., Saavedra, J., Perl, J., et al. (2007). The use of neural network technology to model swimming performance. *Journal of Sports Science and Medicine, 6*(1), 117–125.
Silva, R. M., Manage, A. B. W., & Swartz, T. B. (2015). A study of the powerplay in one-day cricket. *European Journal of Operational Research, 244,* 931–938.
Srinivas, A., & Vivek, T. (2009). *IPL an inside story of cricket and commerce*. The Lotus Collection, New Delhi: Roli Books Pvt. Ltd.
Suleman, M. T., & Saeed, M. A. (2009). *Option on human performance: A case study of Indian premier league*. Retrieved December 21, 2009, from http://papers.ssrn.com/sol3/…/SSRN_ID1474093_code1208558.pdf.
van Staden, P. J. (2009). Comparison of cricketers' bowling and batting performances using graphical displays. *Current Science, 96*(6), 764–766.
Wood, G. H. (1945). Cricket scores and geometric progression. *Journal of Royal Statistical Society, 108,* 12–22. Series A (Statistics in Society).
Young, W. A., & Weckman, G. R. (2008). Evaluating the effects of aging for professional football players in combine events using performance-aging curves. *International Journal of Sports Science and Engineering, 2*(3), 131–143.
Zhang, H. (2004). The optimality of Naïve Bayes. In *American Association for Artificial Intelligence*. Available at http://www.aaai.org. Retrieved on November 19, 2010.

Chapter 4
Fielding Performance Measure in Cricket: Issues, Concerns and Solution

4.1 Introduction

Performance statistics is often used to indicate the level of achievement of the players in sports (Clarke, 2007). Cricket is a team sport that prides itself on the fact that statistics is generated in each and every ball of the game in terms of players, teams, games, series, and seasons. This game heavily relies on performance statistics to evaluate its players. A scorecard of a cricket match always helps the reader to understand performance of the cricketers in a given match. Over the decades, summary of the statistics, viz. batting average, strike rate, economy rate, etc., have been published in many sporting chronicles. These chronicles publish many traditional facts and figures in terms of players' as well as teams' performances, usually at the end of a series/tournament. For example—strike rate, batting average, percentage of contribution for batsmen, bowling average, economy rate, and bowling strike rate for bowlers. These performance statistics, however, do not provide a fair reflection of a player's true ability of performance (Lewis, 2005). Thus, some alternative measures have been proposed by various authors by combining the traditional performance statistics (cf. Table 4.1) in order to determine the more useful measures of players' performance.

Cricket is a game of numbers like runs scored and conceded, balls faced or delivered, wins and losses, etc. All these are turn around batting as well as bowling traits of the game. However, people say catches win matches, and they indeed may, but we do not count them well enough to check it. Also, we do not count them at all when they are not taken. This is one of the reasons why batsmen-wicketkeepers have replaced in the place of specialist wicketkeepers in a team because we would like to count only runs. Nowadays, you will see the number of catches and stumping executed by a wicketkeeper at the scorecard of the match. Furthermore, a run out executed by a piece of fielding brilliance is basically supposed to be the fielder's wicket. We knew that Jhonty Rhodes and Yuvraj Singh both were great fielders, but who has executed more run-outs? Can anyone answer that? No, because fielding statistics just do not exist in proper way. However, sometimes the challenge is to give credit the most

H. Saikia et al., *Cricket Performance Management*, Indian Statistical Institute Series,
https://doi.org/10.1007/978-981-15-1354-1_4

Table 4.1 Different performance measures in cricket

Author(s)	Year	Purpose of the Study	Comment(s)
Wood	1945	Defined a batting average by considering a player's not-out scores as completed innings	Performance measurement for batsman
Kimber and Hansford	1993	Proposed an alternative batting average by considering a player's completed innings and not-out innings separately	Performance measurement for batsman
Lemmer	2002	Devised a bowling performance measure called combined bowling rate (CBR), which comprises of bowling average, economy rate, and bowling strike rate	Performance measurement for bowlers
Barr and Kantor	2004	Proposed a batting performance measure that is a weighted product of batting average and strike rate	Performance measurement for batsman
Lemmer	2004	Suggested an alternative batting performance measure called BP, which is calculated as a product of exponentially weighted batting average (EWA), standardized consistency coefficient (SCC), and standardized strike rate (SSR)	Performance measurement for batsman
Lewis	2005	Proposed a performance measure for cricketers using ball-by-ball information of the match based on the Duckworth/Lewis method	Can be applied for batsman, bowler and all-rounder but ball-by-ball information of the match is required which is very difficult to attain
Damodaran	2006	Suggested an alternative method of batting average to handling a player's not-out scores using Bayesian approach	Performance measurement for batsman

(continued)

Table 4.1 (continued)

Author(s)	Year	Purpose of the Study	Comment(s)
Gerber and Sharp	2006	Proposed various performance indices for measure the batting, bowling, all-rounder, fielding, and wicket keeping performance of the players in cricket separately	These indices are based on the number of specialist players available in the respective skills
Maini and Narayanan	2007	Proposed another solution to the batting average issue that is based on the number of balls faced by the batsman	Performance measurement for batsman
Suleman and Saeed	2009	Developed a performance index called SS Index to quantify the batting, bowling, all-rounder, and wicket keeping performance in Twenty20 cricket. It is a combination of different factors that are related to the performance measurement of cricketers	Subjective weights are used to measure the performance of batsman, bowler, all-rounder and wicket keeper
Borooah and Mangan	2010	Proposed new batting averages using Gini Coefficient, and this measure also pays attention to the contribution of player's runs to the team total	Performance measurement for batsman
Beaudoin and Swartz	2003	Defined a new statistic that is based on the runs scored or allowed per match to measure the performance of batsmen and bowlers, respectively	Performance measurement for batsman and bowler
Lemmer	2011	Developed a performance measure for wicketkeeping in test cricket by combining the batting performance and dismissal rate. The dismissal rate of the wicketkeeper defined as the number of dismissals (i.e., catches plus stumpings) divided by the number of matches played	This is the first attempt to quantify the performance of wicketkeeper along with his batting ability. The extended measure also included the number of byes conceded

deserving fielder in situations where multiple fielders were involved to execute a run-out. Moreover, we do not even know who the poorest catcher in a cricket team is. That is, who is the cricketer in your team, who has the highest ratio of dropped catches per match? What the drop catch ratio of cricket players was impossible to find out, before the ESPN Cricinfo website provided ball-by-ball information of cricket matches. These are some of the issues related to fielding statistics and analytics that should be available for cricketers in order to measure the feat of fielders.

The game of cricket is yet to have a proper measure for quantifying fielding performance. The first pioneering and successful attempt in quantifying fielding performance measure is that of Saikia, Bhattacharjee, and Lemmer (2012). In the words of Swartz (2017)—'*the first quantitative investigation of fielding was undertaken by Saikia, Bhattacharjee, and Lemmer* (2012). *Their approach involved the subjective assessment of every fielding play to provide a weighted measure of fielding proficiency*'. This chapter mostly based on the paper of Saikia, Bhattacharjee, and Lemmer (2012) proposes several changes in the point system to further simplify the matter so that it can be easily applied to live cricket matches. A fielder in a cricket match always tries to prevent the scoring of runs that the batsman desires to get by quickly returning the ball to the stumps. He also tries to get the batsman out by catching the ball or by executing a run out. In cricket, great feats of batting and bowling usually hit the headlines, but it is often ignored that good fielding too can make a crucial contribution to a team's success (Knight, 2006). With the advent of ODI and the recent format of Twenty20 cricket, fielding has become more athletic and is developing as a discipline to rival batting and bowling in the interest of spectators (Hagen, Roach & Summers, 2001). Players too are becoming more and more concerned about their fielding performance; they have started believing the phenomenon that effective and tight fielding can help their team to win a match (Knight, 2006).

Usually, the fielding performances that are perceivable from the scorecard of a match are number of catches and run-outs. But a fielder does more than that in the field which is not expressed in the scorecard unlike the other prime skills of the game like batting and bowling. International Cricket Council (ICC) releases the ranking of teams as well as batsmen, bowlers, and all-rounders periodically on the basis of on-field performance of the cricketers. But, there is no ICC ranking for fielders or wicketkeepers.

Table 4.1 reveals that first step was taken by Gerber and Sharp (2006) to define a fielding performance index in cricket. Prior to that, most of the performance statistics developed were concerned with the batting and bowling performances only. According to Gerber and Sharp (2006), the fielding ability of the i^{th} player can be defined as,

$$\text{FLD}_i = \left(\frac{\text{dismissal rate of fielder} \quad i}{\text{sum of the dismissal rates of all the fielders}}\right) \times \text{number of specialist fielder} \tag{4.1}$$

Here, the dismissal rate of a fielder is defined as the average number of fielding dismissals (i.e., run outs and catches) made by the player per match. According to that measure, the fielding ability of a player is subject to the number of specialist fielders that are available in the team. Since fielding of a player mostly depends on his own skill based on which the player performs independently and not on the skills of any other specialist fielder, the measure has got an obvious fault and limitation. Considering this limitation, we have developed a fielding performance measure in this chapter.

4.2 The Fielding Performance Measure

Regarding fielding in cricket, a fielder always tries to limit the number of runs that the batsman can score and also tries to get the batsman out by catching the ball or by executing a run out. Traditionally, only two factors have been considered as the performance indicator of fielders in cricket—number of catches taken and run outs accomplished. These two factors can be considered as performance indicator of fielders in cricket. The availability of only these two factors in the scorecard of a match is the prime reason behind the use of this approach. However, it can be easily observed that there are several other factors involved in fielding. Even in considering catches and run outs, other factors such as the difficulty level and the accuracy of such actions need to be considered. For example—while taking a catch, a fielder may have to dive toward his natural side or his wrong side. In case of a run out, a fielder has to throw hard to the appropriate (i.e., wicketkeeper/bowler) end, or depending on the situation, needs to hit the stumps directly. Therefore, to provide an accurate measure of fielding performance, one has to take into account all the subtle nuances involved in the process of fielding. Such an approach which considers all these factors will require ball-by-ball information of a match. According to Srinivas and Vivek (2009), statistician needs to evaluate every single ball bowled in Twenty20 and ODI for better understanding of players' performance. A measure that does not take all these acute aspects into account will be incomplete in determining the performance of a cricketer. Therefore, the fielding performance statistic defined in this chapter for the game of cricket is based on ball-by-ball information of a match or series of matches.

In cricket, the fielding positions can basically be divided into two parts, viz. close fielding within the 30 yards circle and distant fielding outside the 30 yards circle (Fig. 4.1). Within these two fielding positions, three different aspects of fielding (i.e., catch, picking the ball, and run out) can be observed as on-field performances of the cricketers. A good and active fielder always tries to catch the ball in flight and/or throw it as hard and quickly as possible, irrespective of the fielding position. An unbelievable catch or a good throw in a cricket match may even reverse the outcome of the match. To consider all these aspects of fielding, ball-by-ball information of a player in a match is required. This can be obtained by observing the match live. Ranges of parameters are considered, and scores are assigned to these parameters

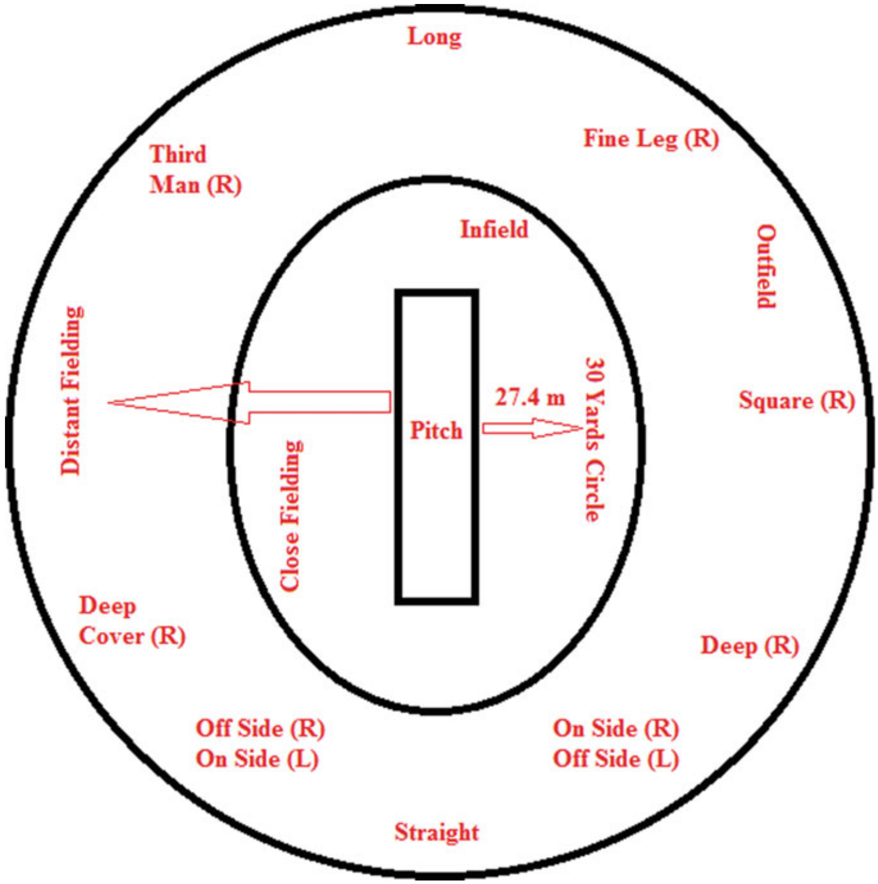

Fig. 4.1 Layout of a cricket field

based on the on-field performances of the fielders. The parameters considered for the study and their corresponding scores are outlined below.

4.2.1 Different Parameters (X_{ij}) Considered Under Fielding

1. **Close fielding within 30 yards circle** (**including bowler and wicketkeeper**)

   (i) **Catch**

      I. Regulation catches [Difficult (1), Easy (0.8), Missed (−0.3)]

      II. High lofted catch

         A. Running/diving backwards [Difficult (1), Easy (0.8), Missed (−0.3)]

- B. With slight adjustment of position [Difficult (1), Easy (0.8), Missed (−0.3)]
- C. With considerable running/diving [Difficult (1), Easy (0.8), Missed (−0.3)]

III. Take on a hard hitted stroke catch
- A. With slight adjustment [Difficult (1), Easy (0.8), Missed (−0.3)]
- B. With dive toward natural side [Difficult (1), Easy (0.8), Missed (−0.3)]
- C. With dive toward wrong side [Difficult (1), Easy (0.8), Missed (−0.3)]
- D. Catching with an upward leap [Difficult (1), Easy (0.8), Missed (−0.3)]

IV. Catch with running/diving
- A. Toward natural side [Difficult (1), Easy (0.8), Missed (−0.3)]
- B. Toward wrong side [Difficult (1), Easy (0.8), Missed (−0.3)]
- C. Toward forward [Difficult (1), Easy (0.8), Missed (−0.3)]
- D. Toward the off-side of stump [Difficult (1), Easy (0.8), Missed (−0.3)]
- E. Toward the leg-side of stump [Difficult (1), Easy (0.8), Missed (−0.3)]
- F. Toward backward [Difficult (1), Easy (0.8), Missed (−0.3)]
- G. With an upward leap [Difficult (1), Easy (0.8), Missed (−0.3)]

V. Direct catch from defensive stroke [Difficult (1), Easy (0.8), Missed (−0.3)]

VI. Lofty ball caught with forward dive [Difficult (1), Easy (0.8), Missed (−0.3)]

VII. Lofty ball caught running backward [Difficult (1), Easy (0.8), Missed (−0.3)]

(ii) **Fielding**

I. Regulation fielding with little adjustment of position and return the ball [Difficult (1), Easy (0.5), Missed (−0.2)]

II. Fielding with dives
- A. Toward natural side [Difficult (1), Easy (0.5), Missed (−0.2)]
- B. Toward wrong side [Difficult (1), Easy (0.5), Missed (−0.2)]
- C. Forward/Backward [Difficult (1), Easy (0.5), Missed (−0.2)]

III. Considerable running to field and return the ball (left/right/forward) [Difficult (1), Easy (0.5), Missed (−0.2)]

IV. Considerable running behind the ball toward the boundary [Difficult (1), Easy (0.5), Missed (−0.2)]

V. Collection of throws from fielders
- A. Regulation collection [Difficult (1), Easy (0.5), Missed (−0.2)]
- B. Throws away but collectable [Difficult (1), Easy (0.5), Missed (−0.2)]

VI. Collection of the ball

A. Considerable running to collect the ball [Difficult (1), Easy (0.5), Missed (−0.2)]
B. With dive toward off-side of the batsman [Difficult (1), Easy (0.5), Missed (−0.2)]
C. With dive toward leg-side of the batsman [Difficult (1), Easy (0.5), Missed (−0.2)]

(iii) **Run out**
I. Direct throw to the stumps [Difficult (1), Easy (0.8), Missed (−0.3)]
II. Indirect throw to wicketkeeper/fielder near the stumps [Difficult (1), Easy (0.8), Missed (−0.3)]
III. Directly taking the ball into the stumps [Difficult (1), Easy (0.8), Missed (−0.3)]
IV. Collect throws and hit stumps with the ball in hands [Difficult (1), Easy (0.8), Missed (−0.3)]

(iv) **Stumping**
I. Batsman left the crease and did not attempt to return [Difficult (1), Easy (0.8), Missed (−0.3)]
II. Batsman stepped forward and tried to return [Difficult (1), Easy (0.8), Missed (−0.3)]

2. **Distant fielding outside 30 yards circle**

(v) **Catch**
I. Regulation catch with adjustment in position [Difficult (1), Easy (0.8), Missed (−0.3)]
II. Catch with considerable running [Difficult (1), Easy (0.8), Missed (−0.3)]
III. High lofted catch
A. With running/diving backward [Difficult (1), Easy (0.8), Missed (−0.3)]
B. With slight adjustment of position [Difficult (1), Easy (0.8), Missed (−0.3)]
C. With considerable running/diving [Difficult (1), Easy (0.8), Missed (−0.3)]
IV. Catch over the fielders head [Difficult (1), Easy (0.8), Missed (−0.3)]
V. Catch with upward leap near the boundary [Difficult (1), Easy (0.8), Missed (−0.3)]

(vi) **Fielding**
I. Regulation fielding with slight adjustment of positions and return the ball [Difficult (1), Easy (0.5), Missed (−0.2)]
II. Considerable running/diving to field and return the ball [Difficult (1), Easy (0.5), Missed (−0.2)]
III. Jump upward/natural side/wrong side with or without running near the boundary [Difficult (1), Easy (0.5), Missed (−0.2)]

(vii) **Run out**
  I. Direct throw to the stumps [Difficult (1), Easy (0.8), Missed (−0.3)]
  II. Indirect throw to wicketkeeper/fielder near the stumps [Difficult (1), Easy (0.8), Missed (−0.3)]

Now, the prime challenge is to list all the possible acts of fielding and then to assign weights to the fielding acts as per their degree of difficulty. Finally, a formula or measure needs to be defined by combining all the acts of fielding as well as weights appropriately. The weighting process is broken up into phases in order to simplify matters. The main categories are catches, run outs, and fielding. The fielding category consisting of all the other aspects not contained in the 'catch' and 'run out' category. Within each category (e.g., catches), weights must be allocated to the different acts of fielding. Fielding that leads to the loss of the wicket of a top-order batsman shall have a higher weight compared to a batsman lower down the order. To include this aspect into the measure, suitable weights are used according to the quality of the dismissed batsman. Accordingly, two fielding performance measures are introduced in this work, viz. a preparatory measure (FP^1) and a fairer fielding measure (FP^2). Details are given in the next subsections.

4.2.2 Preparatory Fielding Performance Measure (FP^1)

The parameter X_{ij} is defined to reflect the achievement of the j^{th} player ($j = 1, 2, \ldots, k$) in the i^{th} ball that he fielded ($i = 1, 2, \ldots, n_j$). If he executes the action successfully, $X_{ij} = 1$ and if he fails, $X_{ij} = 0$. For intermediate success, other values between 0 and 1 are allocated. Consider a potential catch, to be precise, one that can be considered to be within the reach of the fielder. If he catches the difficult ball, $X_{ij} = 1$, if he catches the easy ball, $X_{ij} = 0.8$; if he missed the catch completely then $X_{ij} = 0$. For other details, see Sect. 4.2.1 above.

The fielding performance measure (FP_j^1) of j^{th} player can be defined as

$$\mathrm{FP}_j^1 = \frac{1}{n_j} \sum_{i=1}^{n_j} \left[(1 - \theta_i) w_{ij} X_{ij} + \theta_i b_i w_{ij} X_{ij} \right] \text{ for } n_j \geq 1 \tag{4.2}$$

where n_j is the number of balls fielded by the j^{th} player, w_{ij} represents the weight for the fielding activity performed by the j^{th} player in the i^{th} ball fielded, b_i is the weight for the batsman dismissed in the i^{th} ball, and θ_i is an indicator variable where

$$\theta_i = \begin{cases} 1, & \text{if the fielding activity on the } i^{\text{th}} \text{ ball leads to a dissmissal} \\ 0, & \text{otherwise} \end{cases}$$

The higher the value of the fielding performance measure (FP_j^1) the better is the fielder. A fielder who catches a ball that has been hit by a top-order batsman should be rewarded more than one who catches a similarly hit ball by a tail-ender batsman. To quantify this in the model making use of the wicket weights for ODIs developed by Lemmer (2005), the weighting factor b_i is also incorporated into the formula (cf. Table 4.2). These weights depend on the batting position of the batsman dismissed. The wicket weights for the different batting positions are based on the average of batting averages of a large data set of players, taking into account their average batting positions. Sometimes, it may so happen that a batsman gets injured and retires from batting but continues his innings lower down the order. In such a case, his wicket weight shall correspond to that of his normal batting position.

The wicket weights are converted to weight b_i, as it is felt that weights larger than 1 are justified for the wickets of top-order batsmen, but for lower-order batsmen, just the number of wickets are counted without weighting them. This is also done as it is often difficult to categorize the tail-order batsmen on the basis of their batting skills, and the batting order of such batsmen keeps on changing. Sometimes, a tail ender can hit a big six or even edge a four despite his lack of skills as a batsman.

A further justification of using adjusted weights is as follows. Consider a 'perfect' fielder who is only involved in catches being taken (i.e., $X_{ij} = 1$) during a match. Then, the value of $\theta_i = 1$ for all his actions and the formula (4.2) reduces to

$$FP_j^1 = \frac{\sum_{i=1}^{n_j} b_i w_{ij}}{n_j} \tag{4.2a}$$

If all the catches taken by the fielder are of the most difficult type, then all w_{ij} are equal to 1 (i.e., $w_{ij} = 1$) and the formula (4.2a) becomes

$$\mathrm{FP}_j^1 = \frac{\sum_{i=1}^{n_j} b_i}{n_j} \tag{4.2b}$$

Table 4.2 Wicket weights (b_i) according to batting position of batsman as in Lemmer (2005) and on necessary adjustment for lower order batsmen

Batting position	Wicket weights	
	Lemmer (2005)	Adjusted weights (b_i)
1	1.30	1.30
2	1.35	1.35
3	1.40	1.40
4	1.45	1.45
5	1.38	1.38
6	1.18	1.18
7	0.98	1.00
8	0.79	1.00
9	0.59	1.00
10	0.39	1.00
11	0.19	1.00

If the hypothetical fielder took the catches of lower order batsmen in the match and the weights in column 2 of Table 4.2 are used, Eq. (4.2b) will result in a fielding performance value less than 1 (i.e., $\mathrm{FP}_j^1 < 1$). After taking the most difficult of catches, a player's fielding performance value less than 1 may not be reasonable. By taking a catch, a wicket is lost by the opposing team, and that team accordingly loses some resources. Therefore, for the wickets of lower-order batsmen, the original weights below 1 are replaced by 1 to arrive at the weights b_i in Column 3 (cf. Table 4.2). On the other hand, if the fielder only took catches of higher order batsmen, then his fielding performance value can justifiably be greater than 1 (i.e., $\mathrm{FP}_j^1 > 1$).

4.2.3 *Fairer Fielding Performance Measure (FP^2)*

Keeping the value of performance measure within the range of [0, 1] makes interpretation easier. The upper limit of the preparatory fielding performance measure (FP^1) is greater than 1. Hence, to limit the range of measure FP^1 between 0 and 1, Eq. (4.2) is redefined as,

$$\mathrm{FP}_j^2 = \frac{\sum_{i=1}^{n_j}\left[(1-\theta_i)w_{ij}X_{ij} + \theta_i b_i w_{ij} X_{ij}\right]}{\sum_{i=1}^{n_j}\left[(1-\theta_i)w_{ij} + \theta_i b_i w_{ij}\right]} \text{ for } n_j \geq 1 \tag{4.3}$$

In order to get clarity regarding the working of FP^2, it is useful to consider the above-mentioned special case. In a match, if a 'perfect' fielder is involved in the most difficult catches being taken (i.e., $X_{ij} = 1$ and $w_{ij} = 1$), then with the value of $\theta_i = 1$ for all his actions, Eq. (4.3) becomes

$$\mathrm{FP}_j^2 = \frac{\sum_{i=1}^{n_j} b_i}{\sum_{i=1}^{n_j} b_i} = 1$$

That means the maximum fielding performance value of a player is equal to 1, if he does everything perfectly for all of his actions irrespective of his fielding positions and irrespective of the relative weights. Thus, FP^2 is a fairer measure (and hence the name) than FP^2 in the sense that it does not discriminate between fielding actions. A fielder at third man can have a very high value due to good fielding, as can one who takes remarkable catches.

Another perspective is that a player, whose actions all have a very high degree of difficulty, is disadvantaged in Eq. (4.3) by division by the sum of all the high weights $\left(\text{i.e. } \sum_{i=1}^{n_j}\left[(1-\theta_i)w_{ij} + \theta_i b_i w_{ij}\right]\right)$ compared to one whose actions were easy. In Eq. (4.2), the division is by n_j, so the fielding performance with higher weight gets a better value than one with lower weight. This is an argument in favor of the preparatory measure (FP^1) because the choice of weights reflects the importance of the specific actions.

4.2.4 Determination of Weights

Among the three aspects of fielding, viz. catch, run-out, and picking the ball, the first two lead to a wicket being taken. Thus, their weights must be equal and higher than the weight of 'picking the ball' because saving runs are not as important as taking a wicket.

Suppose 'a' is the weight for 'catch' and also for 'run-out' and 'b' is the weight for 'picking the ball' with $b < a$. A convenient value for 'a' is $a = 1$. Consider the option 'catch' in close fielding within the 30 yards circle. Let d be the weight for a regulation catch. Suppose the most difficult option of catching is II(C), a catch after considerable running (cf. Sect. 4.2.1a), with weight $2d$ allocated to it. Then, put $2d = a$ (i.e., $d = a/2$, where $a = 1$). Weights between d and $2d$ are allocated to the other alternatives under the option 'catch'. Similarly, weights are allocated for all the alternatives under the options close fielding within 30 yards circle and distant fielding outside 30 yards circle (cf. Sect. 4.2.1). Note, however, that for 'catch' and 'run-out' the highest weight of the most difficult option is denoted by $2d = a$, but for 'picking the ball' (i.e., fielding) the highest weight of the most difficult option is denoted by $2d = b$.

In order to decide on the relationship between a and b, consider the most stringent acts in each of the categories 'catch', 'run out', and 'fielding'. A catch after considerable running gets a weight 1 and obviously, a run out with a direct throw also gets a weight 1. Fielding with a dive to the wrong side has a weight $2d = b$. After considering various values of b, it was decided to use $b = \beta.a$ with $\beta = 0.8$, i.e., $b = 0.8$ because in a limited over match the restriction of runs is also very important, but not as important as taking a wicket. Thus, $b = 0.8$ is the highest weight (or value) that a fielder can obtain unless he is also involved in catches and/or run-outs. It should be noted that the weight obtained by a player for one fielding operation must emerge from one type of action and not from multiple actions. All the weights for different alternatives under various options can be seen in Table 4.3.

4.3 Application and Result

Suppose a cricketer took a catch after considerable running and then diving outside the 30 yards circle during a cricket match. One has to recognize which option is appropriate for that catch (cf. Sect. 4.2.1b). Suppose his action leads to the option 'catches with running and diving' and if he took the catch with a fumble, then his weighted fielding performance score on that ball is $2d \times 0.8 = 1.6d = 0.8a = 0.8$ (for weights cf. Table 4.3). The following example shows how this performance measure can be used to quantify the fielding performance of a cricketer. Let us consider a hypothetical fielder who has fielded only six balls in a Twenty20 cricket match, then by his on-field performances, the weighted fielding performance scores in the six different balls are provided in Table 4.4.

Table 4.3 Weights (w_{ij}) of the different fielding parameters

Alternatives	Weights	Express all d's in terms of 'a' ($2d = a$ and $a = 1$)	Relative weights
Catch: Close fielding within 30 yards circle			
I	d	$a/2$	0.5
II(A)	$1.2d$	$1.2 \times (a/2)$	0.6
II(B)	$1.1d$	$1.1 \times (a/2)$	0.55
II(C)	$2d$	a	1
III (A)	$1.2d$	$1.2 \times (a/2)$	0.6
III(B)	$1.4d$	$1.4 \times (a/2)$	0.7
III(C)	$1.5d$	$1.5 \times (a/2)$	0.75
III(D)	$1.3d$	$1.3 \times (a/2)$	0.65
IV(A)	$1.2d$	$1.2 \times (a/2)$	0.6
IV(B)	$1.3d$	$1.3 \times (a/2)$	0.65
IV(C)	$1.1d$	$1.1 \times (a/2)$	0.55
IV(D)	$1.2d$	$1.2 \times (a/2)$	0.6
IV(E)	$1.3d$	$1.3 \times (a/2)$	0.65
IV(F)	$1.5d$	$1.5 \times (a/2)$	0.75
IV(G)	$1.4d$	$1.4 \times (a/2)$	0.7
V	$1.2d$	$1.2 \times (a/2)$	0.6
VI	$1.1d$	$1.1 \times (a/2)$	0.55
VII	$1.3d$	$1.3 \times (a/2)$	0.65
Fielding: Close fielding within 30 yards circle			
I	d	$0.8a/2$	0.4
II(A)	$1.7d$	$1.7 \times (0.8a/2)$	0.68
II(B)	$2d$	$0.8a$	0.8
II(C)	$1.8d$	$1.8 \times (0.8a/2)$	0.72
III	$1.2d$	$1.2 \times (0.8a/2)$	0.48
IV	$1.5d$	$1.5 \times (0.8a/2)$	0.6
V(A)	$1.1d$	$1.1 \times (0.8a/2)$	0.44
V(B)	$1.2d$	$1.2 \times (0.8a/2)$	0.48
VI(A)	$1.4d$	$1.4 \times (0.8a/2)$	0.56
VI(B)	$1.2d$	$1.2 \times (0.8a/2)$	0.48
VI(C)	$1.3d$	$1.3 \times (0.8a/2)$	0.52
Run-out: Close fielding within 30 yards circle			
I	$2d$	a	0.8
II	$1.2d$	$1.2 \times (a/2)$	0.6
III	$1.7d$	$1.7 \times (a/2)$	0.85

(continued)

Table 4.3 (continued)

Alternatives	Weights	Express all d's in terms of 'a' ($2d = a$ and $a = 1$)	Relative weights
IV	$1.5d$	$1.5 \times (a/2)$	0.75
Stumping: Close fielding within 30 yards circle			
I	$1.5d$	$1.5 \times (a/2)$	0.75
II	$2d$	a	0.8
Catch: Distant fielding outside 30 yards circle			
I	$1.1d$	$1.1 \times (a/2)$	0.55
II	$1.2d$	$1.3 \times (a/2)$	0.65
III(A)	$1.5d$	$1.5 \times (a/2)$	0.75
III(B)	$1.2d$	$1.2 \times (a/2)$	0.6
III(C)	$1.3d$	$1.3 \times (a/2)$	0.65
IV	$2d$	a	1
V	$1.8d$	$1.8 \times (a/2)$	0.9
Fielding: Distant fielding outside 30 yards circle			
I	$1.2d$	$1.2 \times (0.8a/2)$	0.48
II	$2d$	$0.8a$	0.8
III	$1.5d$	$1.5 \times (0.8a/2)$	0.6
Run-out: Distant fielding outside 30 yards circle			
I	$2d$	a	1
II	$1.5d$	$1.5 \times (a/2)$	0.75

Once the ball-by-ball weighted performance scores of a cricketer in a match are available, then anyone can easily calculate the sum of $w_{ij}X_{ij}$ and $b_i w_{ij}X_{ij}$ for a given player. Based on the above data, the weighted sum of the fielding performance score of the player using Eqs. (4.2) and (4.3) are

$$\begin{aligned} FP_j^1 &= \frac{1}{n_j}\sum\nolimits_{i=1}^{n_j}\left[(1-\theta_i)w_{ij}X_{ij} + \theta_i b_i w_{ij}X_{ij}\right] \\ &= \frac{1}{6} \times (1.615 + 0.6072) = 0.370 \end{aligned} \tag{4.4}$$

$$\begin{aligned} FP_j^2 &= \frac{\sum\nolimits_{i=1}^{n_j}\left[(1-\theta_i)w_{ij}X_{ij} + \theta_i b_i w_{ij}X_{ij}\right]}{\sum\nolimits_{i=1}^{n_j}\left[(1-\theta_i)w_{ij} + \theta_i b_i w_{ij}\right]} \\ &= \frac{1.615 + 0.6072}{3.03 + 0.44} = 0.640 \end{aligned} \tag{4.5}$$

For a cricketer, both fielding performance measures FP^1 and FP^2 are capable of analyzing the different aspects of fielding according to his given abilities like taking diving catches, throwing at the stumps directly, limiting the opposition's three runs

Table 4.4 Ball-by-ball fielding performance scores for a hypothetical cricketer

Ball No.	Parameters (cf. Section 4.2.1)	Description	Fielding Score (X_{ij})	Weights (w_{ij}) (cf. Table 4.3)	$w_{ij}X_{ij}$	$b_i w_{ij} X_{ij}$
1	(ii).I	Stopped the ball with little adjustment of position within 30 yards circle and return the ball to the wrong end	0.5	0.4	0.2	
2	(ii).II (A)	Field the ball within 30 yards circle by diving toward his natural side which leads to the restriction of runs	0.5	0.68	0.34	
3	(V).I	A catch is being taken with slight adjustment of position outside 30 yards circle with fumble to dismiss the number 5 batsman	0.8	0.55		0.6072[a]
4	(ii).IV	Field the ball as a bowler with considerable running and return it within 30 yards	1	0.6	0.6	
5	(vi).II	Considerable running to field the ball with fumble outside 30 yards circle and return it	0.8	0.8	0.64	
6	(i).II(B)	High lofted catch with slight adjustment inside the 30 yards to dismiss the number eight batsman, but he missed it	-0.3	0.55	-0.165	
				Total	1.615	0.6072

[a]In this table, the values of b_i (i.e. wicket weights according to batting position of the batsmen) come from Table 3.2. Here $b_i = 1.38$ means 5th batting position of the batsman

Table 4.5 Fielding performance of cricketers in Nidahas tri-series trophy 2018

Sl. no.	Players name	Country	Fielding measure (FP^2)	Ranks
1	Suresh Raina	India	1	1
2	Mehidy Hasan Miraz	Bangladesh	0.928	2
3	Sabbir	Bangladesh	0.8116	3
4	NLTC Perera	Sri Lanka	0.7593	4
5	Shikhar Dhawan	India	0.7015	5
6	KL Rahul	India	0.6744	6
7	Rohit Sharma	India	0.6727	7
8	BKG Mendis	Sri Lanka	0.6379	8
9	MK Pandey	India	0.6127	9
10	MD Gunathilaka	Sri Lanka	0.5432	10

into twos, etc. A fielder who has executed the most difficult fielding operations will have a relatively large value. If the purpose of a study is to compare the fielding performances of players irrespective of their fielding positions, then the fairer measure is the appropriate measure to use. If the purpose is to identify the fielder who had the greatest effect on outcome of the match, then preparatory measure should be used.

The defined fielding performance measures are demonstrated below based on data from the Nidahas Twenty20 Tri-series tournament among the teams India, Sri Lanka, and Bangladesh. This tournament was started from March 6, 2018 with a match between India versus Sri Lanka, and the final match of the tournament was played between India and Bangladesh on March 18, 2018. Based on the parameters considered under Sect. 4.2.1, the relevant ball-by-ball information of the matches for all the cricketers are collected from the website www.espncricinfo.com. Accordingly, the fielding performances of the cricketers are quantified using Fairer Fielding Performance Measure (FP^2). The performance of the first ten (10) cricketers in terms of their fielding score can be seen in Table 4.5.

Table 4.5 shows that Suresh Raina was the best fielder of the tournament followed by Mehidy Hasan of Bangladesh. Also, India has five fielders in the top 10 followed by Sri Lanka three, and Bangladesh two. It may be noted that as ground fielding comes into play in the model, more number of catches or run outs implemented is not the only criteria for ranking of the fielders. Once again, as the wicketkeepers have a different role to play, the wicket keeping performance was measured separately and can be seen in Table 4.6. Among the wicketkeepers, KD Karthik was the best wicketkeeper of tournament followed by Mushfiqur Rahim from Bangladesh.

Table 4.6 Fielding performance of wicketkeepers in Nidahas tri-series trophy 2018

Sl. no.	Wicket keepers	Country	Fielding measure (FP^2)	Ranks
1	KD Karthik	India	1	1
3	Mushfiqur Rahim	Bangladeshh	0.490	2
2	MDKJ Perera	Sri Lanka	0.465	3
4	LD Chandimal	Sri Lanka	0.127	4

4.4 Conclusion

With the increasing popularity of cricket, issues related to quantitative analysis of cricket data has gained momentum. Statisticians and data analysts are using such data for predicting match outcome, selecting optimized balanced team, determining the value of players, and so on. Most of these analyses are based on precise quantification of players' performance. While most of the measures of quantifying players' performance take into consideration batting and bowling skills, fielding skill is often ignored. With increasing importance of limited overs game, especially the Tewnty20 format of cricket, saving of runs has become equally important as scoring them. Thus, the necessity of including the fielding ability of a player in the evaluation of performance seems essential. But in the absence of any appropriate tool to quantify fielding performance, this has not been properly attempted.

The fielding performance measures FP defined in this chapter is based on ball-by-ball information of the match. Collecting ball-by-ball information is a tedious job. In the absence of any software or optical observation system, a researcher has to watch matches ball-by-ball and perform the scoring. Thus, given the time constraint and the type of activity involved, it is unlikely that information from large number of matches can be included. It is further improvised by including the fielding performance of cricketers through run-outs and catches (obtained from the scorecard of the other matches of the series). These FP measures can be used to quantify the fielding performances of cricketers for a series of matches, whether it is test, ODI, or Twenty20 cricket. Individual fielding performance scores can then be aggregated to obtain the overall fielding performance of a team. Such an attempt can depict the best fielding team among the teams that participated in a tournament.

However, benefits of fielding performance measure discussed in this chapter can be realized by coaches, players, and team management only when software that can perform all necessary calculations is developed. Capturing the number of runs saved by each fielder accurately may require high precision online time and motion study by high-class technology but not impossible though. With the development of such software, we could work out important but previously unanswerable questions like—how far the ball was from the fielder, accuracy and speed of the throw, how often a second run was successfully made if the fielder took off slowly, etc. Once the enough data is brought in, we could identify which fielder makes an impact in a match and which player only executes chances that are been offered (See Table 4.3).

References

Barr, G. D., & Kantor, B. S. (2004). A criterion for comparing and selecting batsmen in limited overs cricket. *The Journal of the Operational Research Society, 55*(12), 1266–1274.

Beaudoin, D., & Swartz, T. (2003). The best batsmen and bowlers in one-day cricket. *South African Statistical Journal, 37,* 203–222.

Borooah, V. K. & Mangan, J. E. (2010). The "Bradman class": An exploration of some issues in the evaluation of batsmen for test matches, *Journal of Quantitative Analysis in Sports*, *6*(3), 1877–2006. https://doi.org/10.2202/1559-0410.1201.

Clarke, S. R. (2007). *Studying variability in statistics via performance measures in sport*. Netherland: International Statistical Institute.

Damodaran, U. (2006). Stochastic dominance and analysis of ODI batting performance: The Indian cricket team, 1989–2005. *Journal of Sports Science and Medicine, 5*(4), 503–508.

Gerber, H., & Sharp, G. D. (2006). Selecting a limited overs cricket squad using an integer programming model. *South African Journal for Research in Sport, Physical Education and Recreation, 28*(2), 81–90.

Hagen, K. V., Roach, R., & Summers, B. (2001). The sliding stop: A technique of fielding in cricket with a potential for serious knee injury. *British Journal of Sports Medicine, 34,* 379–381.

Kimber, A. C., & Hansford, A. R. (1993). A statistical analysis of batting in cricket. *Journal of the Royal Statistical Society, 156*(3), 443–455.

Knight, J. (Ed.). (2006). Fielding made easy. In: *Cricket for dummies* (p. 131). Chichester, West Sussex, England: Wiley.

Lemmer, H. H. (2002). The combined bowling rate as a measure of bowling performance in cricket. *South African Journal of Research In Sports, Physical Education and Recreation, 24*(2), 37–44.

Lemmer, H. H. (2004). A measure for the batting performance of cricket players. *South African Journal for Research in Sport, Physical Education and Recreation, 26*(1), 55–64.

Lemmer, H. H. (2005). A method for the comparison of the bowling performances of bowlers in a match or a series of matches. *South African Journal for Research in Sport, Physical Education and Recreation, 27*(1), 91–103.

Lemmer, H. H. (2011). Performance measures for wicket keepers in cricket. *South African Journal for Research in Sport, Physical Education and Recreation, 33*(3), 89–102.

Lewis, A. J. (2005). Towards Fairer measures of player performance in one-day cricket. *The Journal of the Operational Research Society, 56*(7), 804–815.

Maini, S. & Narayanan, S. (2007). The flaw in batting averages. *The actuary,* May 30–31.

Saikia, H., Bhattacharjee, D., & Lemmer, H. H. (2012). A double weighted tool to measure the fielding performance in cricket. *International Journal of Sports Science and Coaching, 7*(4), 699–713.

Srinivas, A., & Vivek, T. R. (2009). *IPL an inside story cricket & commerce*. New Delhi: The Lotus Collection.

Suleman, M. T. & Saeed, M. A. (2009). Option on human performance: A case study of indian premier league. Retrieved 21 December, 2009, from http://papers.ssrn.com/sol3/…/SSRN_ID1474093_code1208558.pdf.

Swartz, T. B. (2017). Research directions in cricket. *Handbook of statistical methods and analyses in sports*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods, 445–460.

Wood, G. H. (1945). Cricket scores and geometric progression. *Journal of Royal Statistical Society*, Series A (Statistics in Society), *108*, 12–22.

Chapter 5
Performance-Based Market Valuation of Cricketers

5.1 Introduction

The Indian Premier League (IPL) is a franchisee-based Twenty20 cricket tournament that was initiated by the Board of Control for Cricket in India (BCCI) in 2008. In IPL, the franchises form their teams by competitive bidding from a collection of Indian and international players and the best of young upcoming Indian talents once in every three years. According to Australian batsman Hayden (2011)—'The IPL was modeled on the line of English Premier Football League, where international players add talent and glamor to the local teams.' Nothing of the kind had been attempted in cricket earlier. Team owners bid for the services of cricketers for a total of US$42 million. Each team can purchase a maximum of eight overseas players though only four players can be considered in the playing eleven in a match. Every player has a base price fixed by the IPL authorities, and there is no upper limit. The scale of operations was unheard of and the millions of dollars being bandied about a left large section of people unconvinced about whether the league was sustainable. But Lalit Modi, the founder IPL Commissioner, ploughed on; the players brought their A-game to the party, and the spectators lapped it up (Vasu, 2011). The commercial revenues rose through the league structure and sale of inventory make it the world's richest cricketing tournament (Kitchin, 2008). The overwhelming success of the first season of IPL and sky-high bid prices of the players encouraged many renowned players from Australia, England, South Africa, New Zealand, and Sri Lanka to join the league instead of playing for their domestic leagues (Depken & Rajasekhar, 2010). Five out of the eight teams that participated in IPL had a designated icon player, who was paid an amount 15% higher than the highest-paid player in that team (Rastogi & Deodhar, 2009), while the salaries of the other players are to be decided in the auction. After the end of the first three seasons of IPL, the salary contract of the players with the franchise terminates. Before the beginning of the fourth season, all the players were again sent for auction. During the fresh auction, it was found that the salary offer and even the franchisee of most of the players changed. A similar scenario was observed when the fresh auction of all the players took place every three years. Continuing

H. Saikia et al., *Cricket Performance Management*, Indian Statistical Institute Series,
https://doi.org/10.1007/978-981-15-1354-1_5

with the same trend, a fresh auction of all the players materialized before IPL 2014. Therefore, an attempt is being made to estimate the relative market valuation of the cricketers based on their on-field performances in IPL 2014, IPL 2015, and IPL 2016 seasons. The market valuation shall then be compared with the bid price to find if the bid price is justified by the player's on-field performance during the said IPL seasons.

5.2 Valuation of Players and Competitive Balance

The IPL had a unique method of buying players through an open auction process. With a salary cap, all the franchisee would get an opportunity to buy the best players for their team. This salary cap mechanism was introduced in the 1970s in North America, and the reason behind is that it ensures a more level playing field among participating teams (Mitra, 2010). This is the point that Lalit Modi took into account while implementing the salary cap in IPL because the salary cap mechanism helps to maintain the competitive balance and more scope for an exciting contest among the teams. In order to maintain interest in any league, it is a necessity to maintain competitive balance (Quirk and Fort, 1992). If there are some very strong teams, it will take away interest from the weak teams, and over a period of time, this domination by the strong teams will take away interest from their fans as well (Mitra, 2010).

The player's auction fits in well with some underlying principles in the IPL such as uncertainty of outcomes and competitive balance. Consequently, to fulfill these requirements, some economically irrational salaries have been paid to certain players (Mitra, 2010). It may be recalled that five out of the eight teams had icon players in the first season of IPL. The icon players were R. Dravid for Bangalore, Y. Singh for Punjab, S. Ganguly for Kolkata, S. Tendulkar for Mumbai, and V. Sehwag for Delhi who captained their respective teams. They got a salary of 15% more than the highest-paid player on their team. Chennai had no icon player, and they bid for M. S. Dhoni (who eventually captained the team) at an unbelievable price of US$1.50 million, which made him the highest-paid player of the tournament. Some contrasting results were seen at the auction like—R. Ponting, the captain of the Australian cricket team, was bid for US$0.40 million only, while I. Sharma, with an experience of only one overseas series in 2008, was bid for US$0.95 million. Some other highest-paid players other than the icon players were A. Symonds (US$1.35 million), S. Jayasuriya (US$0.975 million), I. Pathan (US$0.925 million), J. Kallis and B. Lee (both US$0.90 million). Two other fast bowlers, Shane Bond and Kemar Roach, were sold for US$0.75 million and US$0.72 million, respectively. Under normal circumstances, a player like Shane Bond with a strong reputation and an excellent cricketing record would have earned more than a novice like Roach (Mitra, 2010). This is where the peculiarities of the auction come into play. Lalit Modi believes this is nothing but a simple case of market economics, and the players will get paid with respect to the demand for their services as well as value (Mitra, 2010).

There are two ways of looking at a player's performance related to his valuation. First, how it affects the performance of the team and secondly if a superstar effect exists in the IPL. There is no evidence as yet of superstar effect boosting revenues in IPL; however, it is clear from the study by Rastogi and Deodhar (2009) that a superstar or iconic effect does exist. This could explain why franchises like Deccan Chargers and Royal Challengers Bangalore are ready to pay over a million dollars to acquire players like Symonds and Pietersen, respectively, (Mitra, 2010). The second way of judging performance related to valuation is to evaluate the on-field performance of the cricketers in relation to their salary. Since the salary offer to the players was valid for three years only, so in the fresh auction process, the salary offers to the available cricketers would also change. Thus, such a change of the salary should be related to the on-field performance of the cricketers.

In spite of the availability of several works focused on performance measurement of cricketers and ranking them, attempt to relate their performance with pay package is relatively few. Reason for that is the absence of any franchisee-based cricket prior to the start of IPL, where each player has a salary tag attached to him. In the incidence of such a salary tag, instead of ranking of players in terms of performance only, it is necessary to evaluate their performance in relation to their pay package. Two works addressing this issue of—'to what extent the auction-based player salaries are justified by player performances' is of Suleman and Saeed (2008) and Dalmia (2010) both based on the first two seasons of IPL only. While the former work attempted to test if the performances of professional cricket players can be treated as a real option, the later addressed if the players' salaries are justified by their performance. However, the performance indicators of cricketers used in those works require improvement. While Suleman and Saeed (2008) developed a performance index, called the SS Index, which had a subjective approach; the index used by Dalmia (2010) did not make any attempt to standardize/normalize and stabilize the data in terms of their variances. Also, none of these works tried to classify the players based on a pay-performance two-way classification. Thus, an attempt is made in this chapter to evaluate the performance of cricketers in relation to their pay package by normalizing, combining, and stabilizing the various performance-related statistics based on data from IPL 2014, IPL 2015, and IPL 2016.

5.3 Data and Performance Measure

Many players have played in the different seasons of IPL. The huge set of players would make the calculations of the model difficult. Moreover, expectations regarding the level of performance cannot be gauged fairly from a few matches (Bracewell & Ruggiero, 2009). Thus, the players who satisfied all the following conditions were selected for the study.

- The player was in the playing eleven in at least five matches in each of the IPLs.
- Either the player has bowled at least ten overs or has faced at least 100 balls in each IPL as a batsman or both.

There were only 50 players who satisfied both the mentioned criteria. All these players were considered for the study and information about their performance are collected from the Web site www.espncricinfo.com. The performances of these selected cricketers are quantified using the following performance measurement tool.

The performance measure of the i^{th} player is given by,

$$S_i = S_{i1} + \delta_i \tag{5.1}$$

where

$$\delta_i = \begin{cases} S_{i2}^{a_i} + S_{i3}^{1-a_i} - 1, & \text{if } i^{\text{th}} \text{ player is either a bowler or wicket keeper} \\ 0, & \text{if } i^{\text{th}} \text{ player is neither a bowler nor wicket keeeper} \end{cases}$$

where a_i is an indicator variable with,

$$a_i = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ player is a bowler} \\ 0, & \text{if } i^{\text{th}} \text{ player is a wicket keeper} \end{cases}$$

where S_{i1} represents the performance score for batting, S_{i2} represents the performance score for bowling, and S_{i3} represents the performance score for wicket keeping.

This performance measure is almost the same as defined in Chap. 3 (cf. Sect. 3.6.5.1 of Sect. 3.6.5). However, one rudimentary difference from Eq. (3.61) (cf. Chap. 3) is that the fielding ability of a cricketer is not considered in Eq. (5.1). Now, the information about various factors that are considered to measure the performance of batting, bowling, and wicket keeping and along with their process of normalization can be seen in Sects. 3.6.5.2, 3.6.5.3, 3.6.5.4, 3.6.5.5, and 3.6.5.6, respectively (cf. Chap. 3).

5.3.1 Determination of Weights

As mentioned in Sect. 3.6.5.7 of Chap. 3, if w_{jk} represents the weight of the j^{th} factor under the k^{th} ability, then it is calculated as

$$w_{jk} = \frac{C_k}{\sqrt{\underset{i}{\operatorname{Var}}(Y_{ijk})}}, \quad j = 1, 2, 3, 4 \text{ and } k = 1, 2, 3 \tag{5.2}$$

where $\sum_{j=1}^{4} w_{jk} = 1$ for all k and C_k is a normalizing constant that follows

Table 5.1 Weights of the different factors in different seasons of IPL

	Factors	IPL 2014 weights	IPL 2015 weights	IPL 2016 weights
Batting	Batting average	0.239	0.235	0.345
	Batting strike rate	0.393	0.365	0.397
	Average percentage of contribution to the team total	0.366	0.399	0.258
Bowling	Bowling average	0.272	0.362	0.473
	Economy rate	0.377	0.273	0.527
	Bowler's strike rate	0.349	0.364	0.348
Wicket keeping	Number of catches taken	0.252	0.234	0.331
	Number of stumping	0.266	0.243	0.321
	Number of bye runs conceded	0.241	0.279	0.328

$$C_k = \left[\sum_{j=1}^{4} \frac{1}{\sqrt{\underset{i}{\mathrm{Var}}(Y_{ijk})}} \right]^{-1} \tag{5.3}$$

where Y_{ijk} is the normalized score (cf. Sect. 3.6.5.6 of Chap. 3) of the performance measure obtained under the various factors considered to measure the performance in batting, bowling, and wicket keeping. The choice of the weights in this manner would ensure that the large variation in any one of the factor would not unduly dominate the contribution of the rest of the factors (Iyenger & Sudarshan, 1982); (Table 5.1).

5.3.2 Computation of Performance Score

The performance scores of S_{i1}, S_{i2}, and S_{i3} for batting, bowling, and wicket keeping are computed using the same procedure described in Sect. 3.6.5.8 of Chap. 3.

The batting performance ($k = 1$) of the i^{th} player is calculated by

$$S_{i1} = \sum_{j=1}^{4} w_{j1} Y_{ij1}; \tag{5.5}$$

The bowling performance ($k = 2$) of the i^{th} player is calculated by

$$S_{i2} = \sum_{j=1}^{4} w_{j2} Y_{ij2}; \tag{5.6}$$

The performance ($k = 3$) of the i^{th} wicket keeper is calculated by

$$S_{i3} = \sum_{j=1}^{4} w_{j3} Y_{ij3}; \tag{5.7}$$

Now, on obtaining the values of S_{i1}, S_{i2}, and S_{i3} the performance score S_i of the i^{th} player is computed using Eq. (5.1). The performance score of all the 50 players is computed and then converted into corresponding performance index value P_i for the i^{th} player, given by

$$P_i = \frac{S_i}{\max\limits_i (S_i)} \tag{5.8}$$

The performance index value for a player is a number which is lying between 0 and 1 (i.e., $0 < P_i \leq 1$). A higher value of the performance index indicates better performances of the players.

5.4 The Binomial Option-Pricing Model

Binomial option-pricing model is a discrete-time model, most widely used in real options valuation methods. It is represented through a decision tree, describing price movements over time where the market price either can go up or down with some probabilities associated with each movement. This probability is recognized as a risk-neutral probability that is equivalent to the real probability measure. The risk-neutral probabilities are calculated using Eq. (5.9).

$$\text{rf} = p \times M_{\text{u}} + (1 - p) \times (-M_{\text{d}}) \tag{5.9}$$

where rf is the risk-free rate, M_{d} represents downward market price, and M_{u} represents upward market price. The time preference for money is generally expressed by an interest rate. This rate will be positive even in the absence of any risk (Pandey, 2004). Thus, it is called risk-free rate. This risk-free rate is easily obtainable from any default risk-free government security. The risk-free rate considered for this study is 3.9429% which is equal to 0.039429 (Indian T-bills one-year rate).[1]

Let us consider a multi-period binomial option-pricing model with discrete-time periods 0, 1, 2, and 3. Now, let us suppose that market value \$1 at time 0 can be

[1] http://in.reuters.com/article/domesticNews/idINBOM47576920090617

compounded to [\$1 × (1 + rf)] at time 1, or market value \$1 at time 0 can be reduced to [\$1/(1 + rf)] at time 1. So,

$$M_{\mathrm{u}} = [\$1 \times (1 + \mathrm{rf})] = [\$1 \times (1 + 0.039429)] = 1.03 \tag{5.10}$$

$$M_{\mathrm{d}} = \left[\frac{\$1}{(1 + \mathrm{rf})}\right] = \left[\frac{\$1}{(1 + 0.039429)}\right] = 0.96 \tag{5.11}$$

The probability of up state is found to be $p = 0.503$ and down state $q = 0.497$. This risk-neutral probability is assumed to be the same for time periods 2 and 3.

Now, the players' market value of upward cash flow ($\mathrm{CF_u}$) and downward cash flow ($\mathrm{CF_d}$) on the basis of their performance score after the IPL 2014 can be calculated as

$$\begin{aligned} \mathrm{CF_u} &= \text{Bid Price} \times (1 + p_1) \\ \mathrm{CF_d} &= \text{Bid Price} \times (1 - q_1) \end{aligned} \tag{5.12}$$

Further, the player's market value of upward cash flows and downward cash flows after IPL 2015 and IPL 2016 is calculated by the following equations.

$$\begin{aligned} \mathrm{CF_{uu}} &= \mathrm{CF_u}(1 + p_2) \\ \mathrm{CF_{ud}} &= \mathrm{CF_u}(1 - q_2) \\ \mathrm{CF_{du}} &= \mathrm{CF_d}(1 + p_2) \\ \mathrm{CF_{dd}} &= \mathrm{CF_d}(1 - q_2) \\ \mathrm{CF_{uuu}} &= \mathrm{CF_{uu}}(1 + p_3) \\ \mathrm{CF_{uud}} &= \mathrm{CF_{uu}}(1 - q_3) \\ \mathrm{CF_{udu}} &= \mathrm{CF_{ud}}(1 + p_3) \\ \mathrm{CF_{udd}} &= \mathrm{CF_{ud}}(1 - q_3) \\ \mathrm{CF_{duu}} &= \mathrm{CF_{du}}(1 + p_3) \\ \mathrm{CF_{dud}} &= \mathrm{CF_{du}}(1 - q_3) \\ \mathrm{CF_{ddu}} &= \mathrm{CF_{dd}}(1 + p_3) \\ \mathrm{CF_{ddd}} &= \mathrm{CF_{dd}}(1 - q_3)) \end{aligned} \tag{5.13}$$

where p_1, p_2, and p_3 are the performance scores of a given player in IPL 2014, IPL 2015, and IPL 2016 respectively and q_1, q_2, q_3 are the complementary values of p_1, p_2, and p_3, respectively (i.e., $q_i = 1 - p_i$). The above calculations can be plotted in the tree diagram as provided in Fig. 5.1.

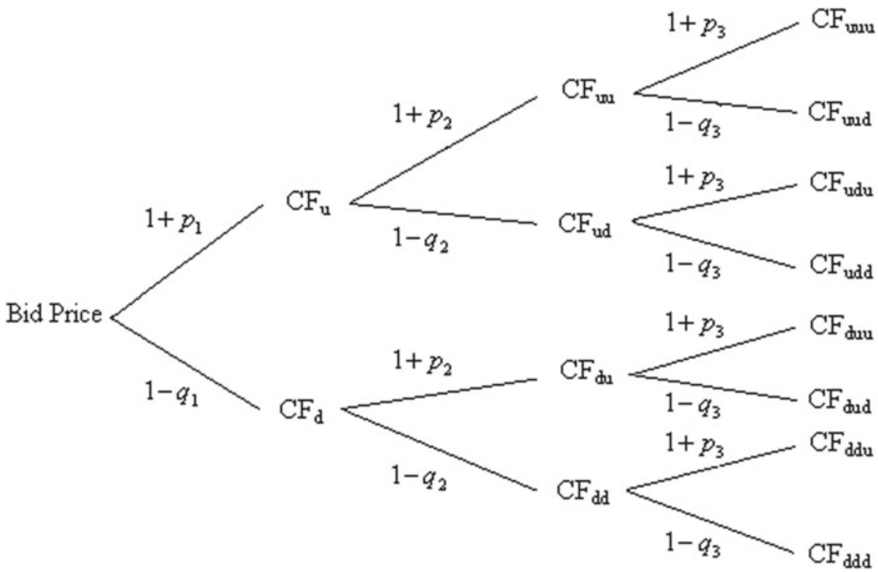

Fig. 5.1 Tree diagram of cash flow

5.5 Market Valuation of Cricketers

The value of each player was determined by using risk neutral real option valuation method on their performance index (P_i) for up state probability for each season. A player can score a maximum of 1 point on the basis of his performance. The performance indicator (P_i) for the i^{th} player shows the probability of an increase in the value of the player, and the remaining ($1–P_i$) is the probability of a decrease in the value of the player. The calculations for determining the present market valuation of the players are done by using the following equations.

$$\mathrm{PV_{uu}} = \frac{p \times \mathrm{CF_{uuu}} + q \times \mathrm{CF_{uud}}}{1 + \mathrm{rf}} \tag{5.14}$$

$$\mathrm{PV_{ud}} = \frac{p \times \mathrm{CF_{udu}} + q \times \mathrm{CF_{udd}}}{1 + \mathrm{rf}} \tag{5.15}$$

$$\mathrm{PV_{du}} = \frac{p \times \mathrm{CF_{duu}} + q \times \mathrm{CF_{dud}}}{1 + \mathrm{rf}} \tag{5.16}$$

$$\mathrm{PV}_{dd} = \frac{p \times \mathrm{CF}_{ddu} + q \times \mathrm{CF}_{ddd}}{1 + \mathrm{rf}} \tag{5.17}$$

$$\mathrm{PV_{u}} = \frac{p \times \mathrm{PV_{uu}} + q \times \mathrm{PV_{ud}}}{1 + \mathrm{rf}} \tag{5.18}$$

$$\mathrm{PV_d} = \frac{p \times \mathrm{PV_{du}} + q \times \mathrm{PV_{dd}}}{1 + \mathrm{rf}} \tag{5.19}$$

$$\mathrm{PV_0} = \frac{p \times \mathrm{PV_u} + q \times \mathrm{PV_d}}{1 + \mathrm{rf}} \tag{5.20}$$

where PV represents the present value of the players, CF represents the cash flow, $\mathrm{PV_u}$ and $\mathrm{PV_d}$ represent the upward present value and the downward present value of the players, and $\mathrm{PV_0}$ represents the neutral present market values of the players. Also, p and q $(= 1-p)$ represent the risk-neutral probability which is calculated by using Eq. (5.13) and assumed to be constant for IPL 2014, IPL 2015, and IPL 2016. The results of the above calculation using a binomial option-pricing model along with the performance scores of the players in said IPL seasons can be seen in Appendix 5.1.

5.6 Classification of Bid Price and Neutral Present Market Value

A meaningful classification for bid price and neutral present market value of the players based on players' on-field performance would be in terms of a suitable interval from an assumed probability distribution. To test the hypothetical distribution of bid price and neutral present market values of the players, the Kolmogorov–Smirnov (K–S) test is used. The test statistic is given by

$$D_{\alpha,n} = \max|S_n(x) - F(x)| \tag{5.21}$$

where $S_n(x)$ and $F(x)$ are empirical and theoretical distribution functions, respectively. However, for performing K–S test the theoretical distribution needs to be completely specified (i.e., the values of the parameters should be known) and so the parameters are estimated from the data. The critical value of D_n for α level of significance depends on the number of observations and may be denoted by $D_{\alpha,n}$. The interval $\left[F(x) - D_{\alpha,n},\, F(x) + D_{\alpha,n}\right]$ provides the 100(1−α)% confidence band for $F(x)$ that can be used to visualize the goodness of fit of $F(x)$. It appears appropriate to assume that both bid price and the neutral present value of the players follow a normal distribution since the p-values of the corresponding Kolmogorov–Smirnov tests are 0.579 and 0.316 respectively. It can also be visualized in the empirical distribution function (EDF), in the graphical depiction of Figs. 5.2 and 5.3 for bid price and neutral present market value, respectively, that the step function is lying within the upper and lower confidence bounds.

Now, if a random variable X follows normal distribution with mean (μ) and variance (σ^2), then the probability density function $f(x)$ can be written as

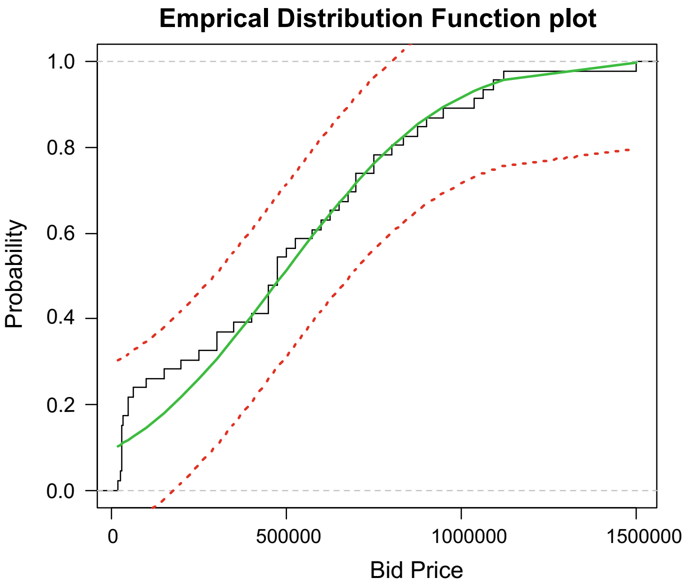

Fig. 5.2 Visualizing the goodness of fit of bid price

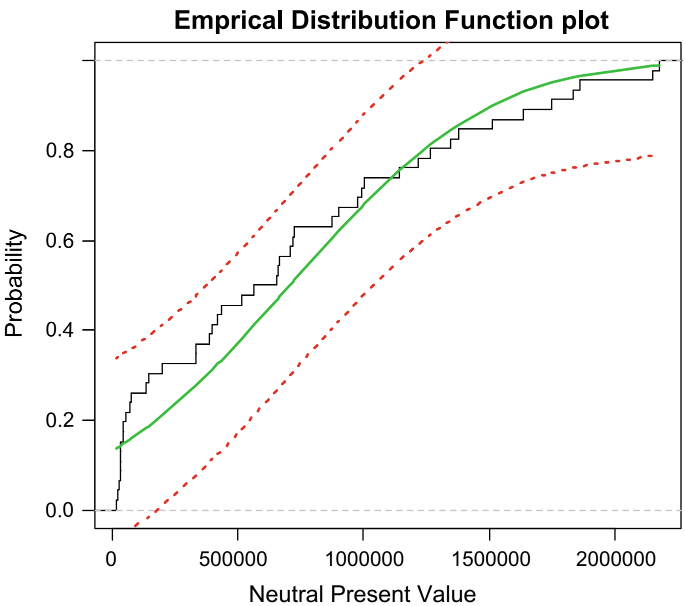

Fig. 5.3 Visualizing the goodness of fit of neutral present market value

Table 5.2 Interval for classification of bid price and neutral present market value

Classification	Bid price	Neutral present value
Very poor	Below 15,952,305.14	Below 11,583,911.49
Poor	15,952,305.14–393,387,402.49	11,583,911.49–36,401,977.58
Average	39,387,402.49–59,572,597.51	36,401,977.58–57,778,355.78
Good	59,572,597.51–83,007,694.86	57,778,355.78–82,596,421.86
Very good	83,007,694.86 and above	82,596,421.86 and above

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \quad \text{where } -\infty < x < \infty \tag{5.22}$$

One can find four real numbers x_1, x_2, x_3, and x_4 to divide the range $[-\infty, \infty]$ into five intervals $(-\infty, x_1)$, (x_1, x_2), (x_2, x_3), (x_3, x_4), and (x_4, ∞) with the same probability weight of 20%. These intervals are used to categorize the various stages of X as

(i) Very Poor if $-\infty < X < x_1$
(ii) Poor if $x_1 < X < x_2$
(iii) Average if $x_2 < X < x_3$
(iv) Good if $x_3 < X < x_4$
(v) Very Good if $x_4 < X < \infty$

The cut-off points x_1–x_4 that are obtained from the assumed normal distribution of both bid price and neutral present market value of the players are given in Table 5.2.

5.7 Results and Discussions

The bid price of the players in IPL 2014 and their corresponding neutral present market values at the end of IPL 2016 is provided in Appendix 5.1. The compensation package of the players is fixed for a period of three years. However, the neutral present values of the players are enhanced because of their performance, which is generally in the increasing note at the end of each season of IPL. The numerical value of the bid prices and the neutral present values shall not be compared but the categories to which they are classified may be. The bid price is the ultimate salary of the player that he receives during every IPL (i.e. once in each of the three seasons of IPL). However, the neutral present value is determined at the end of every three seasons of IPL combining the players' performance. The correlation between the bid price of the players and their neutral present value is as high as +0.89741 implying that there is a high positive relation of the bid price with the present value. However, the performance of the cricketers is not generally related to their compensation packages as evident from the low level of correlation between the bid price and performance levels during the three seasons of IPL (cf. Table 5.3).

Table 5.3 Correlation coefficients with bid price

Neutral present values	Performance score of IPL 2014	Performance score of IPL 2015	Performance score of IPL 2016
0.89741 (0.000)	−0.02852 (0.844)	0.09975 (0.496)	0.29337 (0.119)

In Table 5.3, the values in parenthesis imply the corresponding p-values associated with the correlation coefficient. It can be inferred that in IPL 2014, IPL 2015, and IPL 2016, the performance of cricketers is not significantly related to their bid price. As expected, the neutral present value has a high positive correlation with the bid price of the players. The classification of the players in terms of bid prices and the neutral present market values can be visualized in Fig. 5.4.

Neutral present value \ Bid Price	Very poor	Poor	Average	Good	Very good
Very good				R Ashwin, R Jadeja	G Maxwell, MS Dhoni, R Sharma, S Raina V Kohli
Good			K Sharma, P Chawla, S Watson	D Karthik, S Samson, Y Chahal	C Lynn, KL Rahul, M Pandey
Average		Shakib Al Hasan	A Mishra, K Nair, K Pollard, S Dhawan, U Yadav	A Finch, K Jadav, R Uthappa	
Poor		Rayudu, Cutting, McCullum, Gayle, Miller, Gambhir, H Singh, Jhonson, Shami, M Sharma, S Naddem, Y Pathan, Y Singh	A Rahane, W Saha		
Very poor	H Patel, Duminy, M Vohra, M Singh, M Tiwary M Agarwal, S Binny	F du Plessis, M Vijay, P Patel			

Fig. 5.4 Semi-graphical display for classification of players

In the semi-graphical display (cf. Fig. 5.4), it can be observed that there are 25 cells in all. The rows are used to represent the neutral present values of the players, and the columns are used to represent their bid prices. The five rows represent the five levels of neutral present values, and similarly, the five columns represent the five levels of bid prices as mentioned in Table 5.2.

The cells pertaining to the diagonal are recognized in the thick black-bordered box. These cells are containing those players who had their bid prices and neutral present values at the same level. For example, Harshal Patel, J. P. Duminy, Manoj Tiwary, etc., had a very poor bidding price, and their neutral present value is also very poor. Likewise, Ambati Rayudu, Chris Gayle, Gautam Gambhir, Yuvraj Singh, etc., had received a poor bid and so is their neutral present value. The other players whose names can be seen in the semi-graphical display in black-bordered boxes can be similarly remarked.

The players whose name appeared above the black-bordered boxes have made a considerable increase in their neutral present values compared to their bid prices. This means that the players belonging to this group have performed much better than the compensation packages paid to them. For example, Shane Watson, Piyush Chawla, and Karan Sharma got average bid prices, but they have performed well to attain good neutral present values. Ravindra Jadeja and Ravichandran Ashwin received good bid price, but their performance was at the superlative level, so they can take their bid price to the highest level.

The remaining players considered for the study can be seen below the black-bordered box (cf. Fig. 5.4) implying a lower level of neutral present value compared to their bid prices. Of the eleven players named in the list, F. du Plessis, M. Vijay, and P. Patel are the poorest of performers, considering their bid prices. While K. L. Rahul and Manish Pandey had enjoyed a very good bid price and realized a good present value. Players like R. Uthappa and A. Rahane could only manage average and poor present value, respectively, after attaining a good and average bid price, respectively. It appears from the performance of these eleven players that they are unable to justify their compensation packages by their on-field performances that led to diminished present value.

5.8 Conclusion

The salaries of the professional cricket players that are decided through an auction in the Indian Premier League (IPL) are a way of quantifying players' performance from a financial point of view. As an outcome of the study, one can find out those players who have justified their salaries. Some players have performed much better than what they are paid. They are R. Ashwin, R. Jadeja, P. Chawla, S. Watson, and Shakib Al Hasan. Such players shall be offered higher perks in the upcoming season of IPL. However, players like M. Vijay, P. Patel, A. Rahane, W. Saha, R. Uthappa, K. L. Rahul, M. Pandey, and C. Lynn made a poor display of their skills in the said IPL seasons compared to their bid price. Thus, the franchisee may take a note of that and offer a lower bid price to the said players before making any fresh agreement with them. The remaining set of players had performed as expected from them, in the sense that, their bid price and the neutral present value are at the same level.

Similar studies can be performed to derive the neutral present value of the whole cricket team using real option valuation method by extending this work. The proposed model of performance measurement can also be used in several other professional sports. As the idea of IPL works on any franchisee-based system of hiring players and transfers, the same model is applicable to all such sports like football, rugby, baseball where such a system exists. This shall be helpful to the franchisee to decide about which players to be considered and who are to be dropped for a given price. The players can also use this model to understand what their market price shall be and that they are not underpaid.

Appendix 5.1: Performance Scores, Bid Prices and Neutral Present Values of the Players

Sl. No.	Players Name	Performance scores in			Bid price	PV_0
		IPL 2014	IPL 2015	IPL 2016		
1	Aaron Finch	0.356	0.310	0.539	62,000,000	40,094,918.18
2	Ajinkya Rahane	0.345	0.372	0.614	40,000,000	27,467,554.14
3	Ambati Rayudu	0.371	0.478	0.624	22,000,000	17,471,902.52
4	Amit Mishra	0.514	0.635	0.222	40,000,000	42,835,906.94
5	Ben Cutting	0.743	0.499	0.909	22,000,000	25,423,495.31
6	Brendon McCullum	0.384	0.308	0.810	36,000,000	23,974,429.98
7	Chris Gayle	0.275	0.418	0.387	20,000,000	13,262,912.7
8	Chris Lynn	0.245	0.421	0.554	96,000,000	61,395,092.83
9	David Miller	0.495	0.288	0.283	30,000,000	21,903,461.65
10	Dinesh Karthik	0.363	0.496	0.601	74,000,000	59,228,948.74
11	Faf du Plessis	0.361	0.316	0.842	16,000,000	10,491,275.27
12	Gautam Gambhir	0.363	0.221	0.928	28,000,000	16,246,004.76
13	Glenn Maxwell	0.626	0.844	0.592	90,000,000	126,693,012.2
14	Harbhajan Singh	0.832	0.587	0.251	20,000,000	26,934,016

(continued)

(continued)

Sl. No.	Players Name	Performance scores in			Bid price	PV_0
		IPL 2014	IPL 2015	IPL 2016		
15	Harshal Patel	0.750	0.908	0.567	2,000,000	3,273,149.442
16	JP Duminy	0.875	0.259	0.890	10,000,000	9,719,699.361
17	Karn Sharma	0.877	0.506	0.481	50,000,000	64,439,156.31
18	Karun Nair	0.391	0.367	0.577	56,000,000	40,323,706.58
19	Kedar Jadhav	0.377	0.149	0.627	78,000,000	41,417,406.61
20	Kieron Pollard	0.579	0.306	0.617	54,000,000	43,735,230.24
21	KL Rahul	0.252	0.506	0.745	110,000,000	77,608,216.91
22	Manan Vohra	0.381	0.230	0.407	11,000,000	6,590,662.881
23	Mandeep Singh	0.141	0.376	0.892	14,000,000	7,341,160.091
24	Manish Pandey	0.398	0.351	0.625	110,000,000	78,333,083.12
25	Manoj Tiwary	0.329	0.221	0.683	10,000,000	5,580,269.012
26	Mayank Agarwal	0.269	0.307	0.560	10,000,000	5,789,653.771
27	Mitchell Jhonson	0.817	0.233	0.845	20,000,000	17,980,976.58
28	Mohammed Shami	0.514	0.500	0.726	30,000,000	28,332,147.12
29	Mohit Sharma	0.572	0.478	0.460	24,000,000	23,445,334.15
30	MS Dhoni	0.587	0.568	0.458	150,000,000	162,004,519.1
31	Murali Vijay	0.287	0.199	0.432	20,000,000	10,276,715.99
32	Parthiv Patel	0.309	0.331	0.461	17,000,000	10,652,805.92
33	Piyush Chawla	0.876	0.668	0.650	42,000,000	62,772,736.08
34	Ravichandran Ashwin	0.798	0.738	0.857	76,000,000	113,533,669
35	Ravindra Jadeja	0.922	0.829	0.866	70,000,000	123,006,542.6
36	Robin Uthappa	0.475	0.392	0.680	64,000,000	51,831,478.95
37	Rohit Sharma	0.942	0.379	0.832	150,000,000	176,936,833.1
38	Sanju Samson	0.359	0.418	0.333	80,000,000	58,806,790.97
39	Shahbaz Nadeem	0.555	0.383	0.742	32,000,000	27,773,953.65
40	Shakib Al Hasan	1.000	0.823	0.674	20,000,000	36,905,342.53
41	Shane Watson	0.916	0.861	0.879	40,000,000	71,688,664.56

(continued)

(continued)

Sl. No.	Players Name	Performance scores in			Bid price	PV_0
		IPL 2014	IPL 2015	IPL 2016		
42	Shikhar Dhawan	0.371	0.446	0.766	52,000,000	39,932,859.89
43	Suresh Raina	0.679	0.427	0.537	110,000,000	111,968,385.9
44	Stuart Binny	0.883	0.219	1.000	5,000,000	4,630,241.026
45	Umesh Yadav	0.569	0.627	0.833	42,000,000	47,071,505.33
46	Virat Kohli	0.372	0.460	0.653	170,000,000	132,568,614
47	Wriddhiman Saha	0.416	0.296	0.539	50,000,000	33,996,235.74
48	Yusuf Pathan	0.422	1.000	0.614	19,000,000	24,439,254.11
49	Yuvraj Singh	0.934	0.202	0.624	20,000,000	18,760,600.57
50	Yuzvendra Chahal	0.696	0.512	0.222	60,000,000	67,617,801.41

References

Bracewell, P. J., & Ruggiero, K. (2009). A parametric control chart for monitoring individual batting performances in cricket. *Journal of Quantitative Analysis in Sports, 5*(3), 1–19.

Dalmia, K. (2010). The Indian premier league: Pay versus performance, an honors thesis. In N. Leonard (Ed.), *Stern school of business*. New York: New York University.

Depken, C. A. & Rajasekhar, R. (2010). Open market valuation of player performance in cricket: Evidence from the indian premier league. Available at: http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID1593196_code1712.pdf. Accessed on 29 Apr 2010.

Hayden, M. (2011). *Standing my ground* (p. 346). India: Harper Collins Publishers.

Iyenger, N. S. & Sudarshan, P. (1982). A method of classifying regions from multivariate data. *Economic and Political Weekly* Dec 18, pp. 2048–2052.

Kitchin, P. (2008). The development of limited overs cricket: London's loss of power. *London Journal of Tourism, Sports and Creative Industries, 1*(2), 70–75.

Mitra, S. (2010). The IPL: India's foray into world sports business. *Sport in Society, 13*(9), 1314–1333.

Pandey, I. M. (2004). *Financial management*. New Delhi: Vikas Publishing House Pvt. Ltd.

Quirk, J., & Fort, R. D. (1992). Pay dirt: The business of professional team sports. Princeton, NJ: Princeton University Press.

Rastogi, S. K., & Deodhar, S. Y. (2009). Player pricing and valuation of cricketing attributes: Exploring the IPL Twenty20 vision. *Vikalpa, 34*(2), 15–23.

Suleman, M. T. & Saeed, M. A. (2008). Option on human performance: A case study of indian premier league. Available at: http://papers.ssrn.com/sol3/…/SSRN_ID1474093_code1208558.pdf. Accessed on 21 December 2009.

Vasu, A. (2011). Return of the league. *Sports Illustrated India, 2*(7), 52–55.

Chapter 6
Impact of Age on Performance of the Cricketers

6.1 Introduction

The natural process of getting older has a significant impact on everyone, and so the effects of aging on sportspersons are no exception. As per the demand of sports like hockey, soccer, basketball, the physical aspect of players dominates their performance. It has been widely reported as relative age effect (RAE) in sports. The presence of age effect on the performance of the cricketers is also conveyed continuously in different newspapers and magazines. In cricket, physical as well as the mental ability and technical skills separate the best-performing players from the rest. In Indian Premier League (IPL), as the franchisees formed their teams by competitive bidding—both retired players and emerging talented players get a platform to perform and learn from each other. The former Australian spin bowler Shane Warne mentioned that the IPL is a great platform for forthcoming players rather than the veteran cricketers[1]. IPL certainly helped some young cricketers to make a mark (Raman, 2009a). It is indeed true that a handful of young players (e.g., P. Valthaty, S. Marsh, S. Hasan, A. Rahane) have been able to draw the attention of selectors and onlookers toward themselves through their noteworthy performances in IPL.

The Twenty20 cricket is not particularly taxing to a professional sportsperson or even someone who has crossed age 40 and is in the second life of his career (Krishnaswamy, 2010). However, the performances of senior players are noteworthy in all the seasons of IPL played so far. The best example is that of Shane Warne, the former captain of Rajasthan Royals (RR) in IPL till 2011. This second highest wicket-taker in test cricket history went on to have a superb tournament in 2008, who led his team to win the IPL with some innovative captaincy (Mitra, 2010). Though Shane Warne's performance in the second and third seasons of IPL was less remarkable, there were other players, viz. Adam Gilchrist, Mathew Hayden, etc.,

[1] Twenty20 great platform for youth, retrieved from http://www.iplt20.com/news/index.php?id=585 on February 9, 2010.

H. Saikia et al., *Cricket Performance Management*, Indian Statistical Institute Series,
https://doi.org/10.1007/978-981-15-1354-1_6

who had shined. The performances of these veteran cricketers in IPL refuted that Twenty20 cricket is a game young cricketers' only.

The senior players mark a massive impression in the IPL usually due to their skills. This impression helped to make them everyday cricketers unlike the ones who fire only for a while (Raman, 2009b). This has also been shown by players like S. Tendulkar, R. Dravid, S. Warne, A. Kumble, D. Vettori. Though the players like S. Tendulkar, R. Dravid have not played much of the Twenty20 format of cricket, their skill sets are so comprehensive that they can crack the code in no time (Moonda, 2011). The performances of skilled players in Twenty20 cricket were a lesson for youngsters that there is no substitute for skill if anyone is to succeed in any format of the game. The skills required for batsmen are that they must keep their heads, choose the right shot, and resist pace, swing, and wrist spin as well as for bowlers that they must move the ball around, possess a wide range of deliveries, and retain control under intense pressure (Roebuck, 2008). However, as a player gets older, he becomes less fit for the game and more prone to injury (Lucifora and Simmons, 2003). Thus, because of age effect, the demand for players like S. Warne, A. Gilchrist, S. Ganguly, S. Tendulkar, and M. Hayden may have failed and the fact that they were no longer playing for their national side (Mitra, 2010). These discussions set the background of this chapter, which aims to test 'Is Twenty20 cricket a game for young cricketers?'

6.2 Age Effect on Players' Performance in Different Sports

Relative age effect in sport is a worldwide phenomenon, and it exists in many, but not all, competitive sports (Musch & Grondin, 2001). The empirical evaluation of the aged effect on performance is indeed challenging in sports. It is well understood that the performances of the players reach uttermost at a given age and then show decreasing, but the determination of the peak performing years of the players is always a challenging task. No one had an answer to the question of how to determine the peak performance age of the players over the time period. Also, the level of performance of the players in a team game can vary depending on the strength of the opponents and different playing conditions. Therefore, the determination of peak performance age for each player of the teams over chronological time is not an easy task.

Meanwhile, the effect of age has been studied by various authors in different kinds of sports. Most of these studies revealed linear, curvilinear, or exponential trends when modeling effects of aging (Young & Weckman, 2008). Linear trend was found when evaluating the effects of aging on freely chosen walking speed (Himann, Cunningham, Rechnitzer, & Paterson, 1988). Schulz and Curnow (1988) had examined the age of peak performance in a broad range of athletic events. They had noted that the absolute levels of peak performance among superathletes have increased vividly but the stability of peak performance could not be ascertained. The curvilinear trend was observed when investigating competitions in indoor rowing events (Seiler, Spirduso, & Martin, 1998), and in freestyle swimming performances, it was found decreased exponentially (Tanaka & Seals, 1997). Percentage decline in master's superathlete with increasing age in track and field performance was examined by Baker, Tang, and Turner (2003). They found that track running records declined with aging in a curvilinear fashion as $y = 1 - \frac{\exp(T - T_0)}{\tau}$, whereas in field events it declined in a linear way as $y = \alpha[T - T_0]$. Also, they had reported that decline with aging was greater for females and longer running events. In the case of baseball, Fair (2008) estimated the effects of age using nonlinear fixed effect regression and found that aging effects are larger for pitchers than for batters. The peak age of performance for professional baseball pitchers and batters are 26 and 28 years, respectively. In many sports, the classification system (i.e., senior, young, peak performance age, through cutoff values) based on biological age is difficult to organize (Musch & Grondin, 2001). The effect of aging for professional football players using performance-aging curves was evaluated by Young and Weckman (2008). Addona and Yates (2010) examined the relative age effect (RAE) in the National Hockey League (NHL). An analysis of master athletes in running, swimming, and cycling was performed by Ransdell, Vener, and Huberty (2009) by age-group and gender. They had also examined how physiological, sociological, and psychological factors affect master-level athletes' performance in the USA. Lehto (2015) investigated the age-related changes in the endurance performance among male amateur marathon runners from 1979 to 2014. He found a quadratic relationship $t = a + bx + cx^2$(running time t as a function of age x). The fitted quadratic model indicates that the marathon performance of the average runner improves up to age 34.3 ($\pm$2.6) years, and thereafter, the performance starts to decline. A similar study was performed by Radek (2014) to evaluate peak performance age in track and field athletes. The study includes a total of 6314 athletes (3474 male and 2840 female) from World Championships, European Championships, and Olympic Games. He found that the peak performance age for male and female are 25 and 26, respectively.

Though several measures of performances were reported in the literature surrounding the game of cricket, the impact of aging on players' feats did not enrich our search. However, a study was found in terms of cricket by Hazra & Biswas, (2018) to compare the mental skill ability of the cricket players as per different age-level categories. They performed one-way analysis of variance (ANOVA) and revealed that age may be the predictor of mental skill ability of the cricket players. In cricket, it is difficult to generalize a relation between aging and performance, as cricketers with different expertise attain their peak performance at different ages. Batsmen tend to reach their peak in their late twenties after their technique has matured through experience and conversely, and fast bowlers often reach their peak in between early to mid-twenties when they are at the height of their physical capacity. Other bowlers, mostly spinners, even fast bowlers who can 'swing' the ball, are most effective in the later part of their career (Saikia & Bhattacharjee, 2011).

In most sports, if we look at the career of a thriving sportsman, the general trend is that in his/her initial days he/she is recognized in the sports arena as a good performer and then with time he/she gradually improvises himself/herself and reaches to his physical pinnacle and also to the highest level of performance in the career. Then with maturity, the skill and experience of the player increase and the player performs at the highest level successfully. After that, at some stage of his/her career, the aging of the player has its impact on the performance level of the players. The skill and experience are not complemented sufficiently by the physical ability, and the performance level shows a downward trend. This is the time when players think of retiring from sports or are not considered any further by the selectors to represent their team. Several players at this stage of their career change their role and become coach or mentors or TV commentators or settle down in other professions. But there are players who do not abide by such average laws and have a different story to tell. Some players start their career with a dramatic entry and then after some initial successes get lost once for all, and there are others who keep playing successfully as if the age for them is just a number. Both these cases are outliers (not statistical outliers) in the sea of sportsmen that we come across in different ball games. In cricket too, such outliers do exist.

6.2.1 A Case Study on the Test Career of Irfan Pathan

The first case that we are going to take up is that of Irfan Pathan born on October 27, 1984, and hailing from Baroda, Gujarat; this left-handed all-rounder got into the Indian test team at the age of 19 directly from the under-19 squad with only one year experience of playing first-class cricket. He played the first test for India in December 12, 2003, at Adelaide in Australia and the test in which India made history by defeating mighty Australia in their soil after 23 years. Though not much in terms of dismissals, Pathan kept his mark as a bowler in the series with the ability to swing the ball both ways and also swing the old ball. He gained the responsibility of opening the bowling for India in his first test. His ability both in batting and in bowling reminded many of Kapil Dev, and fans thought that the draught of having an all-rounder in Indian cricket is going to be over soon. Being a left-handed all-rounder with the ability to swing the ball both ways, some compared him with the greats like Wasim Akram. He had got success in other formats of cricket as well, both with the bat and the ball. He was the man of the finals of the Twenty20 World Cup in 2006 at South Africa which India won. In test cricket, he became a regular member in the Indian squad. Starting from his first test match in December 2003 till June 2006, he missed only four test matches out of the 29 test matches played by India during that period. But soon his injuries and ups and downs in his form started. He was out of the test team for the first time in June 2006 and then again reconsidered in December 8, 2007. A few months later he played his last test on April 3, 2008, in Ahmadabad, against South Africa. Since then almost twelve years have passed, there is no sign of any comeback by him to the Indian test side. The player who played his first test at the age of 19 for India was never seen again in the test side. He played his last test for India before he was 24 years old, an age at which many players make their test debut.

The most talked about performance of Pathan in test cricket was his hattrick against Pakistan in the third test of the series of India's tour to Pakistan in 2006. He took the wickets of batting greats like Salman Butt, Younis Khan, and Mohammad Yousuf in the fourth, fifth, and sixth deliveries of his first over as well as the first over of the test. That is still a world record to have a hat trick in the very first over of a test match. Another of his remarkable achievement in test cricket was his all-round performance at the end of his test career at Perth against Australia in 2008. With 28 and 46 runs with his bat and 5 wickets (both Australian openers in both innings), he paved the way to victory for India and for him to the man of the match award. Pathan's career is an interesting case to study. At the age in which many cricketers start their test career, Pathan played his 29th and the last test of his life. Even a former national lady cricketer from Gujarat, Dr. Tanveer Seikh (Assistant Professor in Physical Education) has written her Ph.D thesis leading to the identification of the reasons for the untimely end of the cricket career of Irfan Pathan. She has identified excessive cricket and injury as the main reasons for the downfall of Irfan (Table 6.1).

If we look at the career graph of Irfan Pathan, it is not the way we expected. We thought that it would be high at the beginning and low at the end. But if we plot the number of wickets taken by Irfan Pathan in the different test matches played by him, we can see something like a cyclical pattern (cf. Fig. 6.1).

But definitely, a decreasing trend is also visible toward the end of his career. During the first 5 test matches of his career, he claimed 16 wickets, and in the last 5 tests of his career, he claimed 11 wickets.

Figure 6.2 shows us that Pathan matured as a batsman in the latter half of his career. Though, initially, he made his entry into the team as a swing bowler in due course, he identified his batting skills and surely he did not excel in a similar manner as a bowler. Being a better bowler, everyday was expected from him. India being a batting heavy team and comparatively more in need of pace bowlers, the expectations of the selectors from Pathan in terms of bowling skills were more than what he performed. This might have led to the untimely end of the test career of Pathan. To understand the career bowling performance of Pathan, we compute the combined bowling rate (Lemmer, 2005) for each test and plot it in Fig. 6.3. To normalize the peaks and troughs, we simultaneously plotted a three-point moving average of the values of CBR. Remember, that CBR combines the economy rate, bowling strike rate, and the bowling average of the bowler and is a negative measure. The lower the value better is the bowler. The moving average of the CBR values in the graph tells the story of Pathan's career performance; when Pathan as a bowler is expected to reach to his physical prowess and highest level of maturity given his international exposure from an early age, the curve remained similar to what it was at the beginning of his career. The moving average of the CBR curve of a successful career is expected to dip further down and move closer to x-axis than what happened in case of Pathan.

6.2.2 A Case Study on the ODI Career of Sachin Tendulkar

Any game is always bigger than the players who play it—how much talented a player may be. But all such popular sayings fall flat when the game is cricket and name of the player is 'Sachin Ramesh Tendulkar' (SRT). The followers of SRT recognized him as 'God of Cricket' not only in India but worldwide. He has been a reason behind

Table 6.1 Career Statistics of Irfan Pathan in test matches

Batting statistics: total tests 29 debut test: Dec 12–16, 2003 last test: Apr 3–5, 2008

Inn.	Not out	Runs	HS	Avg.	SR	50s	100s	4s	6s
40	5	1105	102	31.57	53.22	6	1	131	18

Bowling statistics

Matches	Overs	Runs conc.	Wickets	Bowl avg.	Econ. rate	Bowling SR
29	980.4	3226	100	32.26	3.28	58.8

Source www.espncricinfo.com

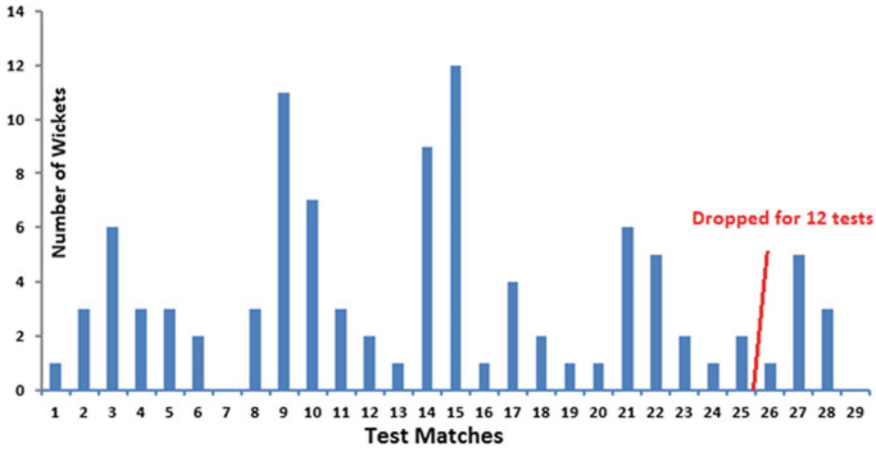

Fig. 6.1 Wickets taken by Irfan Pathan in different test matches

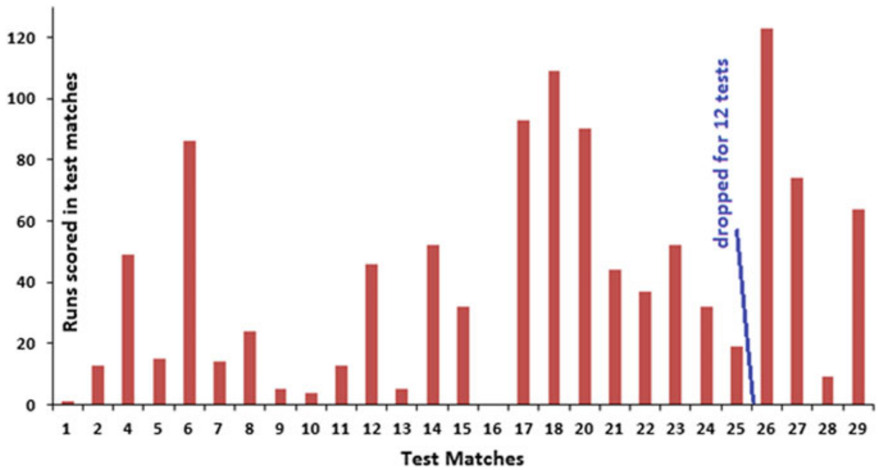

Fig. 6.2 Runs scored by Irfan Pathan in different test matches

the eruption of the popularity of the game of cricket in India. His performance as well as his achievements, consistent over more than two decades, made him a giant player regardless of age. He was a complete batsman with perfect balance in stroke-making and staying in the crease. This quality of his made him the prolific run-maker of all times and surely the biggest cricket icon around the world.

He was only 16 years old, when he made his test debut against Pakistan. His debut in the year 1989 was a blazing introduction to international cricket as he had to face the might of Wasim Akram, Imran Khan, and Waqar Younis in their own courtyard. He was hit on the face by Waqar Younis but continued to bat in a blood-soaked

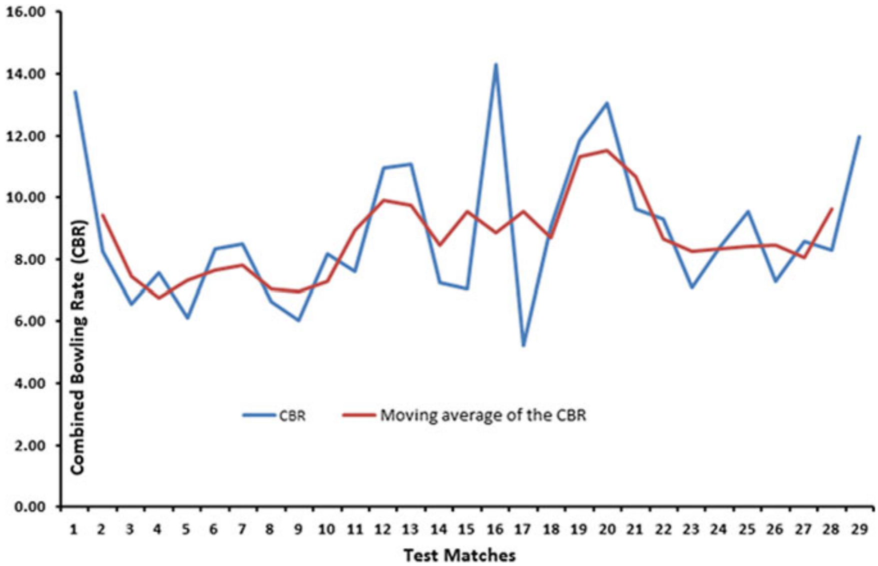

Fig. 6.3 Combined Bowling Rate (CBR) in all test matches

shirt. His first, as well as match-saving test hundred at Old Trafford, came up against England when he was 17 years old. Later on, he scored a few centuries during his tours to England and Australia, with the national team. His finest performance during his initial days was acknowledged by cricket experts, when he made a century as 19 years old on a lightning fast pitch at the WACA in Perth, against the tremendously dominant team Australia of those days. After those innings, the cricket experts of both the countries (England and Australia) agreed that a batting genius was born.

In 2000, he became the first batsman to have scored 50 international hundreds. In the year 2008, he had crossed the record of Brian Lara as the leading test run-scorer and then after a few years he went past 13,000 test runs, 30,000 international runs, and 50 test hundreds. He holds currently the remarkable record of most hundreds in both tests and ODIs. He retired from test cricket after a memorable 200th test on his home ground at the Wankhede Stadium against West Indies. The most interesting fact is that—though he retained a heavenly enthusiasm for the game till his last match he scored his first ODI hundred in the 79th match of his career. Like test debut, his ODI debut was also against Pakistan on December 18, 1989. He played a total of 463 ODI matches in his career with 18,426 runs and one double century (200) against South Africa in Gwalior, that too at the age of 36 years. He was the first player who broke the individual double hundred barriers in ODI in 2008. In the year 2012, he became the first player to score 100 international centuries, a mind-blowing achievement in the history of cricket (Table 6.2).

As a captain of the Indian cricket team, Sachin Tendulkar was not very successful during his two tenures. In 1996, when he took the charge of captaincy, the expectation was very high with enormous hopes. Yet the team was showing very poor

Table 6.2 Career statistics of Sachin Tendulkar in ODI matches

Batting statistics: total ODIs 463 debut ODI: December 18, 1989 last ODI: March 18, 2012									
Inn.	Not out	Runs	HS	Avg.	SR	50s	100s	4s	6s
452	41	18,426	200*	44.83	86.23	96	49	2016	195
Bowling statistics									
Matches		Balls	Runs conc.		Wickets		BA	ER	BSR
463		8054	6850		154		44.48	5.10	52.2

Source www.espncricinfo.com
*means "Not out"

performance by 1997. During his captaincy, out of 73 matches, the Indian team won only 23 matches and lost 43 matches with a winning percentage of 31.5. His batting performance also deteriorated when he was the captain. His batting average was only 36.33 because of captaincy pressure; it is very low as compared to his career batting average of 44.83. Thus, after a test series defeat under his captaincy at home against South Africa, Tendulkar resigned and Sourav Ganguly took over the charge of captaincy for team India. In May 2005, Greg Chappell was appointed as the coach of Indian national cricket team for a two-year term until the 2007 World Cup. After Chappell took the charge as a coach, he was interfering with an impeccable batting line up of team India through his heretical cricket coaching methods. Thus, several senior players, including Sachin Tendulkar, were not happy with his coaching styles. As a result, team India packed up from the first round of World Cup in 2007. Though his performance waned (as batting average was 34.96, much less in comparison with his career batting average of 44.83) during the period of Greg Chappell after that, he had shown his ability and proved that age is just a number. In 2011, when team India won the world cup, he was the ambiance for the teammates to make his dream come true. He had accomplished remarkable 100 international centuries in 2012, just before the completion of his 39th birthday and created history for the world cricket. In the year 2013, he retired from international cricket. After his retirement, the Indian government conferred him the highest civilian award '*Bharat Ratna*' for the first time to a sports person.

If we look at the career graph of Sachin Tendulkar, it is not the situation as we expected. Generally, a cricketer who plays for a long duration scores more runs at the beginning of the career and toward the end of his career. The scores show a declining trend. If we plot the number of runs scored by Tendulkar in all the ODI matches played by him, we shall get the following graph (cf. Fig. 6.4). In the average runs

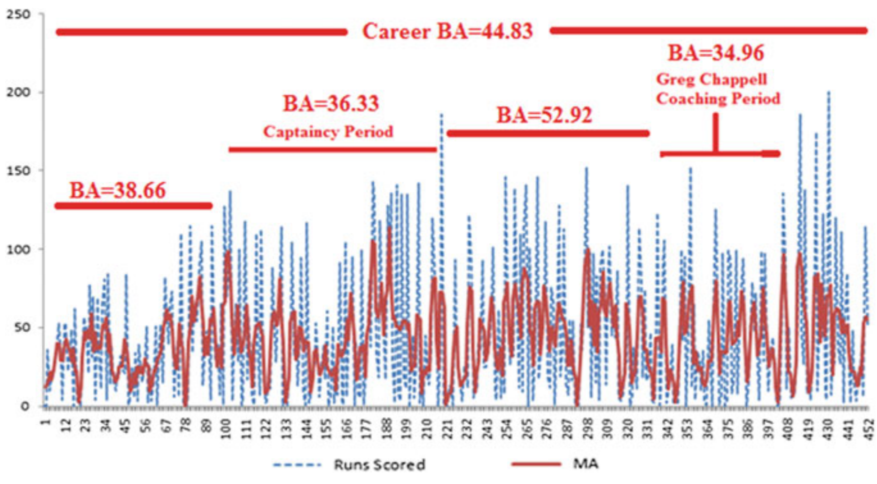

Fig. 6.4 Runs scored by Sachin Tendulkar in different ODI matches

shown in the graph, for not out innings, runs scored by SRT were adjusted using the method developed by Damodaran (2006). The details of this method can be seen in Appendix. Here again, Tendulkar has shown that he was a genius in his chosen sport.

6.2.3 A Case Study on the ODI Career of Dinesh Karthik

Apart from standing behind the stumps, the role of a wicket keeper has long been undecided. For example, where should he bat and how shall he approach the game while batting? Should he open the innings? Should he be aggressive with his bat? Or should he bat somewhere in the lower middle and aim at adding runs and extend the innings as long as possible? etc. Earlier, the wicket keeper of a team would occupy the number six or seven positions of the batting lineup. Then his prime responsibility was to take catches and execute stumpings when in the field rather than contribute with the bat. However, as the years passed, things started to change especially after watching Adam Gilchrist's way of being a wicket keeper. He was a great wicket keeper and a devastating opener, of the likes of Sir Vivian Richards, who tore the bowling of the opposition apart, especially taking the full benefit of the powerplay overs. As spectators believed that the ability to score runs as a wicket keeper is equally important with his ability to successfully stand behind the stumps. This had already been proved by a few new generation wicket keepers apart from Adam Gilchrist, viz. M. S. Dhoni, Kumar Sangakkara, Brandon McCullum, etc.

As an important case, we introduce here the story of Dinesh Karthik, a player whose main role was to be with the team, waiting in the wings. This wicket keeper–batsman was born on June 1, 1985, in Chennai, Tamil Nadu. He made his debut for the Indian national cricket team on September 5, 2004, against England. The reason for his selection in the national team was his remarkable performance in the 2004 under-19 World Cup in Bangladesh. Soon after his debut in ODI matches, he got the chance to start his test career against Australia on November 3, 2004. He had all the potentials with his technical nuances to be a reliable player in both wicket keeping and batting departments. He proved his ability in both the departments during the three tests series against England in 2007 and India experienced a historic test victory on foreign soil. However, despite his prolific performance in the domestic circuit, his international career never really took off. Now with 14 years from his debut, there is hardly any hope for him to become a regular in the Indian side.

Before Karthik capitalized his chance and became a regular member of the Indian team, India found a giant player Mahendra Singh Dhoni (MSD), who could score runs consistently through his bat and took due care of the wicket from behind the stumps. After the emergence of MSD in the Indian team, Karthik's international career became irregular. It can be placed as Dhoni's gain was Dinesh Karthik's loss. Being a superlative wicket keeper–batsman, MSD blocked the position of not only Dinesh Karthik but several other promising wicket keepers like Parthiv Patel, Deep Dasgupta, Wriddhiman Saha who were in and out of the Indian team. Both Dinesh and Parthiv started their career in such an unfortunate time when the place of MSD in the team was fixed not only as wicket keeper–batsman but also as a captain of the team. However, Karthik got the chance to come back into the national team occasionally, usually when MSD missed a match or a series of matches and sometimes as a specialized batsman. Despite his in and out of the national squad and with the limited opportunities he got, he had performed splendidly behind the stumps as well as with the bat (Table 6.3).

The trend of in and out from the national team continued for him, as he waited for his chance for two years after 2014 (i.e., 2015–16) and was again recalled to the national team in the year 2017. He was possibly the only player who was overlooked at the highest level of cricket as a regular player but continues to be the ideal replacement player for India for such a long period of 14 years. If we look at Fig. 6.5, we could see the gap period for Dinesh Karthik with number of matches for which he was dropped from the Indian squad. It is hardly a surprise that, with opportunities so few and far between, Karthik failed to excel at the highest level.

6.3 Data and Performance Measure

The data relevant to the age and performances of the players in all the seasons of IPL played so far (i.e., from 2008 to 2018) was collected from the website www.espncricinfo.com. It is necessary that players' statistics for a large number of games should be considered to measure the performance of players. The actual quality of a player may not be properly judged from one or two games. Thus, some selection condition needs to be set up while considering the players. Accordingly, the players are selected based on the following criteria.

- The player who was in the playing eleven in at least five matches in IPLs.
- Either the player has bowled at least 10 overs in IPLs as a bowler or has faced at least 100 balls in IPLs as batsman or both.

There were 776 players who had satisfied both the criteria. All these players were considered for the study. The performance measure of the i^{th} player is computed using the same procedure as defined in Chap. 3 and it is given by

$$S_i = S_{i1} + S_{i2} + \delta_i \tag{6.1}$$

where $\delta_i = \begin{cases} S_{i3}^{a_i} + S_{i4}^{1-a_i} - 1, & \text{if } i^{\text{th}} \text{ player is either a bowler or wicket keeper} \\ 0, & \text{if } i^{\text{th}} \text{ player is neither a bowler nor wicket keeeper} \end{cases}$

and a_i is an indicator variable with,

$$a_i = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ player is a bowler} \\ 0, & f\ i^{\text{th}} \text{ player is a wicket keeper} \end{cases}$$

Table 6.3 Career statistics of Dinesh Karthik in ODI matches

Batting statistics: total ODIs: 80 debut ODI: September 5, 2004 Last ODI: July 17, 2018

Inn.	Not out	Runs	HS	Avg.	SR	50s	100s	Ct	St
68	17	1517	79	29.74	73.03	9	0	53	7

Source www.espncricinfo.com

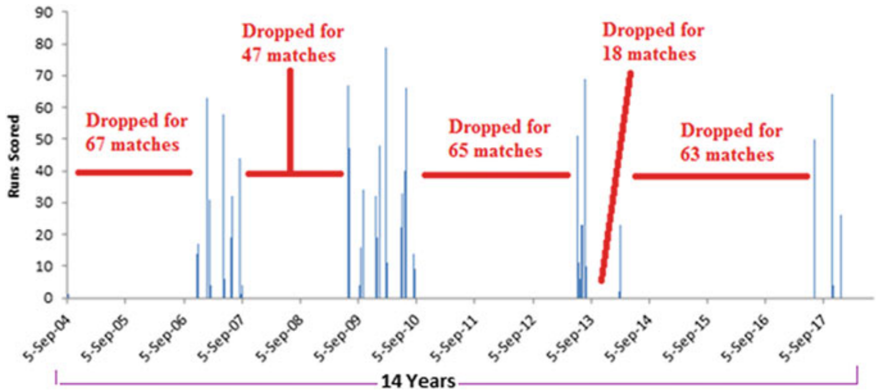

Fig. 6.5 Runs scored by Dinesh Karthik in ODI Matches till 2018

with S_{i1} be the performance score for batting, S_{i2} be the performance score for fielding, S_{i3} be the performance score for bowling, and S_{i4} be the performance score for wicket keeping. The factors that are considered to measure the batting performance of the players are *batting average*, *batting strike rate,* and the *average percentage of contribution to the team total*. For measuring fielding performance, the factors that we have considered are *number of catches taken* and *number of run outs executed*. For bowling performance measure of the players, the factors are *bowling strike rate*, *economy rate,* and *bowling average*. Along with the factors that are considered under batting performance measure and to measure the performance of wicket keepers, the additional factors that we have considered are a *number of catches taken*, *a number of stumpings,* and *number of bye runs conceded*. The normalization procedure of different factors along with their weights under the ability of batting, fielding, bowling, and wicket keeping is calculated following the same way as already discussed in Chap. 3 (cf. subsection of 3.6.5). The performance score of S_{i1}, S_{i2}, $S_{i3,}$ and S_{i4} for batting, fielding, bowling, and wicket keeping is then computed. Finally, on obtaining the values of S_{i1}, S_{i2}, $S_{i3,}$ and $S_{i4,}$ the performance score S_i of the i^{th} player is computed using Eq. (3.66) (cf. subsection 3.6.5.8 of Chap. 3). The performance score (P_i) of all the players is then computed using Eq. (3.67) (cf. subsection 3.6.5.8 of Chap. 3), and its value for a given player is always between 0 and 1. As a large number of players (i.e., 776) for performance measurement are considered here, therefore the details of their performance scores and weights of the different factors are not shown.

6.4 Regression Model for Age Effect on Cricketers' Performance

Suppose we have two variables (say) X and Y where Y is dependent on X (independent). When a relationship is to be established between said dependent and independent variables, a functional relationship (i.e., $Y = a + bX$) concerning dependent to independent is called simple or bivariate linear regression model. In literature, the variables Y and X have several different names that are used interchangeably. Some of them are, Y is called the dependent, explained, response predicted, or regressed variable, and X is called the independent, explanatory, control, predictor, or regressor variable. Multiple regression is an extension of the bivariate regression technique which includes several independent variables, unlike the bivariate regression. However, there are some assumptions that have to be fulfilled. These assumptions are imperative, as it is only under these assumptions that the ordinary least square estimates yield unbiased, consistent, and efficient estimates of the unknown parameters.

6.4.1 *Using the Multiple Regression Model*

In this section, we have fitted a regression model based on panel data to examine the impact of age on performance of the cricketers. In this model, the dependent variable is inverse normal cumulative distribution function of performance score values of the players and independent variables are Age, Age^2, and IPL seasons. Since there are eleven seasons of IPL played so far, therefore to capture the seasonal effect, we have introduced a few dummy variables (d_t) in the model and they are stated below.

$$d_t = \begin{cases} 1, & \text{if a player played in } t^{\text{th}} \text{ IPL season} \\ 0, & \text{otherwise} \end{cases} \quad \text{where } t = 1, 2, \ldots, 10$$

Now the model to examine the impact of age on performance of the cricketers can be written as

$$\varphi^{-1}(P_{it}) = \alpha + \sum_{t=1}^{10} \beta_t d_t + \lambda_1 A_{it} + \lambda_2 A_{it}^2 + \in_{it} \tag{6.2}$$

where P_{it} = Performance score of the i^{th} player at season t

α = Intercept of the model
A_{it} = Age of the i^{th} player at season t
β_t = The coefficients of the dummy variables d_t (i.e., $t = 1, 2, \ldots, 10$)
λ_1, λ_2 = The coefficients of A_{it} and A_{it}^2
$\in_{it}$ = The error term for i^{th} player at season t

The intercept and coefficients of the model are estimated using ordinary least squares method. It has been mentioned earlier that there are some assumptions which need to be fulfilled in an ordinary least squares regression method. One of the most

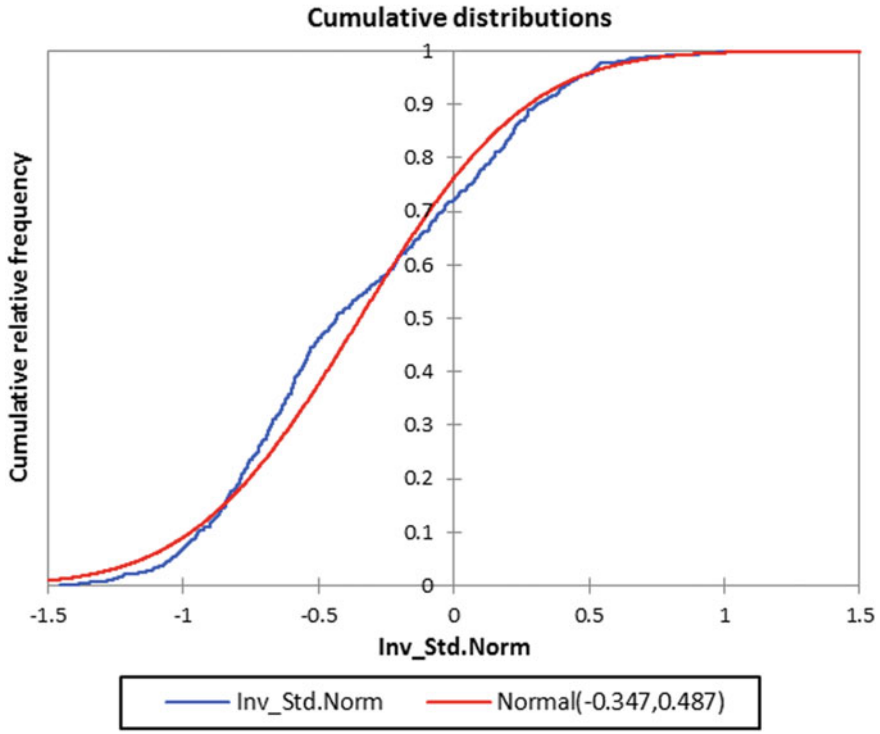

Fig. 6.6 Normality check for the dependent variable of the model

important assumptions is that the dependent variable should be normally distributed. The validation of this assumption is depicted in the following graphical representation (Fig. 6.6).

The estimated model parameters are reported in the following table.

In Table 6.4, it has been observed that the p-values corresponding to IPL2008, IPL2012, IPL2017, and age are 0.000, 0.007, 0.036, and 0.008, respectively. The positive coefficient value of age (λ_1) implies that age has played a significant role in the performance of the cricketers. That means one year increase in age of the players, 0.095 unit will be the increase in their performance. But the quadratic term of the age coefficient (i.e., λ_2) signifies that if age increases by one year, then the performance of the players will decrease by 0.004 (2×0.002) units. Thus, on one hand, we have a positive coefficient of age, and on the other hand, we have a negative coefficient of quadratic age term. Now to look at the net effect of age on performance of the players, we shall draw a scatter plot (cf. Fig. 6.7) by considering 'age' of the players on the x-axis and 'Performance' of the players on the y-axis. The performance that we have considered under y-axis is obtained from the estimated regression line (i.e., Performance $= 0.095A_{it} - 0.002A_{it}^2$).

Table 6.4 Estimated parameters from the model

	Coefficients	Standard error	t	Pr > |t|
Intercept (α)	−1.789	0.503	−3.555	**0.000**
d_1	0.385	0.087	4.423	**0.000**
d_2	0.091	0.087	1.053	0.293
d_3	0.048	0.087	0.554	0.580
d_4	−0.131	0.080	−1.631	0.103
d_5	0.194	0.072	2.706	**0.007**
d_6	0.125	0.070	1.797	0.073
d_7	0.141	0.074	1.884	0.060
d_8	−0.012	0.074	−0.167	0.867
d_9	−0.082	0.075	−1.101	0.271
d_{10}	0.159	0.076	2.104	**0.036**
λ_1	0.095	0.001	−2.548	**0.008**
λ_2	−0.002	0.036	2.647	**0.011**

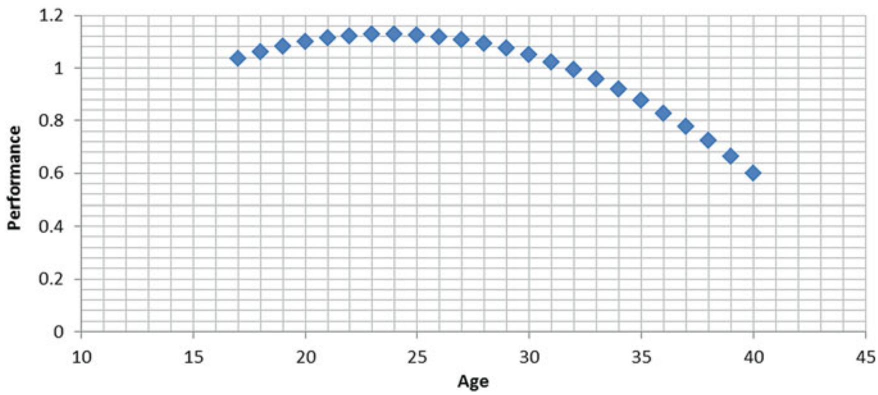

Fig. 6.7 Scatter plot for net effect of age on cricketers' performance

The curvilinear nature of the relationship between age and performance of the players in the above graphical representation confirms that players in the game of cricket reach their peak performance in the late twenties after they have been technically matured enough through experience. However, one can clearly observe from the graph that the performance of the players will start to decrease at one point of time due to the escalation of age. As mentioned earlier, the determination of exact peak performance age of the players is always challenging but then again here we are trying to identify the peak performance age of the players through segmented or piecewise regression. It is because from the nonlinear regression, we could not identify the exact change point or break point of the age after which their performance shows a decline.

6.4.2 Using the Segmented Regression Model

Segmented or piecewise regression is a method which is done by performing a separate linear regression to the data [(x, y) say] with x values (independent variable) smaller and greater than a certain partition value. The partition values between the linear segments are known as change or break points. These break or change points are very essential in decision making and can be interpreted as a threshold value, beyond as well as below which desired effects occur. Like linear regression analysis, the least squares method is applied separately to each segment. The two regression lines are made to fit the data set as closely as possible. As per our data set, we shall try to fit the following two regression lines.

$$\left.\begin{array}{ll}\text{Performance} = C_1 + \beta_1 \text{Age} & \text{for} \quad \text{Age} < K \\ \text{Performance} = C_2 + \beta_2 \text{Age} & \text{for} \quad \text{Age} > K\end{array}\right\} \tag{6.3}$$

where K is the break point, β_1 and β_2 are regression coefficients which are also indicating the slope of the line segments and C_1 and C_2 are called the intercepts of the segmented regression lines. Now to identify the break point for Eq. (6.3) on the basis of piecewise regression, we have used a package called 'segmented' through R programming language which is a free software environment for statistical computing and graphics. The fitted piecewise regression with the estimated break point can be seen in Fig. 6.8. The estimated slopes and intercepts of the segmented or piecewise regression are presented in Table 6.5.

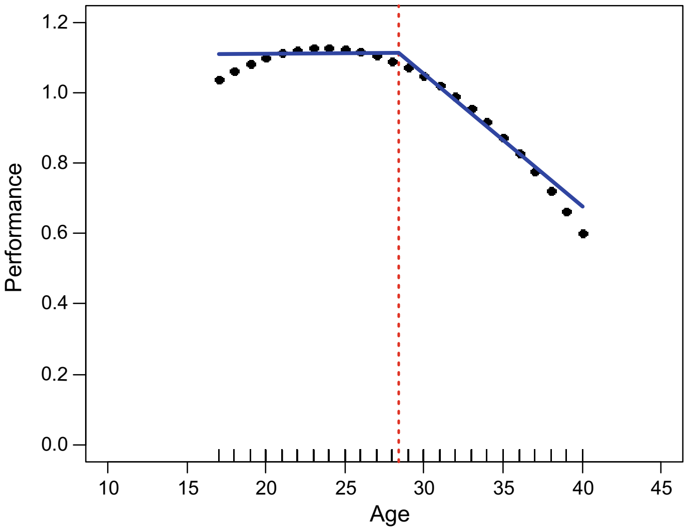

Fig. 6.8 Break point estimation using segmented regression

Table 6.5 Estimated slope and intercepts of segmented regression

	Coefficients	Standard error	t-value	Break point (K)
Intercept (C_1)	1.1025			28.39
Slope (β_1)	0.0004	0.0002	1.4063	
Intercept (C_2)	2.1862			
Slope (β_2)	−0.0377	0.0003	−117.77	

The break point that we have identified from the piecewise regression is 28.39, which means that the performance of the cricketers in terms of Twenty20 cricket shall increase up to the age 28.39 years and after that, it will start to decrease.

Since the peak performance age of a player in terms of team game through a point estimate (e.g., 28.39 years) cannot be admissible, we have to consider an interval estimate. To obtain that interval estimate, we have calculated the standard deviation of the age of the players and it is found to be 4.69. Now we can conclude that the peak performance of the cricketers in terms of Twenty20 cricket is 28.39 ± 4.69. It means that any given player in Twenty20 cricket can perform at his best between the ages 24 and 33 years.

6.4.3 *Doing Separate Analysis for Batsmen and Bowlers*

As we have mentioned earlier, batsman and bowler reach their peak performance at different ages; therefore, to examine that the regression model (6.2) has been applied separately for batsmen and bowlers. The all-rounders are included in both the categories (i.e., batsman and bowler) since they are reasonably good in terms of performance with bat and ball. There were a total of 563 batsmen and 328 bowlers who had participated in all the 11 seasons of IPL played so far. The estimated model parameters through the least squares method for batsmen are given below.

In Table 6.6, it is observed that the p-values corresponding to IPL2008, IPL2012, and age are 0.000, 0.045, and 0.045, respectively. The positive coefficient value of age (λ_1) implies that age has played a significant role in the performance of the batsmen. That means one year increase in age of the batsmen, 0.087 unit will be the increase in their performance. But the quadratic term of the age coefficient (i.e., λ_2) signifies that if age increases by one year, then the performance of the batsmen will decrease by 0.002 (2 × 0.001) units. Now to look at the net effect of age on performance of the batsmen, we have drawn a scatter plot (cf. Fig. 6.9) by considering 'age' of the batsmen on the x-axis and 'Performance' of the batsmen on the y-axis. The performance that we have considered under y-axis is obtained from the estimated regression line $\left(\text{i.e., Batting Performance} = 0.087A_{it} - 0.00143A_{it}^2\right)$.

As discussed earlier, due to the curvilinear nature of the relationship between age and batting performance, piecewise regression is being applied to determine the peak performance age of the batsmen (cf. Eq. (6.3)). The fitted piecewise regression

Table 6.6 Estimated parameters from the model for batsmen

	Coefficients	Standard error	t Stat	P-value
Intercept (α)	−1.76664	0.614199	−2.87634	0.004179
d_1	0.402511	0.108076	3.72434	**0.000216**
d_2	0.133486	0.107698	1.239448	0.215708
d_3	0.10321	0.107573	0.959448	0.337755
d_4	−0.06678	0.092072	−0.72527	0.468597
d_5	0.17304	0.086341	2.004147	**0.045544**
d_6	0.124348	0.086345	1.440127	0.1504
d_7	0.081244	0.091649	0.886466	0.375754
d_8	−0.01834	0.090171	−0.2034	0.838896
d_9	−0.14266	0.091226	−1.56381	0.118438
d_{10}	0.086587	0.093481	0.926256	0.354719
λ_1	0.087026	0.043397	2.005327	**0.045417**
λ_2	−0.00143	0.000758	−1.89161	**0.049069**

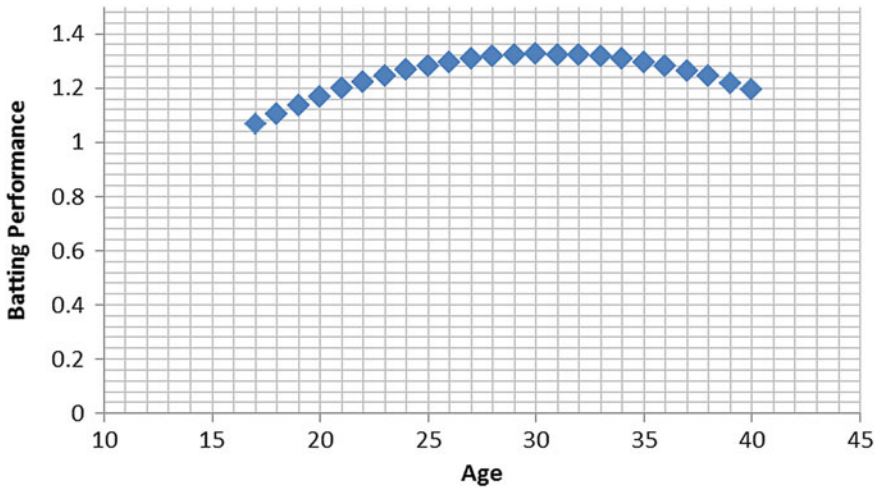

Fig. 6.9 Scatter plot for the net effect of age on batting performance

lines with estimated break point can be seen in Fig. 6.10. The estimated slopes and intercepts of the piecewise regression are presented in Table 6.7.

The break point from the piecewise regression is 28.49, which means that the performance of the batsmen in Twenty20 cricket shall increase up to the age of 28.49 years, and after that, it will start to decrease.

To obtain an interval estimate, we have calculated the standard deviation of the age of the batsmen and it is found to be 4.62. Now it can be concluded that any given

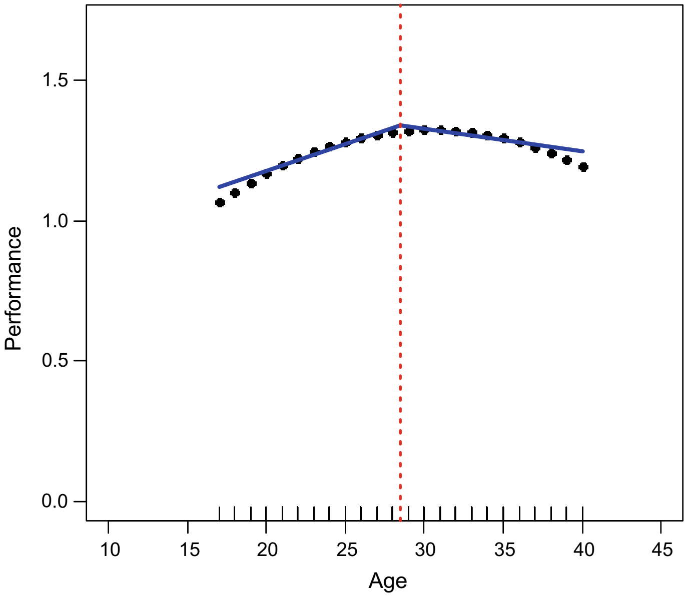

Fig. 6.10 Break point for batsmen using segmented regression

Table 6.7 Estimated slope and intercepts of segmented regression

	Coefficients	Standard error	t-value	Break point (K)
Intercept (C_1)	0.797			28.49
Slope (β_1)	0.018	0.00025	75.097	
Intercept (C_2)	1.567			
Slope (β_2)	−0.008	0.00026	−30.613	

batsmen in Twenty20 cricket can perform at his best between the ages 23.87 and 33.11 years.

Similarly, for bowlers, the estimated model parameters using Eq. (6.2) through least squares method are given in Table 6.8.

In Table 6.8, it is observed that the p-values corresponding to IPL2008, IPL2012, IPL2013, IPL2014, IPL2017, and age are 0.000, 0.000, 0.012, 0.001, 0.000, and 0.000, respectively. The positive coefficient value of age (λ_1) implies that age has played a significant role in the performance of the bowlers. That means one year increase in age of the bowlers, 0.173 unit will increase in their performance. But the quadratic term of the age coefficient (i.e., λ_2) signifies that if age increases by one year, then the performance of the players will decrease by (2×0.002) 0.004 units. Now to look at the net effect of age on performance of the bowlers, we have drawn a scatter plot (cf. Fig. 6.11) by considering 'age' of the bowlers on the x-axis and 'Performance'

Table 6.8 Estimated parameters from the model for bowlers

	Coefficients	Standard error	t Stat	P-value
Intercept (α)	−2.8885	0.681597495	−4.23784	2.97E−05
d_1	0.446299	0.11246945	3.968181	**8.97E−05**
d_2	0.083442	0.112005817	0.744977	0.456841
d_3	−0.03568	0.11191099	−0.31882	0.750077
d_4	−0.03714	0.123996556	−0.29956	0.764713
d_5	0.48374	0.100281677	4.823808	**2.2E−06**
d_6	0.219661	0.087453793	2.511735	**0.012514**
d_7	0.317018	0.097321663	3.257427	**0.001247**
d_8	0.148458	0.091195697	1.627905	0.104545
d_9	0.06967	0.096406185	0.722675	0.470415
d_{10}	0.330755	0.09548517	3.463938	**0.000606**
λ_1	0.173878	0.049206079	3.533665	**0.000471**
λ_2	−0.00284	0.000876881	−3.23601	**0.001341**

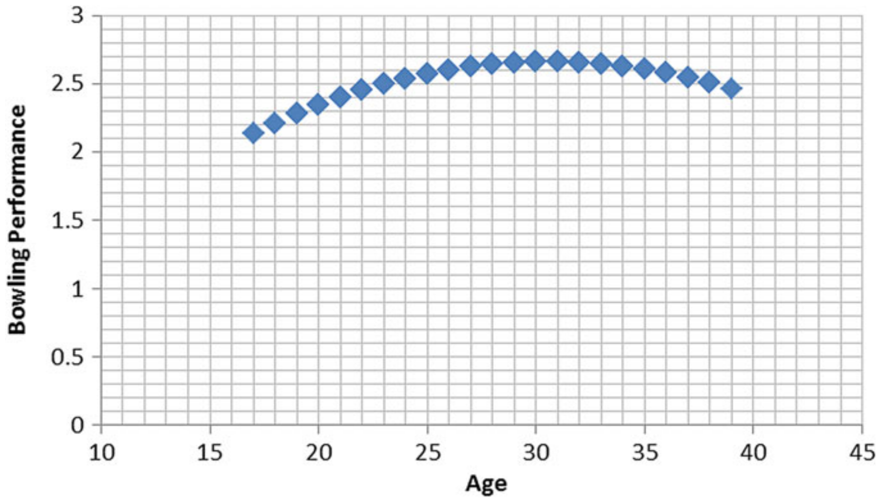

Fig. 6.11 Scatter plot for the net effect of age on bowling performance

of the bowlers on the y-axis. The performance that we have considered under y-axis is obtained from the estimated regression line (i.e., Bowling Performance $= 0.173A_{it} - 0.00284A_{it}^2$).

Again, to capture the curvilinear relationship between age and bowling performance, piecewise regression is applied to determine the peak performance age of the bowlers (cf. Eq. (6.3)). The fitted piecewise regression lines with estimated break point can be seen in Fig. 6.12. The estimated slopes and intercepts of the piecewise

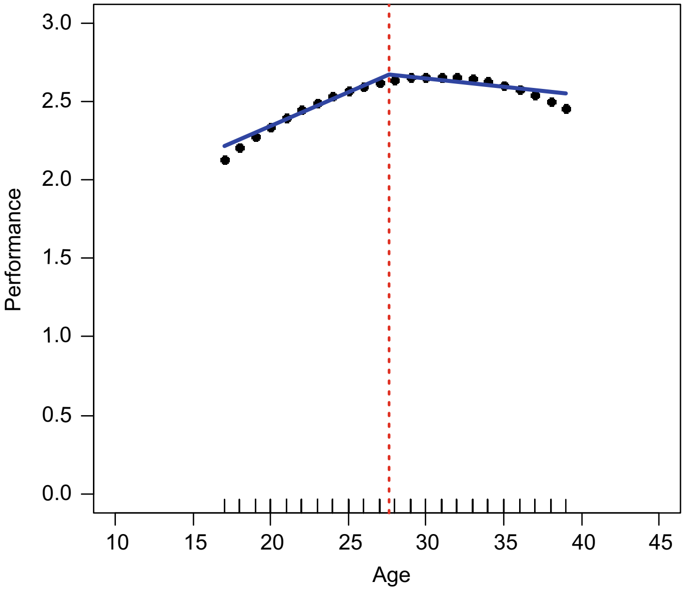

Fig. 6.12 Break point for batsmen using segmented regression

regression are presented in Table 6.9.

The break point from the piecewise regression is 27.63, which means that the performance of the bowlers in Twenty20 cricket shall increase up to the age of 27.63 years and after that, it will start to decrease.

To obtain an interval estimate, we have calculated the standard deviation of the age of the bowlers and it is found to be 4.57. Now we can conclude that any given bowler in Twenty20 cricket can perform at his best between the ages 23 and 32 years.

Table 6.9 Estimated Slope and Intercepts of segmented regression

	Coefficients	Standard error	t-value	Break point (K)
Intercept (C_1)	1.492			27.63
Slope (β_1)	0.042	0.0006	65.519	
Intercept (C_2)	2.974			
Slope (β_2)	−0.011	0.0006	−17.56	

6.5 Age Effect on Cricketers' Performance Using Concentration Index

Here another attempt has been made to examine the age effect on the performance of the cricketers using the concentration curve. It is a common indicator of diversity or inequality in social sciences, and it was first introduced by Wagstaff, Paci, and van Doorslaer (1991), since then it is being used frequently to describe and measure the degree of inequality in healthcare utilization, income or wealth inequality, etc. The concentration curve and index methodology are derived from the most popular Lorenz curve and the Gini index (Atkinson, 1970; Sen, 1973; Anand, 1983). However, it differs from Gini as the ranking variable and variable of interest for which the inequality is evaluated are different and thus concentration index is a bivariate measure of inequality. Since it is a bivariate measure, we have two interrelated variables (i.e., age and performance of the cricketers) to apply this measure. As mentioned earlier, here again, we have selected 776 players considering all the IPL seasons played so far.

Now for each age k (say), let us define

$P(k)$ = cumulative proportion of the player's age $\leq k$
$Q(k)$ = cumulative proportion of the $q(k)$

where the proportion of the players' performance (i.e., $q(k)$) for each k is calculated by

$$q(k) = \frac{\text{aggregate performance score of the players age } \leq k}{\text{aggregate performance score of all the players}} \tag{6.4}$$

If the cumulative proportion of players' performance [$Q(k)$] is plotted against the cumulative proportion of the player's age ordered from lowest to highest [$P(k)$], we shall get the concentration curve (cf. Fig. 6.13). In the graph, the red points represent the values of $Q(k)$ and the blue 45° (or diagonal) dotted line represents the independence case between age and performance. That means if the $Q(k)$ values almost or exactly lie on the blue dotted line, then we can conclude that there is no relationship between age and performance of the players. A very similar situation can be observed explicitly in the following graphical representation. However, some departure can be observed in the extreme left and extreme right, both below and above the 45-degree line. This concentration curve helps us to get a clear visual depiction of performance inequality of players with respect to age.

The same analysis has been performed for batsmen and bowlers separately to examine the impact of age on performance of the cricketers. However, we have observed similar kind of features as in Fig. 6.13 for both batsmen and bowlers.

It is surprising to find that the concentration curves are so close to the 45-degree line, yielding a very low value of the concentration index, whereas, in the case of the regression analysis, we have seen a significant age effect. The explanation of this feature lies in the use of the inverse normal transform in case of the regression

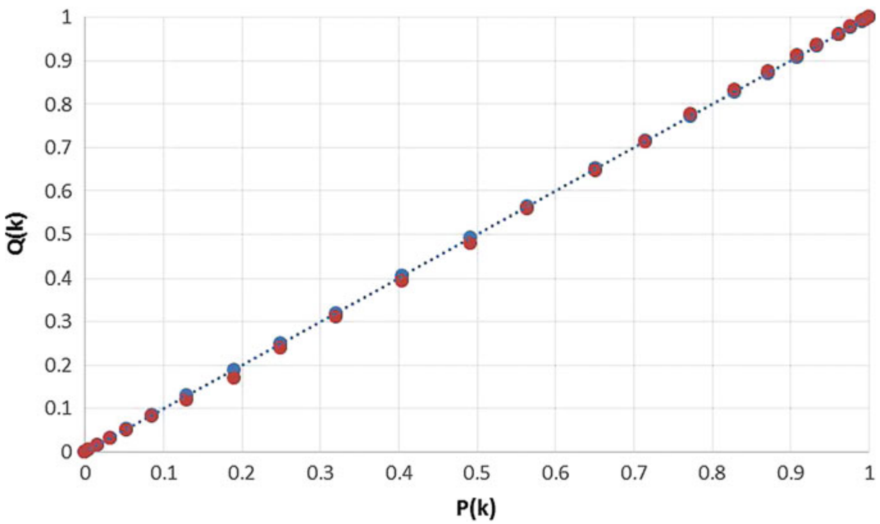

Fig. 6.13 Concentration curve of age and performance

analysis which enabled us to highlight differences which are small in absolute value but are significant on the two extreme ends of the normal curve.

6.6 Conclusion

The results obtained from the regression analysis confirmed that the performances of the cricketers are significantly associated with the age of the players. As an outcome of the study, it has been found that the overall performance of the cricketers as well as batsmen in Twenty20 cricket peak in between the age 24 and 33 years. However, for bowlers, the peak performance age in Twenty20 cricket is between 23 and 32 years. So it is not factual that the younger players perform better in Twenty20 cricket. Therefore, the general belief that Twenty20 cricket is a game of young cricketers only has very little substance.

The results of the concentration curve analysis, which basically measures the performance inequality present in the different skills of the game with reference to age, do not yield analogous significant relationship. From the calculations of the concentration curve, it seems that the impact of aging is very little. But as mentioned earlier, as the calculation of the concentration curve was made on a linear scale and that of the regression analysis on the inverse normal, this difference, in conclusion, is essentially driven by tail behavior of the performance variable.

It is also true that the fitness of the younger players is better than that of the senior players and that young players are often more agile in the field. However, the senior cricketers with their experience can adapt quickly to the Twenty20 format, despite most of them have not experienced much of this format of cricket earlier in their career. It is interesting to note that only senior cricketers with superlative talent like the Tendulkar, Dravid, Warne, and Muralitharan made it to the IPL for first four consecutive seasons. So adjusting to the latest format and performing well may be attributed to the cricketing abilities with which the players are blessed. Many of the younger players who get the opportunity to play in the IPL do not have the same talent as the senior players and very few go through to the national or international level.

In this chapter, the estimation of peak performance age of the cricket players is executed for Twenty20 cricket only, with special reference to the tournament Indian Premier League (IPL). The optimum level of performance of the cricketers between the ages 24 and 33 cannot be generalized to other formats of the game (i.e., one-day international (ODI) and test cricket), but the study can be extended to identify the peak performance age of the cricketers for ODI and test cricket also. It will assist the selectors or board of the management in the decision-making process for team formation of national as well as international tournaments.

Appendix 6.1: The Details of Damodaran (2006) Method

A simple method was defined in Damodaran (2006) that can replace the not out score of a batsman by an adjusted score based on the previous performances of a batsman. The method can be explained as Let R_i be the runs scored by Sachin in the i^{th} innings and that is an innings in which Sachin was not out. The method helps us to estimate the runs Sachin might have scored in that innings if he would have been allowed to bat till he gets out. To each of the $i - 1$ previous innings of Sachin, we associate an indicator variable T_j, where $T_j = R_j$ if $T_j \geq R_i$ and 0 otherwise with, $j = 1, 2, \ldots, i - 1$. Another indicator variable N_j is defined such that $N_j = 1$ if $T_j \geq R_i$ and 0 otherwise. Thus, the adjusted runs R_i^* corresponding to the not out innings where Sachin has scored R_i runs is given by,

$$R_i^* = \frac{\sum_{j=1}^{i-1} T_j}{\sum_{j=1}^{i-1} N_j}$$

In simple words, this is the arithmetic mean of runs scored, more than R_i, by Sachin in all his previous innings.

= 0 =

References

Addona, V., & Yates, P. A. (2010). A closer look at the relative age effect in the National Hockey League. *Journal of Quantitative Analysis in Sports, 6,* 1–19.

Anand, S. (1983). *Inequality and poverty in Malaysia: Measurement and decomposition*. New York: Oxford University Press.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory, 2,* 244–263.

Baker, A. B., Tang, Y. Q., & Turner, M. J. (2003). Percentage decline in masters superathlete track and field performance with aging. *Experimental Aging Research, 29,* 47–65.

Damodaran, U. (2006). Stochastic dominance and analysis of ODI batting performance: The Indian Cricket Team, 1989–2005. *Journal of Sports Science and Medicine, 5*(4), 503–508.

Fair, R. (2008). Estimated age effects in baseball. *Journal of Quantitative Analysis in Sports, 4,* 1–39.

Hazra, M. S., & Biswas, S. (2018). A study on mental skill ability of different age level cricket players. *International Journal of Physiology, Nutrition and Physical Education, 3*(1), 1177–1180.

Himann, J., Cunningham, D., Rechnitzer, P., & Paterson, D. (1988). Age-related changes in speed of walking. *Medicine and Science in Sports and Exercise, 20,* 121–127.

Krishnaswamy, K. (2010). *Of Veterans and Extended Careers.* Retrieved June 28, 2011 from www.tssonnet.com/tss3313/stories/20100401502300400/htm.

Lehto, N. (2015). Effects of age on marathon finishing time among male amateur runners in stockholm marathon 1979–2014. *Journal of Sport and Health Science, 5,* 349–354.

Lemmer, H. H. (2005). A method for the comparison of the bowling performances of bowlers in a match or a series of matches. *South African Journal for Research in Sport, Physical Education and Recreation, 27*(1), 91–103.

Lucifora, C., & Simmons, R. (2003). Superstar effects in sport: Evidence from Italian Soccer. *Journal of Sports Economics, 4*(1), 35–55.

Mitra, S. (2010). The IPL: India's foray into world sports business. *Sports in Society, 13,* 1314–1333.

Moonda, F. (2011). *The IPL XI.* Retrieved May 29, 2011 from www.espncricinfo.com/indian-premier-league-2011/content/story.htm.

Musch, J., & Grondin, S. (2001). Unequal competition as an impediment to personal development: A review of the relative age effect in sports. *Developmental Review, 21,* 147–167.

Radek, V. (2014). Identification of peak performance age in track and field athletes. *International Journal of Performance Analysis in Sports, 14,* 238–251.

Raman, W. V. (2009a). *IPL is Here to Stay.* Retrieved June 28, 2011 from www.tssonnet.com/tss3217/stories/20090523505202800.htm.

Raman, W. V. (2009b). *There's No Substitute for Skill.* Retrieved June 28, 2011 from www.tssonnet.com/tss3217/stories/200904255010004000.htm.

Roebuck, P. (2008). *T20 Will Improve Performance.* Retrieved June 28, 2011 from www.tssonnet.com/tss3121/stories/20080524503602600.htm.

Ransdell, L. B., Vener, J., & Huberty, J. (2009). Masters athletes: An analysis of running, swimming and cycling performance by age and gender. *Journal of Exercise Science and Fitness, 7*(2), S61–S73.

Saikia, H., & Bhattacharjee, D. (2011). On classification of all-rounders of the Indian Premier League (IPL): A bayesian approach. *Vikalpa, 36,* 25–40.

Schulz, R., & Curnow, C. (1988). Peak performance and age among super athletes: Track and field, swimming, baseball, tennis and golf. *Journal of Gerontology: Psychological Sciences, 43*(5), 113–120.

Sen, A. (1973). *On economic inequality*. New York: Norton. https://doi.org/10.1093/0198281935.001.0001.

Seiler, K., Spirduso, W., & Martin, J. (1998). Gender differences in rowing performance and power with aging. *Medicine and Science in Sports and Exercise, 30,* 121–127.

Tanaka, H., & Seals, D. (1997). Age and gender interactions in physiological functional capacity: Insight from swimming performance. *Journal of Applied Physiology, 82,* 846–851.

Wagstaff, A., Paci, P., & van Doorslaer, E. (1991). On the measurement of inequalities in health. *Social Science and Medicine, 33*(5), 545–557.

Young, W. A., & Weckman, G. R. (2008). Evaluating the effects of aging for professional football players in combine events using performance-aging curves. *Internatinal Journal of Sports Science and Engineering, 2,* 131–143.

Chapter 7
Performing Under Difficulty: The Magical Pressure Index

7.1 Introduction

In a limited overs cricket match, let us call the team batting first as Team A and the other team as Team B. To win the match, Team A tries to score as many runs it can against the bowling of team B. Team B on the other hand use all their resources to restrict the run scoring of Team A. At the end of the innings, Team A sets a 'target' for team B, which is one run more than the runs scored by Team A. The target is to be attained by team B within the specified number of overs without losing all its batting resources. In a limited over match, a slow scoring rate or/and losing too many early wickets increases the difficulty for the team batting second in achieving the target. Sometimes to accelerate the rate of scoring runs, the batsmen render themselves to the chance of getting dismissed. This risk can be avoided if the batsmen score slowly but this shall lead to a scoring rate that may be lower than the required rate of scoring runs. This shall make the job difficult for team B. A high scoring over, on the other hand, makes team B's situation more comfortable in the match. In this chapter, we try to introduce a model that can be used to measure the pressure experienced by both teams in the second innings of a limited over match following Bhattacharjee and Lemmer (2016). The pressure on both the teams, batting or fielding, in the second innings may change after every ball. A ball in which no runs are scored (often referred to as a dot ball) or the fall of a wicket increases the pressure on the batting team. A scoring shot, a six or a four, or a productive over of say 20 runs, makes the task of the batting team easier and decreases the pressure. This study attempts to quantify the pressure on both teams in the second innings of a limited over match at any given point of the innings. The pressure on Team B is inversely proportional to (i) the number of balls that remains to be bowled and (ii) the number of wickets yet to fall, but directly proportional to (iii) the runs that are yet to be scored by team B for winning the match.

H. Saikia et al., *Cricket Performance Management*, Indian Statistical Institute Series,
https://doi.org/10.1007/978-981-15-1354-1_7

7.2 Review of Literature

Cricket is a game in which every ball bowled generates huge numerical information. Thus, the game is expected to be a statistician's delight. Data scientists have contributed several works on performance measurement of cricketers, decision making, and team selection. An important and a real statistical problem lies in resetting target scores in matches that are stopped due to bad weather conditions. The well-known Duckworth/Lewis method and now extended to the Duckworth/Lewis/Stern method seems to provide an acceptable solution to the problem after several experimentations. One may look into the Duckworth and Lewis (1998), Jayadevan (2002), de Silva (2011), Preston and Thomas (2002) and Bhattacharya, Gill, and Swartz (2011) for details. Predicting the final score in the second innings of one day international was another interesting and useful exercise taken up successfully by Bailey and Clarke (2006) and De Silva, Pond, and Swartz (2001) when the match was in progress. This in one hand helps to understand the outcome of the match and in the other hand helps to quantify the margin of victory in terms of runs. It may be noted here that—in limited overs cricket, if the team batting second wins the match the magnitude of victory is measured in terms of the number of wickets in hand. But, in case the team batting second loses the match—the magnitude of victory of team batting first is measured in terms of number of runs yet to be scored by the team batting second. Both the works used the Duckworth/Lewis resource table at the end of each over of the second innings predicted the winner of the match along with the innings score. The problem of estimating the magnitude of victory in terms of runs in One-day matches was once again taken up by directly by De Silva et al. (2001) again used the Duckworth/Lewis resource table. As a utility of the study, they suggested that this can be used to break ties between teams in the points table at the end of the league stage of any tournament. Lemmer (2015) developed a performance measure, denoted by PR that can describe the batting progress of the team batting second in a limited overs match.

$$\mathrm{PR} = \mathrm{CRRR} \times \mathrm{RU}$$

where CRRR denotes the current required run rate and RU denotes the percentage of resources that had already been used as per the Duckworth/Lewis system. PR was based on the performance of the batting team and also on its use of resources, i.e., wickets lost and balls already bowled. Shah and Shah (2014) defined a 'pressure index' which can also be used to assess the team's progress during the second innings of a limited overs match. Their pressure index (PI) consists of various parts based on the progress in run scoring and the number of wickets that are lost. Though the idea was praiseworthy, unfortunately, the work lacks motivation and some symbols remained undefined.

Bhattacharjee and Lemmer (2016) describe pressure index as 'a number indicating the relative level of pressure compared with a standard,' where the standard refers to a base or starting value. In this study, it was assumed that the base value will be 1 or

100 and that the index fluctuates around this value according to the pressure exerted by the opponents.

The main purpose of the pressure index introduced by Bhattacharjee and Lemmer (2016) was to quantify the pressure that the batting team or the bowling team experiences at any stage of the match. The pressure on the batting team is obviously determined by a combination of two factors, namely its batting performance and its ability to retain wickets. Similarly, the pressure on the bowling team is determined by its ability to take wickets and to restrict the scoring of runs of its opponents.

7.3 Methods

7.3.1 Criteria for Measuring Pressure

For quantification of pressure in a cricket match, it is necessary to assess the scoring process by considering the required run rate to find out if the scoring rate is satisfactory. Let T be the target runs for Team B. At the start of the Team B's innings, the initial required run rate is

$$\text{IRRR} = T/(\text{Total number of Overs}) = 6 \cdot T/B$$

with B the total number of balls available. During the batting of Team B, the team needs to take into account the current required run rate CRRR where CRRR $=$ $6 \cdot$ Rr/Br with Rr the number of runs still required and Br the number of balls remaining. If this is compared with the initial required run rate, IRRR, it is easy to understand if the progress is satisfactory. The ratio

$$\text{CI} = \text{CRRR}/\text{IRRR} \tag{7.1}$$

starts from the value 1 and it is desirable that it should decrease progressively until it reaches the value 0, but it can become very high if the batting team fails. CI is a suitable measure for batting progress.

The fielding side has to take into consideration two criteria. First, the bowling team wants to take as many wickets as possible so as to weaken the batting team's strength. With the fall of wickets, the batting team's wicket strength deteriorates. But, instead of just counting the number of wickets that have fallen, Bhattacharjee and Lemmer (2016) took into account the ability of the batsmen who were dismissed. Considering the fact that top-order batsmen are better performers than the lower order batsmen, this has been done. In order to take this into account, the wicket weights of Lemmer (2005) are used—see Appendix 7.1. Denote by $\sum w_i$ the sum of the weights of the wickets that had fallen at any stage of the innings. When $\sum w_i$ increases, it is an indication that the team's wicket strength has decreased. Taking into account that the sum of all the wicket weights is equal to 11, the ratio SW $= \sum w_i/11$ reflects how far the wicket strength has deteriorated. As wickets fall, the value of SW will initially increase rapidly, but its growth will slow down when low-order wickets fall. Thus, exp(SW) is used as a factor that can quantify the effect of wickets falling. At the very start of the innings, when no wicket has fallen, the factor exp(SW) has an initial value of 1.

Secondly, the purpose of the fielding team is to restrict the batting team from reaching the target score as its resources are depleted. The Duckworth/Lewis Full Table for Twenty20 matches provides the percentage of resources left to the batting team at the end of each ball depending on the number of wickets they have lost as provided in Bhattacharya et al. (2011). Subtracting the value from 100 provides the percentage of resource already utilized (RU) by the batting team. The value of RU is a function of the number of overs Team B has batted and the number of wickets it has lost. Thus, RU can be used as an alternative to SW and the change in pressure due to resources utilized can be quantified by the factor exp(RU/100).

7.3.2 Pressure Indices

Shah and Shah (2014) defined a pressure index

$$\mathrm{PI} = \mathrm{CI} \times 100 + [(Wk \cdot Wt/180) \times T \times (\mathrm{Br}/B) \times (\mathrm{Rr}/T)] \tag{7.2}$$

The starting value of PI is always 100 and its value will normally fluctuate. When PI was calculated for various matches, it was found that the value of PI decreased when a wicket went down and the number of runs scored was below the required run rate. This is totally unrealistic considering the very purpose of pressure index, and hence, this formula of PI due to Shah and Shah (2014) is not considered any further.

In limited overs cricket, the team batting second needs to achieve the target before their entire resources are exhausted, in terms of balls bowled and without losing all their wickets. In the development of a pressure index, one shall note that the pressure will decrease if the batting progress is good in terms of the required run rate AND if wickets are kept till the target score for victory is reached. Thus, both shall happen simultaneously not the one or the other. In this manner, the measure ought to be the product of the two components and not their aggregate. Two techniques are utilized to evaluate the utilization of resources, first, on the way in which wickets are lost and the other by utilizing the Duckworth/Lewis resource utilization table. This takes us to two different formulae of pressure index as beneath.

The first pressure index is defined as

$$\mathrm{PI}_1 = \left(\frac{\mathrm{CRRR}}{\mathrm{IRRR}}\right) \times \exp\left(\sum w_i/11\right) = \mathrm{CI} \times \exp\left(\sum w_i/11\right) = \mathrm{CI} \times \exp(\mathrm{SW}) \tag{7.3}$$

and the second definition is

$$\mathrm{PI}_2 = \left(\frac{\mathrm{CRRR}}{\mathrm{IRRR}}\right) \times \exp(\mathrm{RU}/100) = \mathrm{CI} \times \exp(\mathrm{RU}/100) \tag{7.4}$$

Factor CI is expected to change after every ball, but exp(SW) changes only on the fall of a wicket but the factor exp(RU/100) changes after every delivery, as the value of RU changes after every ball. At the start of the second team's innings, both pressure indices are equal to 1 and will increase or decrease or fluctuate as the match progresses. At the end of the match, if Team B wins, CI becomes 0 and therefore also PI_1 and PI_2. But, it seems that PI_2 shall be more sensitive compared to PI_1 given the fact that it depends on RU which changes after every ball.

7.3.3 Working of the Pressure Indices

To understand how the indices work, they are applied to an actual match situation. A Twenty20 international match of the recent past having lots of ups and downs in the second innings is considered as an ideal one to understand the working of the indices. The second Twenty20 match of Pakistan versus West Indies, played on March 30, 2017, at the Queen's Park Oval, Port of Spain, Trinidad, is considered. Pakistan batted first and scored 132/10 in 20 overs. The target for West Indies was 133, which seemed to be easily getable considering the batting lineup of West Indies. At the end of the 8th over West Indies was 60/1 and seemed to be comfortably proceeding toward the target. But then started the fall, wickets started tumbling one after another and the target started becoming difficult to attain. At the end of the 19th over, West Indies needed 14 runs to win with 3 wickets in hand. Two consecutive fours in the first two balls made the situation exciting and in favor of West Indies. But only two more runs were scored in the next four balls and West Indies ultimately lost the match by 3 runs. The ball-by-ball values of the pressure indices, i.e., both PI_1 and PI_2, are given in Appendix 7.2. The graphical representation of the same can be seen in Fig. 7.1. The pressure increased almost consistently in the first 10 overs and then a rapid rise is noticed in the last 10 overs.

The graphs in Fig. 7.1 show that there is not much difference in the pressure values obtained through PI_1 and PI_2, and hence, the interpretation is almost the same. After the start of the West Indies innings, the pressure kept on increasing and was boosted up with the fall of wickets at regular intervals. The graph shows that the second wicket partnership between Walton and Samuels lasted longest, followed by the 7th wicket partnership between Holder and Brathwaite. But the partnership of Holder and Brathwaite which scored 33 runs in 32 balls at a rate of little above 6 runs per over was not sufficient given the match situation. Thus, pressure increased during the partnership, which is visualized in the graph. It may be seen that with two consecutive fours in the last over of the match a slight drop in the pressure can be seen but after that, the pressure kept on increasing indicating that the match was lost by Team B.

The two curves of PI_1 and PI_2 delineate the circumstance properly with the pressure on West Indies expanding at the fall of a wicket and the pressure diminishing

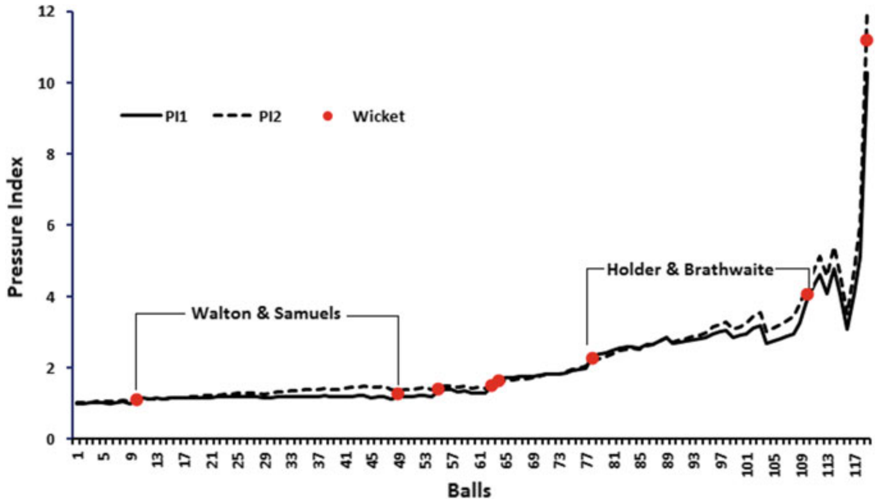

Fig. 7.1 Pressure index graph for the West Indies innings in the international Twenty20 match against Pakistan played on March 30, 2017

after an over where numerous runs were scored. Table 7.1 gives the expansion in the quantities of PI_1 and PI_2 at various circumstances of the match. The expansion in both the lists after high- and low-scoring overs and after the fall of wickets gives fascinating information into the index values. It is seen from Table 7.1, that by and large, the adjustment in the values of PI_1 is not exactly the same as in PI_2. In this manner, PI_2 is by all accounts more touchy than PI_1. The main distinction between PI_1 and PI_2 is in their exponent. After each ball, there is a change in the exponent of PI_2 in view of an adjustment in the value of RU. However, PI_1 changes just at the fall of a wicket. At the point when a wicket falls, the errand of the team that is batting turns out to be notably increasingly troublesome in light of the fact that the new batsman needs to play himself in and the scoring rate generally drops at that stage. PI_1 mirrors this situation very nicely. But for measuring the performance of a partnership, PI_2 is more qualified than PI_1 to mirror the utilization of resources. Ultimately, as a measure for the pressure experienced by a team batting second, the average value of PI_1 and PI_2 is considered. This is defined as

$$\mathrm{PI}_3 = \left(\frac{\mathrm{CRRR}}{\mathrm{IRRR}}\right) \times \frac{1}{2}\left[\exp(\mathrm{RU}/100) + \exp\left(\sum w_i/11\right)\right] \tag{7.5}$$

The curve obtained by plotting the values of PI_3 against the corresponding ball results in a curve called the pressure curve. It is observed that in case the match is won by Team B, the value of pressure computed through any one of PI_1 or PI_2 or PI_3 comes down to zero. However, if the match is lost by Team B, then the pressure curve of PI_1 or PI_2 or PI_3 shall keep on rising as is seen in Figs. 7.1 and 7.2.

Table 7.1 Pressure index values for different events in the second innings of the West Indies vs. Pakistan match

Sit. no.	Event	Situation	PI_1	PI_2	Increase in PI_1	Increase in PI_2
1	Lowest scoring 16th over (0 runs scored)	Opening value	3.141376	3.056936	0.314665	0.235556
2		Closing value	3.456041	3.292492		
3	Highest scoring 8th over (13 runs scored)	Opening value	1.444663	1.321182	−0.14675	−0.15745
4		Closing value	1.297915	1.16373		
5	Lewis out	Before wicket	1.069456	1.039171	0.03651	0.081537
6		After wicket	1.105966	1.120708		
7	Walton out	Before wicket	1.297915	1.16373	0.051243	0.101036
8		After wicket	1.349158	1.264766		
9	Pollard out	Before wicket	1.443951	1.365581	0.07281	0.135233
10		After wicket	1.516761	1.500814		
11	Samuels out	Before wicket	2.064722	2.031855	0.136829	0.231099
12		After wicket	2.201551	2.262954		
13	Brathwaite out	Before wicket	3.803465	3.523094	0.439901	0.548274
14		After wicket	4.243366	4.071368		

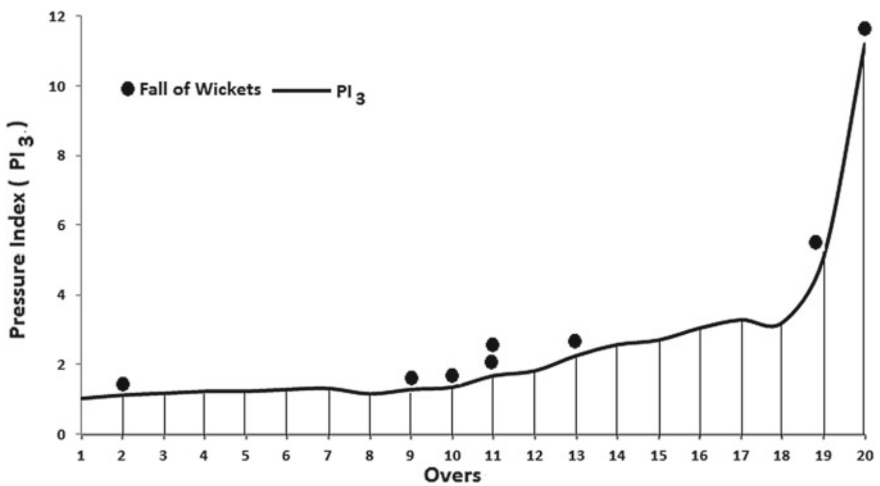

Fig. 7.2 Pressure curve based on the pressure index (PI_3) values at the end of each over

The values of PI_3 are computed at the end of each over and are shown in Fig. 7.2. Plotting at the end of each over, instead of ball-by-ball, smoothens the pressure curve.

The pressure indices can be used for a number of purposes. Some of these are highlighted here.

7.4 Applications of Pressure Index

It is observed that at the end of the match with a rapid increase in the value of CI, the pressure curve shoots up and becomes very unstable. This is not specific to the West Indies vs Pakistan match only but is a common feature to all matches in which the team batting second, i.e., Team B loses the match. This unusual raise in PI_3 values at the end of the innings shall make the pressure difference high at the end of the innings. Thus, the batting partnership at the end of the innings will show high-pressure differences and shall mostly be held responsible for the defeat. This shall remain a common feature to all matches where Team B loses. So, for measuring the performance of batsman or batting partners, some transformation of PI is taken rather than the index directly. The objective of taking the transformation is to provide some symmetry and bring a gradual increase in pressure, even when the chasing team is falling short of their target. One common candidate for transformation shall be the logarithmic transformation (natural log). But when the quantity to be transformed (generally the value of pressure index or a function of it) is less than 1, the transformed value becomes negative. Negative values are difficult to interpret when used for the purpose of performance measurement. Thus, square root transformation is used on pressure index (or its function) to attain the performance of batsmen, bowlers, and batting partners as can be seen in the subsequent sections.

7.4.1 Measuring Batting Performance

The batting performance of the players in a single innings can be compared in terms of runs scored or/and strike rate of the batsman. But here, we propose a new batting performance measure for a single innings as a derivative of the pressure index.

Let $\text{Ave}(PI_3)$ denote the average pressure of the entire second innings. PI_{3i} denotes the pressure level at the end of the i^{th} ball of the second innings of the match. R_{ij} denotes the runs scored by batsman j in the i^{th} ball of the match. Then, the runs scored in the i^{th} ball is transformed to the adjusted runs scored (R^*_{ij}) using the following relation,

$$R^*_{ij} = R_{ij} \times \left[\frac{PI_{3,i-1}}{\text{Ave}(PI_3)} \right]^{0.5} \tag{7.6}$$

That is, the adjusted runs scored in the i^{th} ball is the actual runs scored multiplied by the ratio of the pressure index at the end of the previous ball to the average pressure index value of the entire innings. Summing up all the adjusted runs scored by a batsman (j^{th} batsman, say) his adjusted score of the innings shall be obtained. The adjusted score takes into consideration the actual runs scored and the pressure level under which these runs were scored. We call this his batting contribution AR_j^* where

$$\mathrm{AR}_j^* = \sum_i R_{ij}^* \tag{7.7}$$

The summation extends over those balls that the j^{th} batsman faced. The significance of his batting contribution can be put into perspective by taking into account how many balls he used, say B_j. Define his batting achievement by

$$BA_j = AR_j^*/B_j \tag{7.8}$$

The batting achievement can be considered as a measure of batting performance in a single innings while chase is on. To compute the batting achievement of a given innings, we consider another tightly fought Twenty20 international match played between India and Bangladesh. The two teams locked horns on March 23, 2016, at M. Chinnaswamy Stadium, Bangalore, for the 25th match of the Twenty20 World Cup. It was a Group 2 match of the Super 10 round and a very important encounter in the context of the tournament for both the countries. Bangladesh won the toss and put India to bat. India scored 146 for 7 in 20 overs. When the chase began, the required run rate for Bangladesh was 7.35. The progress was consistent for Bangladesh but at the same time they kept on losing wickets at regular intervals. At the end of the 19th over, Bangladesh's score was 136 for 6. They required only 11 runs in the last over. But could ultimately manage 9 runs in the over, losing the match by one run only. Table 7.2 shows the batting achievements of the Bangladesh innings based on this

Table 7.2 Batting achievement score (BA) of batsmen in the India versus Bangladesh Match of the Twenty20 World Cup in 2016

Batsman	Runs	Balls	Adjusted runs (AR_{ij}^*)	BA_j
Mushfiqur Rahim	11	6	14.91	2.49
Sabbir Rahman	26	15	25.05	1.67
Mashrafe Mortaza	6	5	7.55	1.51
Soumya Sarkar	21	21	25.57	1.22
Shakib-al-Hasan	22	15	20.54	1.37
Mahmudullah	18	22	21.82	0.99
Tamin Iqbal	35	32	27.98	0.87
Md. Mithun	1	3	0.78	0.26

new measure of batting performance as given in (7.8). The batsmen are arranged in descending order of the magnitude of their *BA* values.

Mushfiqur Rahim scored 11 runs in 6 balls, but the runs were scored when the pressure was maximum so his batting achievement score is better than others. Sabbir Rahman scored 26 in 15 balls but when he batted the pressure was not as high when Rahim was batting. However, one may have a restriction on the minimum number of balls faced by a batsman, before his performance being quantified using the '*Batting Achievement Score*.'

7.4.2 Measuring Bowling Performance

The pressure index defined in (7.5) can be used to derive a measure of bowling performance. The pressure index values reflect the pressure exerted on the batting team by the bowling team. The difference between the values of PI_3 at the end and start of the i^{th} over bowled by the j^{th} bowler indicates the increase in pressure on the batting team created while bowler j on bowling his i^{th} over. As discussed earlier, the values of pressure index (PI_3) show an unusual rise at the end of the innings in case Team B falls way behind their target. This will give an advantage to the bowlers bowling at the end of the second innings (when Team B loses the match with a significant difference) if the Bowling Performance Index (BPI) is directly connected to PI_3. Thus, to avoid this, a square root transformation is imposed in the values of PI_3. Let, I_{ij} denotes the difference in the square root of PI_3 values at the end and beginning of the i^{th} over bowled by the j^{th} bowler. Then the average increase in pressure created while bowler j was bowling is given by, I_{ij}

$$\mathrm{BPI}_j = \frac{\sum_{i=1}^{n_j} I_{ij}}{n_j} \tag{7.9}$$

Here, n_j denotes the number of overs bowled by the j^{th} bowler. This can be used as a measure of the bowling performance of the bowler. A good bowler is expected to bowl well and increase the pressure on the batting team by restricting runs and taking wickets. The index measures the change in pressure exerted by the bowler on the batting side by delivering dot balls and through the fall of wickets. The higher the value of BPI_j, better is the performance of the bowler. Thus, the bowler with the highest BPI is the best bowler of the innings. Most of the traditional measures of bowling performance like bowling average, bowling strike rate and economy rate are negative measures, i.e. lower the value better it is. The advantage of BPI_j over others is that it is a positive measure.

Table 7.3 shows that the best bowling performance among the Indian bowlers in the match is by Pandya followed by Bumrah. Since Pandya bowled the last over and was able to bring down the pressure to 0 so his performance is rewarded by being ranked as the best bowler.

Table 7.3 Bowling performance values based on pressure indices in India versus Bangladesh match of the Twenty20 World Cup in 2016

Bowler	Runs conceded (overs bowled)	Wickets taken	BPI
Pandya	29 (3)	2	0.1556
Bumrah	32 (4)	0	0.1481
Ashwin	20 (4)	2	0.0569
Jadeja	22 (4)	2	0.0493
Raina	9 (1)	1	0.0493
Nehra	29 (4)	1	0.0265

Both the bowling performance index (BPI) and the batting achievement score (BA) can be extended to several matches and the performance of batsman or bowlers can be ranked either at the end of the tournament or series.

7.4.3 Finding the Best Partnership of an Innings

The most successful partnership of the match can be identified as the partnership in which the pressure on Team B was decreased to the largest extent per ball played. Instead of taking the direct values of PI_3, a square root transformation is taken to eliminate the abnormal rise of PI_3 at the end of the innings in case Team B falls short of their target. If PI_{si} is the pressure index value at the starting and PI_{ci} as the pressure index value at the closing of the i^{th} partnership, which lasted for b_i balls (say), then we define the batting partnership as,

$$\Delta PI_i = \frac{\sqrt{PI_{si}} - \sqrt{PI_{ci}}}{b_i} \tag{7.10}$$

A positive value of ΔPI_i indicates that the partnership was able to decrease the pressure of Team B and has eased the way for the next partnership but negative value of ΔPI_i implies that the partnership has increased the pressure for Team B. Considering the match between West Indies and Pakistan discussed in Sect. 7.3.3, the performance of batting partners is provided in Table 7.4.

As all the values of column no. 4 of Table 7.4 are negative, this indicates that none of the batting pairs decreased the pressure for West Indies. The exercise shows that during the partnership of MN Samuels and CAK Walton, there was a minimum increase in pressure per ball on the batting team. So, among the different pairs, this is the most effective pair so far as decreasing the pressure on Team B is concerned.

The statistic for measuring the performance of batting partnership, i.e., (Increase in $\sqrt{PI_3}$)/b_i is a negative measure. The lower the value better is the batting partnership. In such cases, it can even take values less than zero, and such values indicate that the batting partnership has eventually diminished the pressure of the batting team. Measuring the performance of batting partners for several matches can be a very

Table 7.4 Pressure index for batting partnerships

Sl no.	Partnership	Partnership statistics	ΔPI_i
1	C. A. K. Walton and E. Lewis	10 runs in 10 balls	−0.0059
2	M. N. Samuels and C. A. K. Walton	50 runs in 39 balls	−0.0017
3	M. N. Samuels and L. M. P. Simmons	5 runs in 6 balls	−0.0110
4	M. N. Samuels and K. A. Pollard	11 runs in 8 balls	−0.0043
5	M. N. Samuels and R. Powell	0 run in 1 balls	−0.0576
6	M. N. Samuels and C. R. Brathwaite	5 runs in 14 balls	−0.0199
7	C. R. Brathwaite and J. O. Holder	33 runs in 32 balls	−0.0248
8	SP Narine and JO Holder	14 runs in 9 balls	−0.1476

useful exercise. At the end of such a study, one can tell the best batting partners of a series or a tournament which is dealt with in the next section. This process can also be developed to find out the best batting partner for a player who had an extended cricket career and has played with several partners during his span.

7.4.4 Best Batting Partnership of a Tournament

Very few studies related to partnerships are found in the available literature. Probably the pioneering work in this regard is that of Pollard, Benjamin, and Reep (1977). Taking data from English County Championships, they made an attempt to quantify batting performance in partnerships using the negative binomial distribution. In order to assess the partnership performance in Test cricket, Valero and Swartz (2012) compared the performance of batsmen with their common partners to identify the ones with whom synergy exist between the players. They defined a bivariate statistic, based on strike rate and a batting average of the players shared with the common partner in the crease, collected over an adequately large number of innings. Thereafter, they have drawn a scatter plot on the basis of the bivariate statistic. The scatter plot is divided into four quadrants to find out any obvious pattern that would probably indicate a presence of synergy between the players. However, the scatter plot did not reveal any such pattern to indicate the presence of synergy.

Usually in one-day international (ODI) cricket, batsmen attempt to score runs at a high rate relative to balls faced while simultaneously avoiding dismissal. It is the combination of wickets available and overs remaining in innings that provide the capacity for scoring runs. To assess the opening partnership performance in ODI, Valero and Swartz (2012) used Duckworth/Lewis (D-L) (1998) method and the method developed by Beaudoin and Swartz (2003). According to Duckworth and Lewis (1998), the combination of wickets and overs is known as resources and the resources consumed by the batting team in ODI during any interval of the match can be determined from the D-L table. Again, following Beaudoin and Swartz (2003), the authors have used the ratio of total runs scored to total resources consumed as the metric for effective batting in one-day cricket.

Scarf, Shi, and Akhtar (2011) analyzed the partnership performance through the ball-by-ball information of 197 test cricket matches. Based on the partnership score data, they have fitted three statistical distributions, viz. geometric distribution, negative binomial distribution, and the zero-inflated negative binomial distribution. They found that zero-inflated negative binomial distribution is the best model for partnership scores among the above-mentioned distributions. Tan and Zhang (2001) fitted an exponential distribution to analyze opening partnership performance using data from test cricket.

On visiting the available literature, it is found that the measurement of performance of batting partners is mostly restricted to opening partners or to test cricket. No such study on Twenty20 cricket or on the batting partnership irrespective of their position enriched our search. Also, none of the previous studies took into consideration the match situation while evaluating the performance of the batting partners. In this context, the current study is placed which formulates a performance statistic for the batting partnership based on a pressure index of Twenty20 matches.

We consider the World Twenty20 tournament played in India in 2016, for measuring the performance of the batting partners, and to find out the best batting partnership in the tournament. For this purpose, we first redefine our pressure index by changing the values of w_i in (7.3) by w_i^*, i.e., the wicket weight of the i^{th} batsman (to be formally defined below). The values of w_i were based on the wicket weights in Lemmer (2005) and are specific for one-day international matches and not appropriate for Twenty20 matches. Also, in Twenty20 matches, as the bowling resources are only 20 overs, so losing all the wickets in 20 overs is not very common. Keeping this in mind, the wicket weights are computed differently here than in Lemmer (2005).

The ICC rating, of all international cricketers, is available at http://www.icc-cricket.com/player-rankings/profile/ at the end of each match. Using this link, the Twenty20 batting rating of all the players, of all the teams who participated in the super 10 round of the Twenty20 World Cup of 2016 was collected. Their rating on March 7, 2016 was considered as it provides the rating of the players a day prior to the start of the world cup. The batting weight of a particular batsman of a team is given by,

$$w_i^* = \frac{r_i}{\text{sum}(r_i)} \times 11 \tag{7.11}$$

where r_i is the batting rating of the player in Twenty20 as provided in the Web link mentioned above and sum(r_i) is the sum of batting rating of all the 11 players in the concerned match. Thus, w_i^* provides the relative importance of the i^{th} batsman in the team. Thus, as wickets keep falling, the value of $\sum w_i^*$ increases, and it is an indication that the team's wicket strength decreases. Accordingly, we redefine the pressure index as,

$$\mathrm{PI}^* = \left(\frac{\mathrm{CRRR}}{\mathrm{IRRR}}\right) \times \frac{1}{2}\left[\exp(\mathrm{RU}/100) + \exp(\textstyle\sum w_i^*/11)\right] \tag{7.12}$$

Defining PI_{si} as the pressure index value (PI*) at the starting and PI_{ci} as the pressure index value (PI*) at the closing of the i^{th} partnership, which lasted for b_i balls (say), we define

$$\Delta \mathrm{PI}_i = \frac{\sqrt{\mathrm{PI}_{si}} - \sqrt{\mathrm{PI}_{ci}}}{b_i} \times 100 \tag{7.13}$$

as a measure of the pressure differential at the starting and at the closing of the i^{th} partnership for each ball faced multiplied by 100. A batting pair that can diminish the pressure, than what it was at the beginning of their partnership, is characterized by a positive value of $\Delta \mathrm{PI}_i$. The best partnership of the match can be identified by the one which has a maximum value of $\Delta \mathrm{PI}_i$. The same process can also be used to find the best partnership in the entire tournament. However, the actual quality of a partnership may not be properly judged by very few balls. A partnership needs to play for a reasonable period or more precisely a significant number of balls for a meaningful judgment. The effects of outstanding or poor performances in very few deliveries are smoothed over a considerable number of deliveries. However, it is difficult to decide on the significant number of balls, but for Twenty20 cricket, 12 balls may be considered as it is equal to 10 percent of the total bowling resources available. So, any partnership that faced at least 12 balls in an innings can be considered for evaluation. Thus, the i^{th} partnership can be defined as a best partnership (while chasing) in a match or in the tournament if,

$$\Delta \mathrm{PI}_i > \Delta \mathrm{PI}_k, \tag{7.14}$$

where k is any other partnership in the match/tournament that lasted for at least 12 balls.

Table 7.5 shows the 10 best partnerships of second innings in the tournament that survived at least 12 balls to get a positive score of $\Delta \mathrm{PI}_i$, indicating the partnerships that could diminish the pressure for the team batting second.

From Table 7.5, it can be seen that the partnership of Samuels and Brathwaite in the final of the tournament has figured as the best followed by the one by Mathew and Dilshan in their group match against Afghanistan. West Indies figured in four partnerships out of nine, which indicates a reason for their success in the tournament.

7.4.5 Finding the Turning Point of the Match

The pressure curve will sometime indicate a clear turning point of the match if Team B wins. To understand this, let us draw the pressure curve of a very competitive match, which Team B ultimately won. For this, we consider the second Twenty20

Table 7.5 Partnership performance scores of World Cup Twenty20 of 2016 during run chase (second innings)

Pair name	Country	Opponent	ΔPI_i	Comment	Runs	Ball	SR
Samuels_Brathwaite	West Indies	England	6.6533	Final	54	25	216
Mathews_Dilshan	Sri Lanka	Afghanistan	6.5229	Match 16	42	22	190.91
Dhoni_Kohli	India	Pakistan	5.3112	Match 19	35	23	152.18
Dhoni_Kohli	India	Australia	5.1802	Match 31	67	31	216.13
Amla_de Villiers	South Africa	Sri Lanka	4.8618	Match 32	46	27	170.37
Russell_Fletcher	West Indies	Sri Lanka	4.0235	Match 21	55	33	166.67
Buttler_Root	England	New Zealand	3.9205	Semi Final	49	29	168.97
Russell_Simmons	West Indies	India	3.7031	Semi Final	80	39	205.13
Russell_Gayle	West Indies	England	3.6349	Match 15	70	35	200
Maxwell_Marsh	Australia	Bangladesh	1.5337	Match 22	22	12	183.33
Root_Ali	England	South Africa	1.1172	Match 18	33	16	206.25

match between Pakistan and Sri Lanka, played on August 1, 2015, at the Premadasa Stadium, Colombo. Sri Lanka batted first and scored 172/7 in 20 overs. The target for Pakistan was 173. The target was attained by them for the loss of 9 wickets with 4 balls to spare. The pressure curve of the Pakistan innings can be seen in Fig. 7.3.

The pressure curve in Fig. 7.3 demonstrates that Pakistan lost wickets at standard interims and the pressure continued to expand. In any case, fortune turned in the seventeenth over in favor of Pakistan, where 21 runs were scored. Following a six from the second ball, the pressure diminished and the pattern proceeded in the rest of the overs. In this manner, the 97th ball (the first of the seventeenth over) of the Pakistan innings (set apart in Fig. 7.3 with a spotted circle) can be considered as the defining moment of the match. It may so happen that in a match, there can be more than one conceivable points, where the fortune of the match swung in favor of one team from the other. In such a case, the pinnacle or trough, which is nearest to the end of the match, may be considered as a definitive turning point.

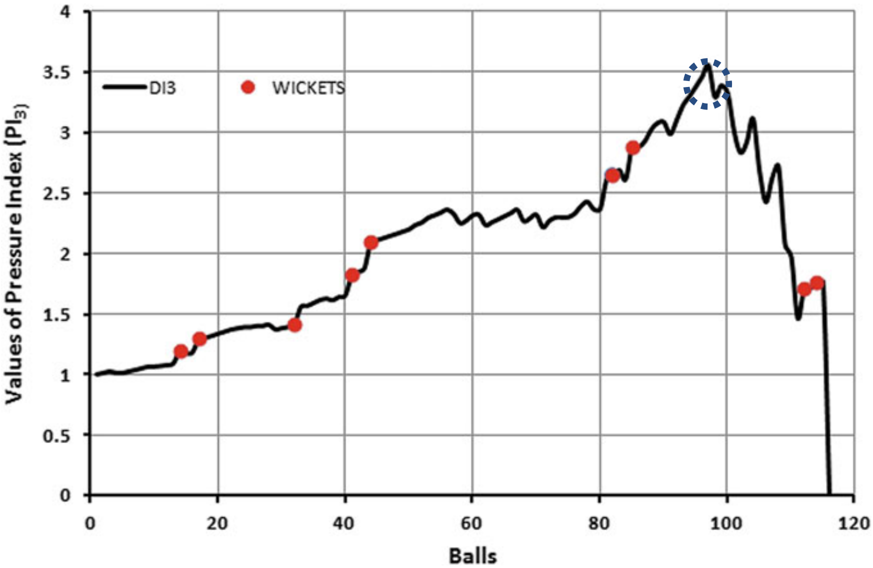

Fig. 7.3 Pressure curve of the Pakistan versus Sri Lanka Twenty20 international match played on August 1, 2015

7.4.6 Comparing Run Chases in Cricket Using the Pressure Index

The pressure index can be used to compare two similar run chases. The idea came to us when one of us went to Vishwa Bharati University, West Bengal, India, and delivered a lecture on the pressure index. After the lecture, a student walked up and told us that the Indian Premier League (IPL) finals of 2012 and 2014 were very exciting and we shall compute the pressure curves for those two matches as well. Suddenly, we remembered that both the finals were won by Kolkata Knight Riders (KKR). In both the matches, KKR won while chasing a huge target and both matches were very evenly poised. There were lots of similarities between the two finals, especially the run chases. Taking a cue from what he told an improvisation was thought. We just pondered, if pressure index can be used to find the similarity between two run chases.

Here, we provide a brief description of the matches to be compared. The IPL final of 2012 was played between Kolkata Knight Riders (KKR) and Chennai Super Kings (CSK) at MA Chidambaram Stadium, Chepauk, Chennai on the May 27, 2012. The toss was won by CSK who decided to bat first. CSK scored 190 runs for the loss of 3 wickets in 20 overs with a run rate of 9.5 per over. The chase started for KKR to achieve a target of 191 runs in 20 overs. KKR chased the target successfully achieving 192 for 5 in 19.4 overs and eventually clinched the league for that year.

After a disappointing campaign in 2013, KKR was once again in the final in the next year, i.e., in 2014. This time their opponent was Kings XI Punjab (KXIP). The final was played at M. Chinnaswamy Stadium, Bangalore, on the June 1, 2014. KKR won the toss and decided to field first. KXIP scored 199 for 4 in 20 overs. The target for KKR was 200 which they achieved in 19.3 overs at the loss of 7 wickets. That year again, KKR won the league.

The pressure curve for the KKR versus CSK final of IPL of 2012 and the KKR versus KXIP of 2014 is plotted in the same graph which can be seen in Fig. 7.4. From the figure, one gets a clear view that the two matches were similar in a number of ways, but the final of 2012 was more competitive. The pressure was similar during the first 100 deliveries of the KKR innings of both the matches. However, after the 100th ball, the pressure increased much more in the 2012 final than that of 2014. Since KKR won both the finals, both the pressure curves ultimately came down to 0. But their path, especially after the 100th delivery, was in two different directions. While the final of 2012 pressure mounted at that stage of the innings, for 2014 it seemed to be more manageable. The curves also show that during the middle period of the match, pressure on the 2014 KKR innings was more than their 2012 innings.

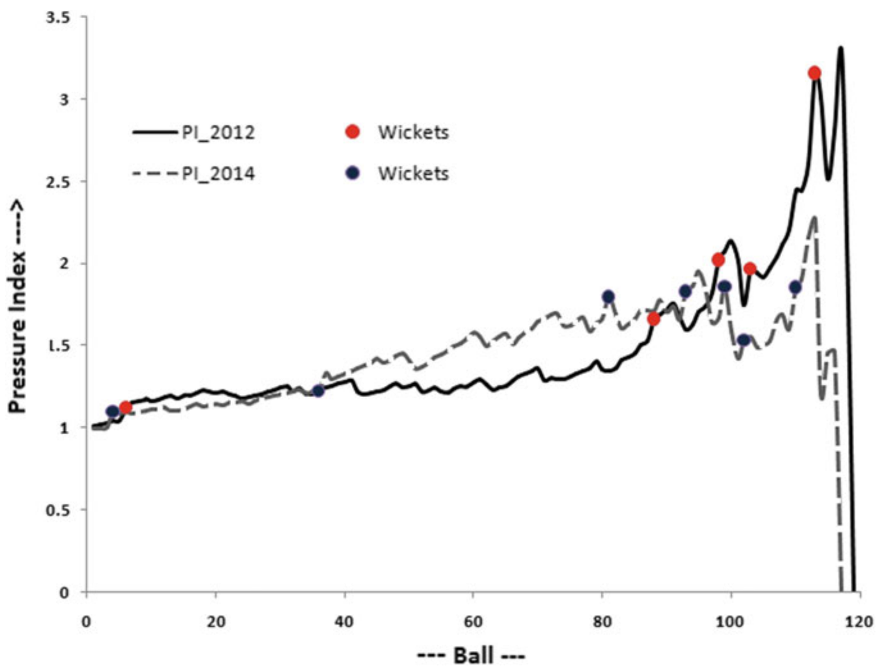

Fig. 7.4 Pressure curves of the two finals won by KKR

7.4.7 Match Outcome Prediction

Though different works on cricket analytics were attempted by researchers at different ages with the pioneering work being that of Elderton and Elderton (1909), but use of analytics for predicting the outcome of matches only appeared in the works of Carter and Guthrie (2002), Allsopp and Clarke (2004), Bailey and Clarke (2006), etc. Of late, some other researchers also ventured in this area, but none of them actually tried to predict the outcome of Twenty20 matches. The Twenty20 format is the shortest version of cricket and has become very popular even in non-cricket playing countries especially among the younger generation because of its speed, short duration, big hitting, and other accessories added to it. The excitement of Twenty20 cricket is more compared to the other formats, as in most games one has to wait until the last over to understand the winner of the match. Most Twenty20 matches head for a close finish and even the under-dogs are found to register wins over the favorites. This makes the job of prediction in Twenty20 matches a difficult proposition.

It is observed that the pressure curve gives a clear indication of how the team batting second is positioned in the match at any given instant. This section uses the pressure index developed by Bhattacharjee and Lemmer (2016) to predict the outcome of international Twenty20 matches.

Predicting the outcome of matches in any sports is a popular exercise these days. With the advances in technology, huge amount of individual, as well as team level, performance-related statistics are now available in case of most of the popular sports. Thus, sports data mining has turned into an emerging area of research. Several papers are produced unearthing the facts that are hidden behind the sports related data.

In tennis, Barnett and Clarke (2002) used two parameters, viz. winning percentage serve of player 1 and winning percentage serve of player 2, as inputs into Markov Chain and predicted the winner of matches played in Australian Open 2003. Clowes, Cohen, and Tomljanovic (2002) used the concept of conditional probability to predict the winner of a tennis match from any stage of the match based on the participating players' probability of winning a point in the game. Somboonphokkaphan, Phimoltares, and Lursinsap (2009), using a set of more extensive features, predicted the outcome of tennis matches with the help of multilayer perceptron and back-propagation learning algorithm. Somboonphokkaphan et al. (2009) also showed that their model provided better performance than the method developed by Barnett and Clarke (2002). Noubary (2005) using long jump data and data for men's 400-m run tried to predict the number of times a new record shall be established in a given time period with the application of Poisson and exponential distribution. Extending the idea of James (1980), Rosenfeld et al. (2010) generalized the Pythagorean win expectation formula by James and estimated the win probabilities of teams in different team sports like basketball, football, and baseball. Heazlewood (2006) made a review of the various prediction models used for quantifying several future Olympic performances in athletics and swimming. Predictions using mathematical models based on pre-1996 athletics and pre-1998 swimming data, the performance for both male and females at the 2000 and 2004 Olympic Games were done. The result showed that the prediction models worked well for short-distance events both in athletics and swimming and the accuracy of predictions decreased with an increase in distance. The factors influencing the probability of victory in major league baseball was studied by Yang and Swartz (2004). Then, using a two-stage Bayesian model and a Markov Chain Monte Carlo algorithm, outcome of future games was predicted.

In football, the application of several machine learning techniques for match prediction has become very popular. Joseph, Fenton, and Neil (2006) studied the performance of an expert constructed Bayesian network compared to other machine learning (ML) techniques for predicting the outcome (win, lose, or draw) of matches played by Tottenham Hotspur Football Club between 1995 and 1997. The study shows that the expert Bayesian network is superior to the other techniques in terms of predictive accuracy. Several other works using machine learning techniques to predict the result of the football match are very frequent. Mention can be made of some of such works like Cui, Li, Woodward and Parkes (2013), Huang and Chang (2010), Hucaljuk and Rakipovic (2011), and Lata and Gupta (2016).

In cricket, not many works related to match prediction enriched our search. The pioneering work in this area in cricket is probably that of Allsopp and Clarke (2004). They used multinomial logistic regression to estimate the probability of team i winning over team j in test matches and one-day internationals (ODI) using the batting, bowling ratings of the i^{th} and j^{th} teams, home team advantage, and first innings lead (only in case of test matches). Scarf and Shi (2005) developed a decision support tool to be used for deciding the timing of declaration of third innings in test cricket. The authors compute the probability of victory for various declaration times using a multinomial logistic regression model. Gathering match information from 2200 one-day international matches and using a multiple linear regression model Bailey and Clarke (2006) found that predicted run totals and match outcome are related to home ground advantage, past performances, match experience, performance at the specific venue, performance against the specific opposition, experience at the specific venue, and current form. Using the model, the margin of victory and team total was predicted. Using binary logistic regression, Bandulasiri (2008) found that the probability of victory in one-day international matches is significantly influenced by home ground advantage, the outcome of the toss, and interaction between toss outcome and day-night matches. Thus, the study suggested that the match outcome can be predicted using these significant factors. Lemmer (2015), taking data from the group stage of 2011 ICC World Cup, using a logistic regression model tried to predict the outcome of the matches played at the knock out stage of the tournament.

Some other work not directly related to match outcome prediction in cricket but related issues comprise of Carter and Guthrie (2002) and Ramakrishnan, Sethuraman, and Parameshwaran (2010). Both the works attempted to predict the target score in cricket matches. O'Riley and Ovens (2006) compared several existing target rescheduling methods and eventually tested if the methods can be extended to predict the final score based on the score at any given point of the innings. They concluded that the Duckworth/Lewis method is readily adaptable both as a predictive and a target resetting tool. Considering longitudinal data from test cricket, Wickramasinghe

(2014) applied a three-stage hierarchical linear model and predicted the performance of batsmen in test cricket.

The Twenty20 format of cricket is the most recent one at the international level. With the format gaining popularity, several franchisees-based Twenty20 cricket tournaments has started around the globe. This has increased the interest of spectators and several studies related to Twenty20 cricket has also figured in research journals. But the issue of match outcome prediction on this format is never attempted. Considering the uncertainty associated with Twenty20 cricket, where most matches go down to the last over to get the winner, match outcome prediction must be a difficult task to attain. Thus, this work based on a pressure index developed by Bhattacharjee and Lemmer (2016) makes an attempt to predict the outcome of Twenty20 cricket matches.

At the start of any run chase during the second innings of limited overs match the value of the pressure index remains at unity. With every delivery, the value of the pressure index is computed using (7.5).

When the pressure index values are plotted against each ball, the pressure curve is generated. As the match situation becomes difficult for the team batting second, the pressure curve keeps on rising and eventually drops down reaches zero if the team batting second wins the match. But, if the team batting second loses the match, the pressure curve rises steeply. Figure 7.5 depicts the pressure curves of two different Twenty20 matches. The first of the matches were played between West Indies and Australia on March 30, 2012 at Kensington Oval, Bridgetown, Barbados. West Indies batted first and were all out for 160 in 19.4 overs. In reply, Australia could score only 146 for 9 in their allotted 20 overs and eventually lost the match. The corresponding

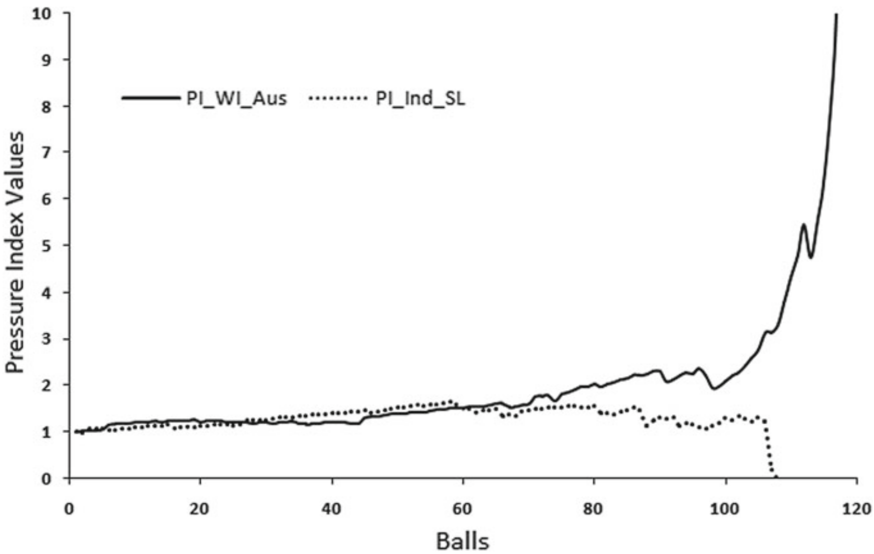

Fig. 7.5 Typical pressure curves of two Twenty20 matches

pressure index values of Australia is plotted to get the pressure curve shown by the dark black line. The pressure curve keeps on rising higher and higher as the match situation becomes more and more difficult for Australia. The other pressure curve is a match between India and Sri Lanka played at Pune on February 9, 2016. India were 101 all out in 18.5 overs and in reply, Sri Lanka easily reached the target in 18 overs with five wickets in hand. The corresponding pressure curve of Sri Lanka's run chase is shown by the dotted line in Fig. 7.5. It can be seen that the pressure curve drops down and eventually reaches zero as Sri Lanka wins the match. A similar trend like that of the dotted line is noted in case of all pressure curves where the team chasing comes out victorious and that of the dark line is observed in case of pressure curves where the chasing team loses.

If the pressure index values of an innings are provided up to a given point, is it possible to predict where the pressure curve shall ultimately go? Thus, given the performance of the team during run chase up to a certain point, say half-way of the second innings, the problem is to predict if the pressure curve shall come down to zero (indicating victory of the chasing team) or show an upward raise at the end of the innings (indicating that the chasing team has lost the match).

Thus, win and loss are the two possibilities and they are complementary to each other. The outcome of the match can be predicted if one can compute the probability of winning a match by a team gave the value of pressure index values up to a given instant of the second innings of that match. The outcome variable is either win or loss in a match and the explanatory variables are Target that is set for Team B and the pressure index values at different instant of time. This calls for the application of binary logistic regression with which the probability of winning a match by a team batting second can be computed. The binary logistic regression can be used if the outcome variable is binary in nature (having only two outcomes—1 and 0) and the explanatory variables can be continuous, binary, or categorical. For more information on binary logistic regression, refer to Appendix 7.2 at the end of this chapter.

In this exercise, we tried to fit three different binary logistic equations. The explanatory variables are—target score set by the team batting first in a Twenty20 match for the team batting second (X_1), pressure index values at the end of the 6th over, 10th over, and the 15th over of second innings of the said match to be denoted as X_2, $X_{3,}$ and $X_{4,}$ respectively. A large sample of complete Twenty20 matches is then considered. The runs scored at the end of first innings of the match are increased by one to get the value of X_1. Based on the ball by ball information of the second innings of these matches the pressure index values during the run chase are computed. The pressure index values of the second innings, at the end of the 6th, 10th, and the 15th over are then retained as values of X_2, $X_{3,}$ and $X_{4,}$ respectively. The logistic regressions are planned as follows:

(a) The probability of winning by Team B as the outcome variable fitted against the target score for Team B (X_1) and pressure index at the end of 6th over (X_2) as the explanatory variables.

(b) The probability of winning by Team B as outcome variable fitted against the target score for Team B (X_1) and pressure index at the end of 10th over (X_3) as the explanatory variables.
(c) The probability of winning by Team B as outcome variable fitted against the target score for Team B (X_1) and pressure index at the end of 15th over (X_4) as the explanatory variables.

Some rationale exists for choosing the 6th, 10th, and 15th over as landmark overs. The 6th over marks the end of the powerplay overs. The first six overs in an innings of a Twenty20 match are called powerplay overs, when there are some fielding restrictions and the batting team tries to take advantage of these restrictions by scoring runs at an additional pace. However, there is no specific reason for choosing the 10th and the 15th over. But at the end of the 10th over, the team batting second completes half of their journey and the target is to be achieved in the remaining 10 overs. Likewise, at the end of the 15th over, the team batting second consumes 75% of their available bowling resources and most of the twists of close or successful run chases appear after that. It shall be interesting to note how successful the predictions are as the match progresses. However, inquisitive readers may try prediction at any other point of the second innings.

To understand the working of the model, we attempted the following exercise. The training sample for the model consists of 118 international Twenty20 matches played between February 2012 to February 2017 and the holdout sample contains all the 23 matches played during the Twenty20 World Cup in March-April, 2016. Obviously, the training sample does not contain those matches which are considered in the holdout sample. The computation is done using the package SPSS 22.0.

The results of the match, either win (1) or loss (0), is measured in terms of Team B, i.e., the team that bats second in a given Twenty20 match. The explanatory variables are as described in (a), (b), and (c) above. The results are as follows:

The explanatory variables in the model (a) are the target score for Team B (X_1) and pressure index at the end of 6th over (X_2). First, we test the independence between target score for Team B (X_1) and pressure index at the end of 6th over (X_2), one of the important assumptions of the technique. The correlation between X_1 and X_2 is 0.18 (p-value of $0.052 < 0.05$) which is insignificant. The values of Cox and Snell R^2 and that of Nagelkerke R^2 are 0.352 and 0.469, respectively, which are relatively high indicating a good fit. Both of them are somewhat similar to that of R^2 of multiple regression and can be termed as pseudo R^2 of logistic regression. Any value of more than 0.3 shall be assumed to provide a good fit. The Hosmer and Lemeshow test[1] results to a value of 12.67, which is a χ^2 statistic with 8 df providing a corresponding

[1]The test is commonly used for assessing the goodness of fit of a model. The test is similar to a χ^2 goodness of fit test and has the advantage of partitioning the observations into groups of approximately equal size. The observations are grouped into deciles based on their predicted probabilities. The test statistic is calculated based on the observed and expected counts for both the wins and losses. The statistic has an approximate χ^2 distribution with 8 degrees of freedom (=10 deciles—2 variables in the model). For this case the p value of the test $= 0.123$ indicates that the number of wins are not significantly different from those predicted by the model and that the model is a good fit.

Table 7.6 Result of binary logistic regression for values of pressure index up to 6th over

Variables in model	B	SE	Wald	df	Significance	Exp(B)
Target (X_1)	−0.039	0.01	16.22	1	0.00	0.962
PI_6 (X_2)	−5.381	1.269	17.98	1	0.00	0.005
Constant	13.163	2.545	26.74	1	0.00	520693.4

p-value of 0.123 (>0.05) favoring the fit. The fitted model could predict 95 out of the 118 cases of the training successfully. The ultimate regression table is as shown in Table 7.6.

It is seen from Table 7.6 that (which is based on the training sample), both the coefficients associated with Target (X_1) and PI_6 (X_2) are significant[2] and negative. This indicates that as pressure index values increase, the probability of winning the match decreases. The same can be said about the variable Target (X_1) as well. The value of Exp(B) corresponding to Target (X_1) is 0.962, this indicates that if the target increases by one run the odds of winning the match decreases by 3.8%. But, if pressure index at the end of 6th over increases by one the odds of winning the match decreases by 99.5% (corresponding value of Exp(B) being 0.005). Thus the ultimate equation for computing the probability of winning a match by the team batting second in Twenty20 match given the target they are chasing (X_1) and the pressure index value at the end of the 6th over of their run chase i.e. PI_6 (X_2) is given by,

$$P(\text{Win}) = p = \frac{e^{13.163-0.039\times X_1-5.381\times X_2}}{1+e^{13.163-0.039\times X_1-5.381\times X_2}} \tag{7.15}$$

The equation can be used to predict the outcome of a Twenty20 match at the end of the 6th over of the second innings given the values of Target (X_1) and PI_6 (X_2). If the value of P(Win) for a given set of values of X_1 and X_2 is more than 0.5 then it means a 'win' is more probable than a 'loss' for the team batting second and vice versa. For example, if a team chasing 165 runs in a Twenty20 match reaches a Pressure Index value of 1.43 at the end of 6th over, then we have,

$$p = \frac{e^{13.163-0.039\times X_1-5.381\times X_2}}{1+e^{13.163-0.039\times X_1-5.381\times X_2}} = \frac{e^{13.163-0.039\times 165-5.381\times 1.43}}{1+e^{13.163-0.039\times 165-5.381\times 1.43}} = 0.2755$$

As the value of p is less than 0.5 so we shall predict that the chasing team is expected to lose the match. The fitted model in (7.15) is now used to predict the outcome of all the 23 matches played during the Twenty20 World Cup in March–April, 2016. These are the matches of the holdout sample as they were not used to estimate the parameters of the model (Table 7.7).

[2]The significance can be ascertained from the Wald statistic which follows χ^2 statistic with 1 df. The Wald χ^2 statistics are used to test the significance of individual coefficients in the model and are calculated by $\left(\frac{B}{SE(B)}\right)^2$

Table 7.7 Performance of binary logistic regression in predicting the outcome of the matches of the training sample

Match number	Teams	Target	PI 6	Actual result	Estimated probability	Prediction	Miss classification
Semi Final 1	NZ_Eng	154	0.9194	Win	0.9011	Win	
Semi Final 2	Ind_WI	193	1.5195	Win	0.0730	Loss	Yes
20	SA_Afg	210	1.1578	Loss	0.2215	Loss	
Final	Eng_WI	156	1.6119	Win	0.1687	Loss	Yes
15	Eng_WI	183	1.2153	Win	0.3743	Loss	Yes
17	NZ_Aus	143	1.1781	Loss	0.7768	Win	Yes
18	SA_Eng	230	1.2182	Win	0.0861	Loss	Yes
19	Pak_Ind	118	1.7428	Win	0.3064	Loss	Yes
21	SL_WI	123	1.1793	Win	0.8829	Win	
22	Ban_Aus	157	1.0988	Win	0.7554	Win	
13	NZ_Ind	127	1.8254	Loss	0.1663	Loss	
14	Pak_Ban	202	1.4859	Loss	0.0624	Loss	
16	Afg_Sl	154	1.2876	Win	0.5568	Win	
23	NZ_Pak	181	1.1062	Loss	0.5378	Win	Yes
24	Eng_Afg	143	1.5990	Loss	0.2654	Loss	
25	Ind_Ban	147	1.1877	Loss	0.7387	Win	Yes
26	Aus_Pak	194	1.5090	Loss	0.0743	Loss	
27	SA_WI	122	1.3538	Win	0.7541	Win	
28	NZ_Ban	145	1.4412	Loss	0.4385	Loss	
29	Eng_SL	172	1.8484	Loss	0.0296	Loss	
30	Afg_WI	124	1.3139	Loss	0.7785	Win	Yes
31	Aus_Ind	161	1.2430	Win	0.5487	Win	
32	SL_SA	121	1.2826	Win	0.8238	Win	

Notes: (i) The team named second in column 2 is the one batting second in the corresponding Twenty20 match and the win or loss is relative to the team batting second
(ii) Match number: The holdout sample contains information about Twenty20 matches that were played in Twenty20 World Cup in 2016 in India. This column contains the identification number of the matches
(iii) Teams: The teams that played against each other in the corresponding Twenty20 matches. The details are Eng: England, SA: South Africa, NZ: New Zealand, Afg: Afghanistan, Pak: Pakistan, SL: Sri Lanka, WI: West Indies, Aus: Australia, and Ban: Bangladesh. The team that batted in the second innings of the match is named later
(iv) Target: The runs to be scored in the second innings of the match by the team batting second in order to win the match
(v) PI6: Value of the pressure index of the team batting second at the end of the 5th over of their run chase
(vi) Actual result: The actual outcome of the match. Here, win or loss is relative to the team batting second
(vii) Estimated probability: The value of p obtained by the application of binomial logistic regression Eq. (7.15)
(viii) Prediction: The predicted outcome of the match obtained from the estimated probability of win from the corresponding logistic regression equation
(ix) Misclassification: 'Yes' indicates that the outcome is misclassified by the prediction model

The success rate is around 61% which is compatible with the cross-classification exercise, for the training sample which resulted in 81% correct classification. We are satisfied with the overall functioning of the prediction rule. Information of the second innings up to 6 overs out of 20 overs is used for predicting the outcome of a match.

Initially, we thought that when the same exercise shall be repeated for (b) and (c), better results shall be obtained. But things happened differently. For (b), the two explanatory variables target score for Team B (X_1) and pressure index at the end of 10th over (X_3) were found to be related to each other with a correlation coefficient of 0.311 (p-value of 0.001). This was a violation of one of the assumptions of logistic regression where the explanatory variables needed to be independent of each other (cf. Appendix 7.2). Hence the model described in (b) cannot be fitted.

Next, we tried fitting (c). This time, the two explanatory variables target score for Team B (X_1) and pressure index at the end of 15th over (X_4) were found to be independent of each other with a correlation coefficient of 0.005 (p-value of 0.956). The values of Cox and Snell R^2 and that of Nagelkerke R^2 are 0.192 and 0.256, respectively, which are relatively low indicating a poor fit (both values less than 0.3). Though, the Hosmer and Lemeshow test results to a value of 11.6, which is a χ^2 statistic with 8 df providing a corresponding p-value of 0.17 (>0.05) favoring the fit. But the ultimate regression table tells that the pressure index at the end of 15th. over (X_4) is an insignificant variable in the model with a p-value of 0.657. Thus, model (c) also could not be recommended for predicting the outcome of matches using the pressure index. However, (a) can be used for prediction and it works well. Readers may try for testing the capability of prediction of the binomial logistic regression at other points of the match. One may also try other multivariate models of classification or discrimination for predicting the results of match outcome based on pressure index.

7.4.8 Identifying the Most Resilient Cricket Team

As cricket is a team game, it is imperative to emphasize the study of team performances. One of the earliest studies in this direction is due to Clarke (1988). In his study, Clarke attempts to predict the optimal scoring rates of a team for successfully chasing a target with the help of dynamic programming. The paper concludes that teams have higher chances of winning a test match if they opt for batting in the first innings and score runs at a faster rate than the expected run rate. In their study, Preston and Thomas (2000) use the data from limited overs cricket matches of English County teams and suggest that teams are at a better position if they start with a slow run rate and gradually accelerate their run rate so that wickets are not lost untimely. In a study in this direction, Allsopp (2005) develops a method for estimating the projected score of the team which has been chasing a prefixed target during the second innings of ODI cricket match under home ground advantage. Barr et al. (2008) put forward a world cricket team by studying the performance of players participated in the 2007 World Cup and provide some strategic suggestions so that other teams can compete well enough with the Australian team. Using an ordered response model, to the data of test cricket matches of the years 1994 to 1999, Brooks, Faff, and Sokulsky (2002) have shown that simple batting and bowling measures can predict test match performance of teams to a large extent. Allsopp and Clarke (2004) analyze the batting and bowling performances of teams playing both innings of one-day international cricket matches and the first innings of test matches with the help of multiple regression techniques. For predicting match outcomes, the authors use different independent explanatory variables, viz. the strength of a team in terms of batting and bowling in the first innings, team's first innings lead, batting line-up and playing a match in a home or away ground. The conclusion is that in test matches, teams batting in the second innings are better placed than the other team batting first and there is no evidence to suggest that teams have an added advantage by winning a toss in case of test cricket matches. Douglas and Tam (2010) have examined the team performances using several key variables in terms of batting, bowling, and fielding for all winning and losing teams played in the Twenty20 World Cup 2009. They suggest that for victory in a Twenty20 match, teams should dismiss wickets fast, bowl dot balls, and emphasize on 50+ partnerships along with hitting boundaries while batting. Daud and Muhammad (2011) introduce an index called Team Index (T-Index) to rank cricket teams through runs and wickets. The teams are ranked by the principle that a team winning a difficult match, i.e., against a stronger opponent would be awarded more points than winning against a weaker team utilizing the number of runs, wickets, and also the outcome of the match (i.e, win or loss). Dey et al. (2015) analyze the team performance with the help of multicriteria decision algorithms under fuzzy environment using the data in case of limited overs cricket. To study the batting performance of Indian cricketers during 1985 to 2005 in one-day cricket, Damodaran (2006) applied stochastic dominance rules, which are normally used in investment management, by viewing players as securities and the team as a portfolio.

The pressure index can also be used to measure the team performance and can be utilized to compare different teams in a given time frame and also can compare the same team at different time periods. The pressure index can be utilized to find out the most resilient team, i.e., a team which has the ability to win the match at difficult as well as easy situations while chasing. The process of finding out the most resilient team is based on the pressure index computed for several matches.

By now, we are aware of the fact that if the pressure values computed based on (7.5) are plotted against each ball, the resultant curve obtained is called the pressure curve. The typical pressure curves—one for a win to the team batting second (the dotted line) and the other for a loss to a team batting second (the solid line)—look similar to the curves as in Fig. 7.5.

Let the pressure curves of all the run chases of a given cricket team in Twenty20 internationals during a time period, say 2012–2016 be considered in the same graph. λ_i and λ_j be the two possible values of the pressure indices with $\lambda_i < \lambda_j$. The closed interval $[\lambda_i, \lambda_j]$ be termed as the pressure zone (i, j). Let X_k represent a binary random variable taking the values,

$$X_k = 1, \quad \text{if the } k^{\text{th}} \text{ match is won by the team batting second}$$
$$= 0, \quad \text{otherwise}$$

Let n be the number of visits and ξ be the number of visits that eventually led to victory, by all the pressure curves, in the given time period, to the pressure zone (*i, j*). Thus, the relative frequency $\frac{\xi}{n}$ acts as an estimate of the probability of winning for a team, given the pressure curve reaches the pressure zone (*i, j*) i.e.,

$$P\left[X_k = 1 | \lambda_i < \text{PI} < \lambda_j\right] = \frac{\xi}{n} \tag{7.16}$$

As per the definition of the pressure index, PI $\in$ [0, ∞]. But on observing the pressure curve of a large number of matches, some common trends are visible:

(i) If the pressure curve during any run chase reaches values less than 0.5, then the match is ultimately won by the team batting second, i.e., $P[X_k = 1|0 < \text{PI} < 0.5] = 1$
(ii) If the pressure curve during any run chase reaches values more than 3.5, then the match is ultimately lost by the team batting second, i.e., $P[X_k = 1|\text{PI} > 3.5] = 0$
(iii) For the different subintervals of PI values between [0.5, 3.5], viz. $[0.5, \lambda_{i_1}], [\lambda_{i_1}, \lambda_{i_2}], \ldots, [\lambda_{i_m}, 3.5]$, the values of $P[X_k = 1|\lambda_i < \text{PI} < \lambda_j]$ generally starts with 1 but gradually keeps on decreasing and eventually reaches 0.

So, when the values of $P[X_k = 1|\lambda_i < \text{PI} < \lambda_j]$ are plotted against the mid value of the pressure zone $[\lambda_i, \lambda_j]$, the graph is expected to produce the mirror image of an elongated '*S*.' However, the observed curve is found to produce a zigzag path, around a hypothetical curve, the mirror image of an elongated '*S*' with constant values at the tails, viz. 1 and 0. Thus, a gradual decline is seen between the constant values, 1 and 0, like that of the decay curve (negative growth). It is obvious from the above discussion that nonlinear regression models shall be used to explore the relationship of the said probability function with the pressure index. An appropriate nonlinear model that can be fitted to such a zigzag line generated by observed values of $(\frac{\lambda_i+\lambda_j}{2}, P[X_k = 1|\lambda_i < \text{PI} < \lambda_j])$ is the Farazdaghi and Harris (1968) yield–density equation. The yield–density equation of Farazdaghi and Harris had its origin from the equation of Shinozaki and Kira (1956) which is the yield–density equation derived from logistic curves of plant population growth or decay. They considered the logistic differential equation

$$\frac{\mathrm{d}Y}{\mathrm{d}t} \cdot \frac{1}{Y} = \lambda\left(1 - \frac{Y}{Y_\infty}\right) \tag{7.17}$$

with Y the crop yield per unit area, X is the density (plant population in number of plants sown per unit area), $w = Y/X$ = mean crop yield per plant and q the growth (rate) coefficient λ is independent of X and both λ and Y_∞ are independent of t, subject to the law of constant final yield,

$$Y_\infty = w_\infty X = k = \text{constant} \tag{7.18}$$

Under the above assumptions, Shinozaki and Kira (1956) derived yield density function as,
$w = \frac{1}{a+bX}$, where a and b are the parameters, such that

$$\frac{\mathrm{d}w}{\mathrm{d}X} = \frac{-b}{(a+bX)^2} \quad \text{and} \quad w \to a^{-1} \quad \text{as} \quad X \to 0 \tag{7.19}$$

Following Shinozaki and Kira (1956), Farazdaghi and Harris (1968) developed their yield–density equation from the same logistic curve subject to the modification that the law of constant final yield holds to,

$$Y_\infty = w_\infty X^c = q = \text{constant} \tag{7.20}$$

It can be observed that (7.20) is equivalent to (7.18) if $c = 1$. Accordingly, Farazdaghi and Harris forwarded their growth model (yield–density curve) as,

$$w = \frac{1}{a + bX^c} \tag{7.21}$$

With a, b, and c as parameters of the model estimated from the data,

$$\text{Now, } \frac{\mathrm{d}w}{\mathrm{d}X} = \frac{-bcX^{c-1}}{(a+bX^c)^2} \langle 0, a, b \rangle \quad \text{and} \quad c > 1 \tag{7.22}$$

Also, $w \to a^{-1}$ as $X \to 0$.

Since the above differential is less than zero, w is a decreasing function of X. Thus, the Farazdaghi and Harris model shows either asymptotic or parabolic yield–density behavior, depending on the value of c. If $c > 1$, then the curve shows a parabolic pattern. Replacing, $P[X_k = 1|\lambda_i < PI < \lambda_j]$ and $(\lambda_i + \lambda_j)/2$ in place of w and X in (7.21), we get the Farazdaghi–Harris curve as,

$$P[X_k = 1|\lambda_i < PI < \lambda_j] = \frac{1}{a + b\left(\frac{\lambda_i+\lambda_j}{2}\right)^c} \tag{7.23}$$

where a, b, and c are constants to be determined from data. A popular method for estimating the parameters of the nonlinear regression function is the method of least squares. According to this method, the estimates of a, b, and c are obtained by minimizing the quantity $S = \sum_{k=1}^{n} e_k^2$, where e_k represents the difference between the actual value and the estimated value of $\mathrm{P}[X_k = 1|\lambda_i < \mathrm{PI} < \lambda_j]$, with n representing

Table 7.8 Twenty20 international cricket matches chased by selected teams from 2012 to 2016

Country	No. of matches with run chase[a]	Victorious	Percentage of victory while chasing	No. of matches batting first[a]	Victorious	Percentage of victory while batting first	Overall percentage of victory
India	25	17	68.00	21	13	61.9	65.21
Australia	23	11	47.83	16	7	43.75	46.15
West Indies	15	9	60.00	29	16	55.17	56.81
New Zealand	17	7	41.18	25	15	60.00	52.38

Source of data www.espncricinfo.com
[a]Tied and truncated matches were not considered

the number of observations. The estimation of parameters of the nonlinear regression model, which cannot be converted to linearizable models through transformation, is not as straightforward as in case of linear/linearizable models. The Farazdaghi and Harris model discussed above is one such model. The estimation of the parameters of such nonlinear models usually requires the use of iterative methods on digital computers, an explicit formula for estimating the parameters of the models are not generally available. Most commonly available statistical software packages provide routines for calculating the estimates of parameters of the model (Graybill & Iyer, 1994). In this case, the package Curve Expert is used to fit the model.

Accordingly, the Farazdaghi–Harris curve as in (7.23) is fitted to the matches played by four international Twenty20 cricket teams, viz. New Zealand, India, West Indies, and Australia during the period 2012–2016. The matches considered comprise of all the full Twenty20 matches in which the aforementioned teams chased the target during the time period mentioned above. The teamwise number of matches and relevant details are provided in Table 7.8.

Ball-by-ball information of all the 80 run-chases is collected from espncricinfo.com and accordingly the pressure index values of all the run chases are computed. Accordingly, using (7.16), the values of $\mathrm{P}[X_k = 1|\lambda_i < PI < \lambda_j]$ are computed for each of the teams separately and are laid down in Table 7.9.

The estimated parameters of the fitted Farazdaghi–Harris curves based on Table 7.9 are presented in Table 7.10. The fitted Farazdaghi–Harris curve is mentioned here as the resilience curve, as the curve gives the idea about the probability of winning of the team at different pressure levels.

The values of coefficient of determination, i.e., R^2, as observed from Table 7.10, indicates that the fitting of the data to the models is good enough.

Figure 7.6 shows all the four resilience curves for the four international Twenty20 teams, viz. India, Australia, West Indies, and New Zealand based on matches played between 2012 and 2016. It is observed that the Australian team performs poorly compared to all other teams. The probability of victory is less than one when the pressure is less and converges to zero quicker than the other teams as pressure increases. India

Table 7.9 Pressure index with corresponding probabilities of victory of different international Twenty20 teams

Pressure index values		The probability of winning of teams			
Lower (λ_i)	Upper (λ_j)	India	New Zealand	West Indies	Australia
0	0.1	1	1	1	0.947368
0.1	0.2	1	1	1	0.923077
0.2	0.3	1	1	1	1
0.3	0.4	1	1	1	1
0.4	0.5	1	1	1	0.692308
0.5	0.6	1	1	1	0.857143
0.6	0.7	1	1	1	0.97561
0.7	0.8	1	1	1	0.955224
0.8	0.9	1	1	0.93333333	0.913043
0.9	1	0.965517	0.711538	0.97260274	0.760331
1	1.1	0.820809	0.511551	0.69361702	0.555276
1.1	1.2	0.65404	0.375969	0.78333333	0.515464
1.2	1.3	0.81701	0.393258	0.64166667	0.56015
1.3	1.4	0.76087	0.536913	0.79746835	0.468085
1.4	1.5	0.696335	0.276923	0.69387755	0.428571
1.5	1.6	0.736434	0.486842	0.74766355	0.329897
1.6	1.7	0.701923	0.573171	0.65432099	0.411765
1.7	1.8	0.476636	0.409091	0.60714286	0.549296
1.8	1.9	0.458333	0.403509	0.5	0.191176
1.9	2	0.56	0.30303	0.46	0.307692
2	2.1	0.489796	0.441176	0.47619048	0.285714
2.1	2.2	0.369565	0.365854	0.56756757	0.142857
2.2	2.3	0.21875	0.285714	0.35714286	0.027778
2.3	2.4	0.294118	0.074074	0.3125	0
2.4	2.5	0.611111	0.086957	0.26829268	0
2.5	2.6	0.285714	0.125	0.23529412	0
2.6	2.7	0.266667	0	0.32142857	0.071429
2.7	2.8	0.384615	0	0.33333333	0
2.8	2.9	0.222222	0	0.2	0.071429
2.9	3	0.166667	0	0.1875	0
3	3.1	0.142857	0.333333	0.42857143	0
3.1	3.2	0.3	0	0.5	0
3.2	3.3	0.076923	0.052632	0.25	0
3.3	3.4	0	0	0.125	0

(continued)

Table 7.9 (continued)

Pressure index values		The probability of winning of teams			
Lower (λ_i)	Upper (λ_j)	India	New Zealand	West Indies	Australia
3.4	3.5	0.090909	0.0625	0.25	0
3.5	3.6	0	0	0.33333333	0
3.6	3.7	0	0	0.33333333	0
3.7	3.8	0	0	0.11111111	0
3.8	3.9	0	0	0.33333333	0
3.9	4	0	0	0	0

Table 7.10 Farazdaghi–Harris model (resilience curve) summary

Teams	Values of coefficients			R^2
	a	*b*	*c*	
India	1.1169	0.0559	4.0772	0.86
Australia	1.2273	0.1675	4.6441	0.84
West Indies	1.1276	0.0946	3.1051	0.91
New Zealand	1.1284	0.2757	3.3459	0.83

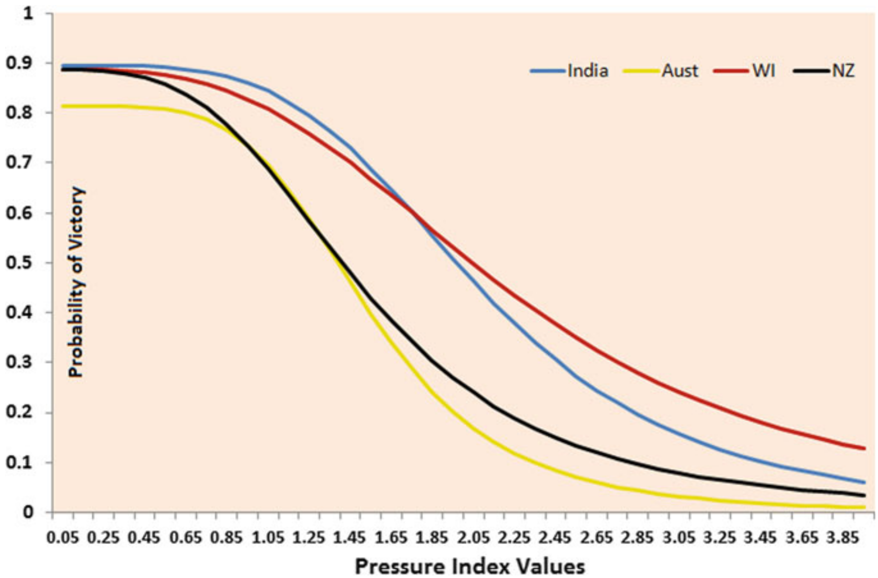

Fig. 7.6 Resilience curves of the four international Twenty20 teams

has the maximum probability of winning till the pressure is manageable around 1.75 but with an increase in pressure, the probability of victory is less than that of West Indies. The resilience curve of the West Indies cricket team is very interesting. It remains below the Indian curve till 1.75 following which it supersedes the Indian curve. This indicates that West Indies has less chance of winning matches from an easy situation while chasing than India but their probability of winning matches from a difficult situation is more than any other country considered in the study. This may be attributed to the fact that West Indies team had several all-rounders like Darren Sammy, Andre Russell, Sunil Narine, Dwayne Bravo, etc., who played between 2012 and 2016. These players were good finishers and have pulled several matches in favor of West Indies while chasing by their capability of slog hitting. But the case of India was different, the lower middle order could not show the capability that West Indians showed and so India could win matches only when the top-order batsmen fired. India has won fewer matches when the pressure increased beyond a limit and so their resilience curve came closer to X-axis as the pressure mounted.

7.5 Computing Pressure Index for the First Innings

With the prolonged discussion on pressure index and its uses, the readers must be asking, 'What about pressure index for the first innings of a limited overs match?' Incapability of computing the pressure index for the first innings shall lead to restricted use of the index. Computing pressure index for the first innings is a challenge and is not as straightforward as it is in the second innings. This is because, unlike the second innings, there is no fixed target in the first innings. The pressure index formula is extensively dependent on the target score. Thus, in the absence of a target score, the pressure index in the first innings seems difficult to attain. But it is not reasonable that the team batting first does not have any target in mind and keeps on hitting the ball all around the park and score maximum runs they can. Indeed, a team that bats first set themselves a target based on their assessment of pitch and match conditions, the strength of their opponents, weather, etc. They start batting and after some time they realize that their target was either too high or too low and they then adjust their target. Several times during the post-match presentation captains expressed such views. One such example is as follows:

In the third Twenty20 of the series played between England versus India at Bengaluru on the of February 1, 2017, India batted first and scored 202 for 6 in their 20 overs and in reply England were all out for 127 in 16.3 overs. India eventually won the match by 75 runs. In that match, while India was batting, at the end of 17th over, the score was 153/3 the run rate was 9 runs per over. But in the 18th over bowled by Chris Jordan, with M. S. Dhoni and Yuvraj Singh at the crease, the runs scored were 1, 6, 6, 4, 6, and 1. A total of 24 runs in that over took the scoring rate close to 10 runs per over. After the 17th over, the possible target at the end of the Indian innings was expected to be around 180, eventually went past 200 in 20 overs. This is also complemented by a comment made by the India captain Virat Kohli during the presentation ceremony of the match, '*... the one over from Chris Jordan to Yuvi, that was the momentum changer for us. We were thinking about 175 to 180 but that's the ability he has. He pushed us up 200.*'

Thus, we assume that the first innings target is not a constant but a variable, and it is rescheduled based on the match situation. But, we need to define an initial target that the team batting first thinks of attaining while they start the innings. It is generally based on the previous international Twenty20 matches played in that venue. The initial target is taken to be $T_0 = \mu + \sigma$, where μ is the average runs scored by the teams batting first in the international Twenty20 matches played in the same venue and σ is the corresponding standard deviation. If the runs scored in the first innings of international Twenty20 matches are considered to be normally distributed, then there is only a 16% chance that a team batting first shall score more than T_0. Accordingly, T_o is assumed to be the initial target for the team batting first in an international Twenty20 match.

With the fall of every wicket, a new partnership starts and so it is expected that the new batting partner comes in with some instructions from the dressing room about how to approach the game given the current match situation.

Accordingly, the following formula for the rescheduled target score at the end of each partnership is proposed:

$$T_1 = T_0$$

and

$$T_t = R_{t-1} + (100 - RU_{t-1}) \times \text{Max}\left(\frac{T_0}{B}, \frac{R_{t-1}}{B_{t-1}}\right) \quad \text{for} \quad t = 2, 3 \tag{7.24}$$

where R_{t-1}, $RU_{t-1,}$ and B_{t-1} denote the total runs scored and the resources utilized and the number of balls bowled up to the end of the most recently concluded partnership, respectively.

However, a situation may so arise that no wicket falls. In such a case, the team batting first shall definitely reschedule its targets depending on the progress of the team at different time points. In such cases, the target rescheduling can be done at the end of 6th over (end of the powerplay), 10th over (half way of their innings), and during the 15th over (3/4th of their innings), using the formula in (7.24).

The other calculations shall remain similar to those of the pressure index defined in (7.5), with T replaced by T_t in the calculation of CRRR. Thus, the basic difference between the computation of the pressure index in the first innings and the second innings is that while the target score remains fixed for the team batting second, for the team batting first, the target is rescheduled based on the match situation. However, the performance statistics computed based on the pressure index of the first innings shall not be compared with those of the second innings. This is because the target score of the first innings is estimated and is updated during the innings and that of the second innings is fixed.

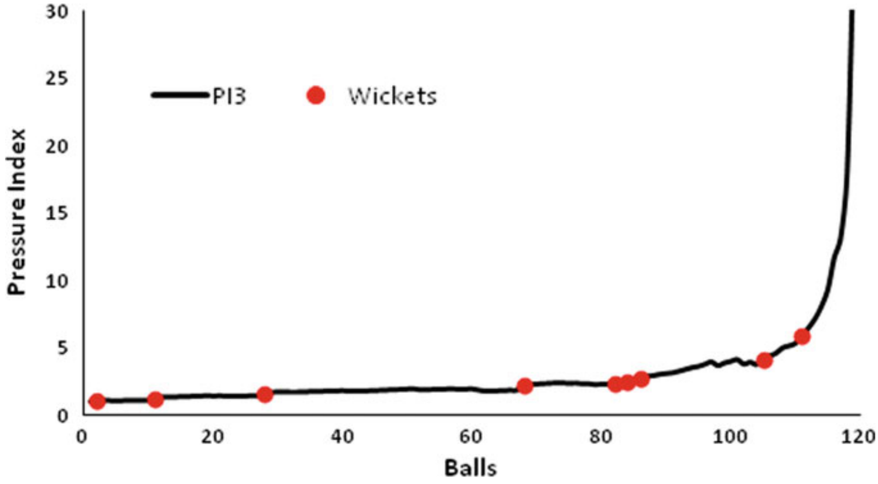

Fig. 7.7 Pressure curve of the first innings of the Twenty20 World Cup Final of 2016

To understand the working of the pressure index, for the team batting in the first innings of a limited over match, let the example of the final match of the Twenty20 World Cup of 2016 be considered. The match was played between England and West Indies. England batting first scored 155 for 9 in 20 overs in that match. The match was played at Eden Gardens, Kolkata, where the average first innings score is 154 with a standard deviation of 29.9. Thus, following (7.24) T_0 for the ground is 184. This means that when England started their innings, their target was 184. But, England lost their first wicket in the very second ball of the match, following which the target of the first innings (T_t) get rescheduled to 169 using the second equation of (7.24). The third, fourth, fifth, sixth, seventh, eighth, and ninth wicket of England fell in the 28th, 68th, 82nd, 84th, 86th, 105th, and 111th balls with the target shifting to 164, 179, 178, 174, 171, 172, and 171, respectively. The other steps in computing the pressure index are the same as the process used in computing the pressure index in the second innings of a limited over match. Figure 7.7 shows the pressure curve of England (First innings) of the Twenty20 World Cup final played at Eden Gardens, Kolkata in 2016.

Thus, the pressure index values obtained for the first innings of a limited overs match can now be utilized to all the applications discussed above.

7.6 Need for Redefining the Pressure Index

The formula for pressure index defined in (7.5) is well accepted. Also, we have found in the different subsections following the introduction of the pressure index formula several applications of the same. The results based on pressure index formula have high predictive power indicating that it is measuring the feature correctly for which it is intended. If Team B wins the match, the pressure index comes down to zero providing an exact destination for the pressure curve. However, the pressure index value is not limited to any fixed quantity if Team B loses the match. Though this lenient nature of the pressure index values (for matches lost by Team B) do not hamper the general working of the index or its predictive power, it gives abnormal values when the pressure index values are extended to measurement of batting or bowling performances. Let us take an example.

This is a Twenty20 match between Bangladesh and Sri Lanka, played at Sher-e-Bangla National Stadium, Mirpur on February 28, 2016. This was the fifth match of Asia Cup of 2016. Bangladesh batting first scored 147 for 7 in their 20 overs and in reply Sri Lanka could score 124 for 8 at the end of 20 overs. Sri Lanka eventually lost the match by 23 runs. The pressure curve for Team B, i.e., Sri Lanka for this case is shown in Fig. 7.8.

It may be observed here that Team B, i.e., Sri Lanka lost the match and so their pressure index values kept creeping up at the end of their innings and even reached values close to 50. The average pressure index value of the entire innings was only 3.2619. Now, when we want to measure the batting and bowling performance based on pressure index, we reach some contrasting result. The results are indicated in Table 7.11.

A look at Table 7.11 reveals the value of performance measure of batsmen and bowlers that we reach based on pressure index, which are difficult to interpret. The computations of adjusted runs are based on (7.6) and (7.7) and the values of BPI utilized (7.9). It can be seen that when Chandimal scored run a ball and reached 37,

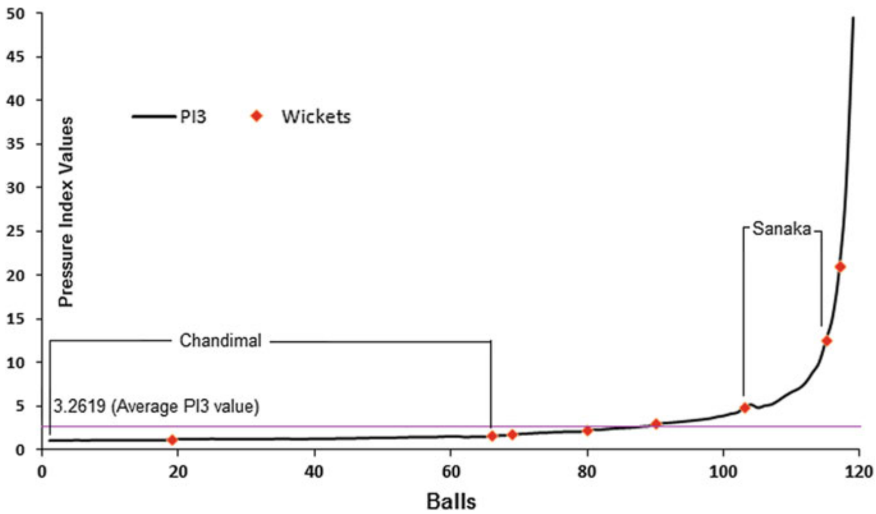

Fig. 7.8 Pressure curve of the Sri Lankan innings for Bangladesh versus Sri Lanka match played on February 28, 2016

Table 7.11 Batting and bowling performance measures of some selected players based on pressure index values for the Bangladesh versus Sri Lanka match

Batting performance measure of the selected batsman (Bangladesh)

Batsman	Runs	Balls	Strike rate	Adjusted runs
Chandimal	37	37	100	23.73
Jayasuriya	26	21	123.8	17.53
Shanaka	14	14	100	17

Bowling performance

Bowler	Overs	Runs conceded	Wickets taken	Economy rate	BPI
Al Amin	4	34	3	8.5	10.17
Mustafizur	4	19	1	4.75	1.47
Shakib	4	21	2	5.25	0.24

the measure adjusted runs scaled it further down to 23.73 but in case of Shanaka scoring at the same rate got his runs scaled further up from 14 to 17. It happened because Shanaka batted when the pressure was more so each runs he scored were scaled higher than the actual runs scored and the reverse happened for Chandimal, as during his batting the pressure was less. This is also the principle of applying the pressure index for measuring batting performance where we need to provide more importance to runs scored when the pressure was more. In this case, as the pressure index rose abruptly because of Team B losing the match by a significant margin, the adjusted runs of batsmen batting at the later part of the innings were boosted more than necessary. On the contrary, batsmen batting at the early part of the innings got their adjusted runs scaled downward to a greater extent. Similarly, deviations can be viewed for bowlers as well. The performance of Al Amin based on BPI is much better than all the other bowlers though he had a poor economy rate compared to others. Actually, Al Amin Hossain bowled the 18th and the 20th over. This was the time when he took all the three wickets and the pressure index values were very high at that stage of the match. The large difference in the performance of Al Amin is because of the abrupt rise in pressure index values rather than his actual performance. Thus, we feel a need to control the abrupt and abnormal raise to the pressure index for matches lost by Team B. This calls for some refinement in the formula. Let us revisit the formula of pressure index (PI_3) once again

$$\mathrm{PI}_3 = \left(\frac{\mathrm{CRRR}}{\mathrm{IRRR}}\right) \times \frac{1}{2}\left[\exp(RU/100) + \exp\left(\sum w_i/11\right)\right]$$

Clearly, the formula is a combination of two parts, viz. $\left(\frac{\mathrm{CRRR}}{\mathrm{IRRR}}\right) = \mathrm{CI}$ and $\frac{1}{2}\left[\exp(RU/100) + \exp(\sum w_i/11)\right] = A(\text{say})$.

Thus, the pressure index is a product of two different indicators—one which deals with how Team B is progressing in terms of runs scored and the other one denotes the rate at which the resources are consumed by Team B mainly in terms of wickets

Table 7.12 Values of CI and A at end of each over in the second innings for Bangladesh versus Sri Lanka match

Over	Value of CI	Value of A	Over	Value of CI	Value of A
1	1.024182	1.019817	12	1.165541	1.599499
2	1.043544	1.04042	13	1.235521	1.640722
3	1.017488	1.062833	14	1.351351	1.794146
4	1.030405	1.165082	15	1.513514	1.947267
5	0.990991	1.188855	16	1.689189	1.997855
6	1.003861	1.215865	17	2.072072	2.052929
7	0.997921	1.244044	18	2.5	2.24992
8	1.036036	1.274696	18.3	3.063063	2.282403
9	1.081081	1.308044	19	4.324324	2.318015
10	1.108108	1.34296	19.3	8.378378	2.518909
11	1.081081	1.462021	19.5	19.45946	2.544792

lost. For the current example, we look at the different values of CI and A at some discrete points of the match and place it in Table 7.12.

A close look at the values of CI and the corresponding values of A expresses the rise in the values of CI compared to that of A at the concluding part of the innings when Sri Lanka was falling reasonably short of their target. Thus, there is a need to control the value of CI in order that the value of PI_3 remains reasonably finite.

Iyengar and Sudarshan (1982), while dealing with the formation of a composite index based on several indicators opines that, large variation in any one of the participating indicators in an index would unduly dominate the contribution of the rest of the indicators and distort the value of the index. Thus, need is felt to introduce a variance stabilization rule such that the variation in CI and A remains approximately equal denying the excessive influence of CI over A in the value of PI_3. Accordingly, we propose the use of a new formula of pressure index (PI). This formula is universally applied to both cases, viz. when Team B wins the match and when the match is lost by Team B. We define,

$$\mathrm{PI} = \left(\frac{\mathrm{CRRR}}{\mathrm{IRRR}}\right)^{\alpha} \times \frac{1}{2}\left[\exp(RU/100) + \exp\left(\sum w_i/11\right)\right] \tag{7.25}$$

where α is chosen using an iterative method in such a way that

$$\mathrm{Var}\left[\left(\frac{\mathrm{CRRR}}{\mathrm{IRRR}}\right)^{\alpha}\right] \cong \mathrm{Var}\left[\frac{1}{2}\left[\exp(RU/100) + \exp\left(\sum w_i/11\right)\right]\right] \tag{7.26}$$

For, the current example, i.e., for Bangladesh versus Sri Lanka Twenty20 match played at Sher-e-Bangla National Stadium, Mirpur on February 28, 2016 we have,

$\mathrm{Var}\left(\frac{\mathrm{CRRR}}{\mathrm{IRRR}}\right) = 4.6879$ and $\mathrm{Var}\left[\frac{1}{2}\left[\exp(RU/100) + \exp\left(\sum w_i/11\right)\right]\right] = 0.196879$ accordingly for $\alpha = 0.47035$ the equivalence defined in (7.26) holds. Plotting PI and PI_3 against the number of balls bowled Fig. 7.9 is obtained. It may be seen from the figure that applying the modification defined above the abrupt rise to PI_3 values at the concluding overs of the innings is restricted largely.

The measures of performance measurement of the crickets are recomputed using the pressure values obtained using (7.25) and are tabulated (cf. Table 7.13). It is seen from the table that though the ranks of the players have not changed, the values of performance measurement are now more reasonable than what it was earlier.

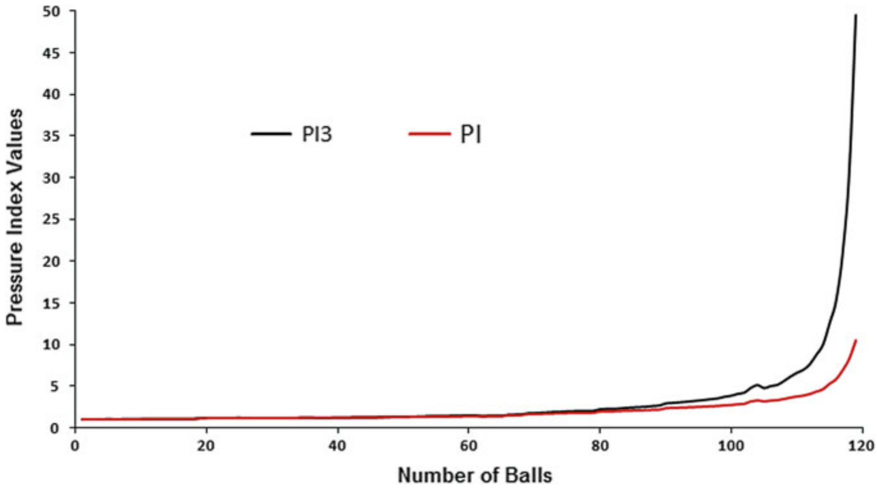

Fig. 7.9 Pressure index values obtained using the modified formula is plotted along with the original pressure index values

Table 7.13 Measures of batting and bowling performance using the modified formula of pressure index

Batting performance measure of the selected batsman (Bangladesh)					
Batsman	Runs	Balls	Strike rate	Adj. runs	Adj. runs (new)
Chandimal	37	37	100	23.73	29.71
Jayasuriya	26	21	123.8	17.53	21.8
Shanaka	14	14	100	17	17

Bowling performance of some Sri Lankan bowlers						
Bowler	Overs	Runs conc.	Wkts taken	Eco. rate	BPI	BPI (New)
Al Amin	4	34	3	8.5	10.17	1.6
Mustafizur	4	19	1	4.75	1.47	0.46
Shakib	4	21	2	5.25	0.24	0.14

7.7 Conclusion

Thus, starting with a simple formula of quantifying the pressure experienced by a team batting in the second innings of a limited overs cricket match, the concept is now extended to different dimensions, performance measures of teams and individuals, match outcome prediction, comparing run chases, and so on. Considering the huge area of applicability of the index, we call it magical. The pressure index also opens up another area of research for people interested in sports analytics—it can be extended to other sports as well. The pressure index can also be modified in a number of ways. The first issue can be to redefine the pressure index formula so that an upper bound can be attained or to restrict the rise of the pressure curve when the team batting second falls short of the target. We believe that the pressure index has in its store several other hidden dimensions which are left for the readers to explore.

Appendix 7.1: Wicket Weights of Different Batting Positions as in Lemmer (2005)

Batting position (i)	Wicket weight w_i
1	1.30
2	1.35
3	1.40
4	1.45
5	1.38
6	1.18
7	0.98
8	0.79
9	0.59
10	0.39
11	0.19

Appendix 7.2: Binary Logistic Regression

In some cases, it so happens that subjects under consideration can be classified into one of the two groups based on their performance on a set of explanatory variables ($\boldsymbol{X}$, say). In such a case, we are rather interested to compute the probability that a subject belongs to one of the two categories given the values of the explanatory variables. Let us call the proportion as p. Estimating the value of p based on a set of explanatory variables ($\boldsymbol{X}$) or building models for estimating the value of p given values of the explanatory variables $\boldsymbol{X}$ is called *logistic regression*.

Binary type data, i.e., data of the form Yes and No comes from a random sample that has a binomial distribution with probability of success p. Often the value of the probability of success is unknown and is to be estimated using a model given the data. Since the binary type data cannot follow a normal distribution, an essential condition for using ordinary regression, so one needs a new type of regression model to do this job called the *logistic regression.*

A *logistic regression* model ultimately gives you an estimate of p, the probability that a particular outcome will occur in a binary set up. The estimate is based on information from one or more explanatory variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_k)$.

Logistic Regression can be used whenever an individual is to be classified into one of two populations. However, if the population into which an individual to be classified is more than two then we reach to the multinomial logistic regression.

Assumptions of Binary Logistic Regression

- Since it assumes that $P(Y = 1)$ is the probability of the event occurring, so it is essential that the dependent variable (also referred to as the response variable) is being coded accordingly. The factor level 1 (i.e. $Y = 1$) of the dependent variable should represent the desired outcome.
- The independent variables (also referred to as the explanatory variables) should be independent of each other.
- It requires quite a large number of samples because the coefficients are estimated by MLE method which is less powerful than the OLS method.

Derivation of the Regression Equation

Here, we try to understand the binomial case:

Let Y be a binary variable having two outcomes 0 and 1, such that,

$$P(Y = 1) + P(Y = 0) = 1$$

We are in search of a relation of the form

$$P(Y = 1) = a + b'\mathbf{X}$$

Let us now introduce the example concept of odds in favor. The odds in favor of an event $Y = 1$ is the ratio between $P(Y = 1)$: $P(Y \neq 1)$. Thus,

$$\text{Odds}(Y = 1) = \frac{P(Y = 1)}{P(Y \neq 1)} = \frac{P(Y = 1)}{P(Y = 0)} = \frac{P(Y = 1)}{1 - P(Y = 1)}$$

Unlike $P(Y = 1)$, odds ($Y = 1$) does not have an upper limit, however like probability odds ratio cannot be negative. However, with log transformation the value of the variable ranges from $-\infty$ to ∞. Thus,

$\log_e$ [Odds ($Y = 1$)] $= \log_e\left[\frac{P(Y=1)}{1-P(Y=1)}\right]$ is called as logit of Y.

So, logit $(Y) = \log_e\left[\frac{P(Y=1)}{1-P(Y=1)}\right]$.

Let us assume that the logit is a function of some independent variables X_1, X_2, …, X_k etc. which may be discrete or continuous, i.e.,

logit $(Y) = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k$

$\Rightarrow$ log [odds($Y = 1$)] $= a + b_1X_1 + b_2X_2 + \cdots + b_kX_k$

$\Rightarrow$ odds($Y = 1$) $= e^{a+b_1X_1+b_2X_2+\cdots+b_kX_k}$

$\Rightarrow \frac{P(Y=1)}{1-P(Y=1)} = e^{a+b_1X_1+b_2X_2+\cdots+b_kX_k}$

$\Rightarrow P(Y = 1) = e^{a+b_1X_1+b_2X_2+\cdots+b_kX_k} - P(Y = 1)e^{a+b_1X_1+b_2X_2+\cdots+b_kX_k}$

$\Rightarrow (1 + e^{a+b_1X_1+b_2X_2+\cdots+b_kX_k})P(Y = 1) = e^{a+b_1X_1+b_2X_2+\cdots+b_kX_k}$

$\Rightarrow P(Y = 1) = \frac{e^{a+b_1X_1+b_2X_2+\cdots+b_kX_k}}{1+e^{a+b_1X_1+b_2X_2+\cdots+b_kX_k}}$

Or more precisely we can write $P(Y = 1)$ as $P(Y = 1|X_1, X_2, \ldots, X_k)$ writing, $a + b_1X_1 + b_2X_2 + \cdots + b_kX_k = z$, we have,

$$P(Y) = \frac{e^z}{1+e^z} = \frac{1}{\frac{1+e^z}{e^z}} = \frac{1}{1+e^{-z}} = \frac{1}{1+\exp\{-(a+b_1X_1+b_2X_2+\cdots+b_kX_k)\}}$$

Thus, the logit model is a form of the logistic curve and hence the name.

Understanding the coefficients of the model

We have,

$$\text{logit}(Y) = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k$$

Now,

$$\begin{aligned} a &= \text{log odds for individual for all } \ X_i = 0 \\ &= \text{the odds that would have resulted for a logistic model} \\ &\quad \text{without any } X \text{ at all.} \end{aligned}$$

Now, in order to understand the b's let us consider the change in logit (Y) when one of the X's varies when all others are kept fixed.

Say: Fix $X_2, X_3, \ldots, X_k$ to 0 and vary X_1 from 0 to 1.

Let, logit $P_1(Y)$ when $X_1 = 1$ and

logit $P_0(Y)$ when $X_1 = 0$, while all others are fixed at 0 as mentioned earlier. Thus,

$$\text{logit } P_1(Y) - \text{logit } P_0(Y) = a + b_1X_1 - a = b_1X_1 = b_1 \ \text{ as } \ X_1 = 1$$

Thus, b_1 represents the change in the log odds i.e. logit that would result from a one unit change in the variable X_1 when other variables are fixed.

References

Allsopp, P. E. (2005). Measuring team performance and modeling the home advantage effect in cricket. Unpublished Ph. D dissertation.

Allsopp, P. E., & Clarke, S. R. (2004). Rating teams and analysing outcomes in one-day and test cricket. *Journal of the Royal Statistical Society*, *Series A (Statistics in Society), 167*(4), 657–667.

Bailey, M., & Clarke, S. R. (2006). Predicting the match outcome in one day international cricket matches, while the game is in progress. *Journal of Sports Science and Medicine, 5,* 480–487.

Bandulasiri, A. (2008). Predicting the winner in one day international cricket. *Journal of Mathematical Sciences & Mathematics Education, 3*(1), 1–12.

Barnett, T. & Clarke, S. R. (2002). Using Microsoft Excel to model tennis match. In G. Cohen & T. Langtry (Eds.), *Proceedings of the 6th conference on Mathematics and Sports* (pp. 63–68).

Barr, G. D. I., Holdsworth, C. G., & Kantor, B. S. (2008). General evaluating performances at the 2007 Cricket World Cup. *South African Statistical Journal, 42,* 125–142.

Beaudoin, D., & Swartz, T. B. (2003). The best batsmen and bowlers in one-day cricket. *South African Statistical Journal, 37,* 203–222.

Bhattacharjee, D., & Lemmer, H. H. (2016). Quantifying the pressure on the teams batting or bowling in the second innings of limited overs cricket matches. *International Journal of Sport Science and Coaching, 12*(1), 683–692.

Bhattacharya, R., Gill, P. S., & Swartz, T. B. (2011). Duckworth-Lewis and Twenty20 Cricket. *Journal of the Operational Research Society, 62,* 1951–1957.

Brooks, R. D., Faff, R. W., & Sokulsky, D. (2002). An ordered response model of test cricket performance. *Applied Economics, 34,* 2353–2365.

Carter, M., & Guthrie, G. (2002). Cricket interruptions, fairness and incentive in interrupted cricket matches. *Journal of the Operations Research Society, 55*(8), 822–829.

Clarke, S. R. (1988). Dynamic programming in one-day cricket-optimal scoring rates. *Journal of the Operational Research Society, 39,* 331–337.

Clowes, S., Cohen, G., & Tomljanovic, L. (2002). Dynamic evaluation of conditional probabilities of winning a tennis match. In *Proceedings of the Sixth Australian Conference* on *Mathematics and Computers in Sport, 1–3 July 2002, at Bond University, Gold Coast, Queensland 4229, Australia* (pp. 112–118).

Cui, T., Li, J., Woodward, J. R., & Parkes, A. J. (2013). An ensemble based genetic programming system to predict English football premier league games. In *Proceedings of the 2013 IEEE Conference On Evolving and Adaptive Intelligent Systems, Eais 2013–2013 IEE Symposium Series On Computational Intelligence* (pp. 138–143). https://doi.org/10.1109/eais.2013.6604116.

Damodaran, U. (2006). Stochastic dominance and analysis of ODI batting performance: the Indian cricket team, 1989–2005. *Journal of Sports Science and Medicine, 5,* 503–508.

Daud, Ali & Muhammad, Faqir. (2013). Ranking Cricket Teams through Runs and Wickets, p. 8210. https://doi.org/10.1007/978-3-319-02750-0_16.

de Silva, R. (2011) A method to calculate the target score in rain interrupted limited over cricket matches. In *International Conference on Mathematics and Science (ICOMSc)* (pp. 1–8), Surabaya, Indonesia.

de Silva, B. M., Pond, G. R., & Swartz, T. B. (2001). Estimation of magnitude of victory in one-day cricket. *Australian and New Zealand Journal of Statistics, 43*(3), 259–268.

Dey, P. K., Ghosh, D. N., & Mondal, A. C. (2015). IPL team performance analysis: a multi-criteria group decision approach in fuzzy environment. *International Journal of Information Technology and Computer Science, 8,* 8–15.

Douglas, J. M., & Tam, N. (2010). Analysis of team performances at the ICC world twenty20 cup 2009. *International Journal of Performance Analysis in Sport, 10*(1), 45–53.

Duckworth, F. C., & Lewis, A. J. (1998). A fair method of resetting the target in interrupted one-day cricket matches. *Journal of the Operational Research Society, 49*(3), 220–227.

Elderton, W. P., & Elderton, E. M. (1909). *Primer of statistics*. London: Black.

Farazdaghi, H., & Harris, P. M. (1968). Plant competition and crop yield. *Nature, 217,* 289–290.

Graybill, F. A., & Iyer, H. K. (1994). *Regression analysis: Concepts and applications*. Belmont, Calif: Duxbury Press.

Heazlewood, T. (2006). Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic games. *Journal of Sports Science and Medicine, 5,* 541–547.

Huang, K. Y., & Chang, W. L. (2010). A neural network method for prediction of 2006 world cup football game. In *The 2010 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8).

Hucaljuk, J., & Rakipovic, A. (2011). Predicting football scores using machine learning techniques. In: MIPRO, 2011, *Proceedings of the 34th International Convention. IEEE, 2011* (pp. 1623–1627).

Iyengar, N. S., & Sudarshan, P. (1982). A method of classifying regions from multivariate data. *Economic and Political Weekly. 18*, 2048–2052.

James, B. (1980). *The Bill James abstract*. Self-published.

Jayadevan, V. (2002). A new method for the computation of target scores in interrupted limited-over cricket matches. *Current Science, 83*(5), 577–586.

Joseph, A., Fenton, N. E., & Neil, M. (2006). Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems, 19*(7), 544–553.

Lata, K., & Gupta, P. (2016). Forecasting English premier league match results. *International Journal of Modern Trends in Engineering and Research, 03*(02), 261–268.

Lemmer, H. H. (2005). A method for the comparison of the bowling performances of bowlers in a match or series of matches. *South African Journal for Research in Sport, Physical Education and Recreation, 27*(1), 91–103.

Lemmer, H. H. (2011). Predicting the names of the best teams after the knock-out phase of a cricket series. European journal of sport science, 14(sup1), S191–S196.

Lemmer, H. H. (2015). A method to measure choking in cricket. *International Journal of Sports Science and Coaching, 10*(2 + 3), 461–477.

Noubary, R. (2005). A procedure for prediction of sports records. *Journal of Quantitative Analysis in Sports, 1*(1), 1–12. Retrieved 27 May. 2017, from https://doi.org/10.2202/1559-0410.1001.

O'Riley, B. J., & Ovens, M. (2006). Impress your friends and predict the final score: An analysis of the psychic ability of four target resetting methods used in one-day international cricket. *Journal of Sports Science and Medicine, 5*(4), 488–494.

Cricinfo: Sri Lanka v. Pakistan, T20I Series. (2015). Retrieved from http://www.espncricinfo.com/sri-lanka-v-pakistan-2015/engine/match/860281.html.

Pollard, R., Benjamin, B., & Reep, C. (1977). Sport and the negative binomial distribution. In S. P. Ladany & R. E. Machol (Eds.), *Optimal strategies in sports* (pp. 188–195). New York: North-Holland.

Preston, I., & Thomas, J. (2000). Batting strategy in limited overs cricket. *The Statistician, 49*(1), 95–106.

Preston, I., & Thomas, J. (2002). Rain rules for limited overs cricket and probabilities of victory. *Journal of the Royal Statistical Society Series D, 51*(2), 189–202.

Ramakrishnan, V., Sethuraman, K., & Parameshwaran, R. (2010). Target score prediction in the game of cricket. Retrieved from https://people.ucsc.edu/~praman1/static/pub/ML_project_CS7641_report.pdf.

Rosenfeld, J. W., Fisher, J. I., Adler, D., & Morris, C. (2010). Predicting overtime with the pythagorean formula. *Journal of Quantitative Analysis in Sports, 6*(2), Article 1.

Scarf, P., & Shi, X. (2005). Modelling match outcomes and decision support for setting a final innings target in test cricket. *IMA Journal of Management Mathematics, 16,* 161–178.

Scarf, P., Shi, X., & Akhtar, S. (2011). On the distribution of runs scored and batting strategy in Test cricket. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 174*(2), 471–497.

Shah, P., & Shah, M. (2014). Pressure index in Cricket. *IOSR Journal of Sports and Physical Education, 1*(5), 9–11.

Shinozaki, K., & Kira, T. (1956). Intraspecific competition among higher plants. VII logistic theory of the C-D effect. *Journal of the Institute of Polytechnics, Osaka City University, D7,* 35–72.

Somboonphokkaphan, A., Phimoltares, S., & Lursinsap, C. (2009). Tennis winner prediction based on time-series history with neural modeling, In *Proceedings of the International MultiConference of Engineers and Computer Scientists 2009, IMECS* (Vol. I, pp. 127–132), March 18–20, 2009, Hong Kong.

Tan, A., & Zhang, D. (2001). Distribution of batting scores and opening wicket partnerships in cricket. *Mathematical Spectrum, 34*(1), 13–17.

Valero, J., & Swartz, T. B. (2012). An investigation of synergy between batsmen in opening partnerships. *Sri Lankan Journal of Applied Statistics, 13,* 87–98.

Wickramasinghe, I. P. (2014). Predicting the performance of batsmen in test cricket. *Journal of Human Sport and Exercise, 9(4)*, 744–751.

Yang, T. Y., & Swartz, T. (2004). A two-stage Bayesian model for predicting winners in major league baseball. *Journal of Data Science, 2*(1), 61–73.

Chapter 8
Decision Making in Cricket: The Optimum Team Selection

8.1 Introduction

Whether in a general conversation between fans or in a public debate, the process of team selection in many team sports finds itself under scrutiny (Brettenny, 2010). Fans often believe that the process of team selection can be done by them in a more meaningful way than those, who are actually responsible for the job. In a team sports, different players have difference in their expertise and so selecting a best possible team shall comprise of choosing the finest of the available players under different expertise. Like any other team game, in cricket too, when the selectors sit for choosing a team (let us call it a squad) they have to pick representatives from all different departments of the game. This is definitely an uphill task for the selectors and probably more difficult and different from several other team games. A balanced cricket squad shall have opening batsmen, middle-order batsmen, spinners, pace bowlers, wicket keepers and all-rounders. Often the format of the game and the country where the series shall be played determines the team composition when the squad is to be selected for an overseas tour. Though in a cricket match, irrespective of the format of the game, eleven players shall take the field but a squad generally comprises of 15 players. Such a squad is generally selected by a group of selectors nominated by the cricket board. The selectors are mostly veteran cricketers who have sufficient expertise and experience for accomplishing this duty. Once the squad is selected—the team management that generally comprises of the captain, vice-captain, and the coach decides on the exact eleven players who shall play a given match during the tour. Though most of the cricket boards rely on the selection committee for picking up the squad, these days several analysts support them with necessary statistics for such decision making.

There are different authors who had addressed this issue of team (or squad) selection mathematically. This is mostly a problem of optimization—optimization of the available resources into a team formation. The processes of team selection using mathematical modeling by different authors have something in common. The first step is to decide about the statistic that can measure the performance of the cricketers

H. Saikia et al., *Cricket Performance Management*, Indian Statistical Institute Series,
https://doi.org/10.1007/978-981-15-1354-1_8

in the consideration set and then to find out an optimum tool that can provide the best mix of players in the squad. Kamble, Rao, Kale, and Samant (2011) presented a selection procedure where they used an analytical hierarchical process using which they choose a subset of players from a universal set of cricketers comprising of batsmen, bowlers, all-rounders, and wicket keepers. Two other works on the same issue are that of Lemmer (2011) and Ahmed, Jindal, and Deb (2012). Lemmer (2011) used integer programming to reach the solution and Ahmed et al. (2012) used evolutionary multi-objective optimization to choose the cricket team. Barr and Kantor (2004) used the concept of portfolio management of share market to determine the batsmen who are supposed to be more suitable for the selection in an world XI based on data from 2003 World Cup. Gerber and Sharp (2006) proposed an integer programming technique in order to select a limited over squad of 15 players instead of selecting just the playing XI. For this purpose, they collected data on 32 South African cricketers and selected a one-day squad. The same idea, but in a larger frame was used by Lourens (2009) and he selected an optimal Twenty20 South African cricket side based on performance statistics of a host of players who participated in the domestic Pro20 cricket tournament of South Africa. Brettenny (2010), using integer programming, selected players for a fantasy league cricket team, under certain prespecified budgetary constraints but with a progressive approach. His optimal team was recalculated at each stage of the tournament, considering the performance of available cricketers till the previous match. Most of the authors who worked on team selection used the binary integer programming tool for the purpose of the selection, but they used different tools for measuring cricketers' performance. Some authors used common measures like batting average, strike rate, etc. for quantifying performance of cricketers, while others tried combination of the common measures to a refined statistic to evaluate players' performance. Lourens (2009) and Brettenny (2010) combined/compared different refined measures that were used for performance measurement of cricketers. Both the works also showed how optimum teams varied when the statistic used for performance measures are changed. The purpose of applying optimization models in selection of a balanced team is explained by Boon and Sierksma (2003). The paper is related to team selection in soccer and volleyball, but it is an excellent example of the application of transportation problem for the purpose of team selection.

In this chapter, we perform two exercises of selecting a squad of 15 players. The first exercise is relatively simple and straight forward and do not use any mathematical tool for optimization. In it, a simple index is developed that maps the performance of a cricketer into a real number in the interval [0, 1]. The higher the value, better is the performance. Then, the squad is selected based on some predetermined number of batsmen, bowlers, all-rounders, wicket keepers, etc. Since the number is predetermined so the task is relatively easier and the need of using binary integer programming (a special case of mathematical programming) is not necessary. The second exercise is relatively more complicated, but is definitely closer to the real situation than the first one. In this method, the need for defining a performance measure as well as an optimization technique under several constraints is discussed. The second method considers the fact that the captain of the cricket team gets an obvious selection in the team and the expertise of the captain influence the selection of the remaining players of the team. The details of both the techniques are placed in the subsequent sections.

8.2 National Team Selection from the Cricketers Participating in Domestic Tournaments

Here, we take up a simple example to explain how we can define a single index for measuring the performance of cricketers and then extending it to select a squad. The objective of the exercise is to keep the methodology simple and with minimum use of advanced statistical/mathematical tool or that of optimization. This can be handled even by the undergraduate students and can be an interesting summer project. Some of these performance measuring tools are similar to those discussed in Chap. 3, but for maintaining continuity they are revisited briefly.

Let us consider a domestic tournament, where cricketers of different expertise have participated. The performance measure of the i^{th} player in the tournament is given by,

$$S_i = S_{i1} + \delta_i \tag{8.1}$$

where

$$\delta_i = \begin{cases} S_{i2}^{a_i} + S_{i3}^{1-a_i} - 1, & \text{if } i^{\text{th}} \text{ player is either a bowler or wicket keeper} \\ 0, & \text{if } i^{\text{th}} \text{ player is neither a bowler nor wicket keeeper} \end{cases}$$

a_i is an indicator variable with,

$$a_i = \begin{cases} 1, \text{ if } i^{\text{th}} \text{ player is a bowler} \\ 0, \text{ if } i^{\text{th}} \text{ player is a wicket keeeper} \end{cases}$$

and

S_{i1} = Performance score for batting
S_{i2} = Performance score for bowling
S_{i3} = Performance score for wicket keeping

8.2.1 Different Factors Considered for Batting

To measure the batting performance of a cricketer in Twenty20 cricket, a number of factors are considered. These factors are number of innings in which the player had actually batted, batting average, strike rate of the batsman, and average percentage contribution to the team total. The values of these factors for each player are normalized, and then, weights of these factors are calculated to stabilize their variance. All the normalized scores for the considered factors are multiplied by their corresponding weights and then added together to get S_{i1} (i.e., the performance score for batting for the i^{th} batsman).

8.2.2 Different Factors Considered for Bowling

The different factors that are considered while computing the performance score of bowlers are number of innings in which the player was bestowed with the responsibility of bowling, bowling average, economy rate, and strike rate of the bowler. The values of these factors, for each player, are normalized, and then, weights of these factors are calculated to stabilize their variance. As earlier, all the normalized scores for the considered factors are multiplied by their corresponding weights and then added together to get S_{i2} (i.e., the performance score for the i^{th}bowler).

8.2.3 Different Factors Considered for Wicket Keeping

To measure the performance of a wicket keeper, the different factors considered are number of matches played as wicket keeper, number of catches taken per match, number of stumping per match, and number of bye runs conceded per match. Here, the phrase ‘per match’ means the number of matches when the player kept the wickets for his team. This is because some of the teams have more than one player in their playing eleven who are capable of wicket keeping.

8.2.4 Normalization

The developed performance measure in this paper is a linear combination of traditional performance measures under the skills of batting, bowling, and wicket keeping. Therefore, to overcome this limitation of different measures having different units the use of normalization is essential in this model. Normalization aids to eliminate the unit of measurement and variability effect of all the traditional performance measures. Based on normalization, the traditional performance measures under the skills of batting, bowling, and wicket keeping come within a similar range from 0 to 1. Since normalization makes the measures unit free, so they can be aggregated through addition.

Let X_{ijk} be the observed values of the i^{th} player for the j^{th} factor (i.e., batting average, strike rate, etc.) of the k^{th} skill (i.e., batting, bowling, and wicket keeping). Out of the different factors considered for performance measurement, some are having positive dimension like batting average, batting strike rate, number of stumping, etc. as they are directly related to the skill of the player. While some of the factors like economy rate, number of bye runs conceded, etc. have negative dimension as they are negatively related to the skill of the player. Now if the factor represents positive dimension, then it is normalized as

$$Y_{ijk} = \frac{X_{ijk} - \min\limits_{i}\left(X_{ijk}\right)}{\max\limits_{i}\left(X_{ijk}\right) - \min\limits_{i}\left(X_{ijk}\right)} \tag{8.2}$$

and if the factor represents negative dimension, then it is normalized as

$$Y_{ijk} = \frac{\max\limits_{i}\left(X_{ijk}\right) - X_{ijk}}{\max\limits_{i}\left(X_{ijk}\right) - \min\limits_{i}\left(X_{ijk}\right)} \tag{8.3}$$

8.2.5 Determination of Weights

The next step would be determination of weights. Basically, simple averages provide equal importance to each of the variables. When variables are weighted to a composite measure, the relative importance of the variables are considered. Other than the arbitrary weighting techniques, there are different statistical methods of weighting as well. Iyenger and Sudarshan (1982) assumed that the weights vary inversely as the variation in the respective variables. This conception has been thoroughly applied in this study to determine the weights of different factors that are associated with the various skills of the cricketers.

Let Y_{ijk} be the normalized value of the i^{th} players for the j^{th} factors of the k^{th} skills where i (=1, 2, …, n) represents players; j (=1, 2, 3, 4) for the four different factors considered under each of the k skills (e.g., under batting skill different factors are number of innings, strike rate, batting average, average percentage of contribution to the team total); $k = 1, 2, 3$ represents the different skills, viz. batting ($k = 1$), bowling ($k = 2$), and wicket keeping ($k = 3$). If w_{jk} represents the weight of the j^{th} factor under the k^{th} skill, then it is calculated as

$$w_{jk} = \frac{C_k}{\sqrt{\operatorname*{Var}\limits_{i}(Y_{ijk})}} \quad j = 1, 2, 3, 4 \quad \text{and} \quad k = 1, 2, 3, \tag{8.4}$$

where $\sum_{j=1}^{4} w_{jk} = 1$ for all k.

C_k is a normalizing constant that follows

$$C_k = \left[\sum_{j=1}^{4} \frac{1}{\sqrt{\underset{i}{\mathrm{Var}}(Y_{ijk})}} \right]^{-1}$$

The choice of the weights in this manner would ensure that the large variation in any one of the factor would not unduly dominate the contribution of the rest of the factors (Iyenger & Sudarshan, 1982).

8.2.6 Computation of Performance Score

The performance scores of S_{i1}, S_{i2}, and S_{i3} for batting, bowling, and wicket keeping, respectively, of the cricketers in the consideration set are computed as follows.

The batting performance ($k = 1$) of the i^{th} player is calculated by

$$S_{i1} = \sum_{j=1}^{4} w_{j1} Y_{ij1}; \quad j = 1, 2, 3, 4 \tag{8.5}$$

The bowling performance ($k = 2$) of the i^{th} player is calculated by

$$S_{i2} = \sum_{j=1}^{4} w_{j2} Y_{ij2}; \quad j = 1, 2, 3, 4 \tag{8.6}$$

The performance ($k = 3$) of the i^{th} wicket keeper is calculated by

$$S_{i3} = \sum_{j=1}^{4} w_{j3} Y_{ij3}; \quad j = 1, 2, 3, 4 \tag{8.7}$$

On obtaining the values of S_{i1}, S_{i2}, and S_{i3}, the performance measure S_i of the i^{th} player is computed using Eq. (8.1). The performance measures of all the players are computed and then converted into corresponding performance index (P_i). The performance index of the i^{th} player is denoted by P_i and is given by,

$$P_i = \frac{S_i}{\underset{i}{\max}(S_i)} \tag{8.8}$$

The performance index for each player is a number lying between zero and unity (i.e., $0 < P_i \leq 1$). Higher value of the performance index, better is the performance of the player.

Now the values of P_i corresponding to the different players are used to rank them. This shall be done separately for the players of different expertise like openers, middle-order batsmen, fast bowlers, spin bowlers, wicket keepers and all-rounders. Now, if the requirement of the squad is well defined, like three openers, four middle-order batsmen, three fast bowlers, three spinners, one wicket keeper, and one all-rounder then one can look up into the expertise-wise P_i table and can choose the best performers under the different expertise categories.

8.3 Data and Team Selection After a Domestic Tournament

This section explains the working of the methodology explained in Sect. 8.2. For our purpose, we use the data of Bangladesh Premier League (BPL) of 2012. Soon after the said season of the BPL, the Bangladesh team was selected for the Twenty20 World Cup. The selectors appointed by the Bangladesh Cricket Board (BCB) are to choose a squad of 15 players for the World Cup. Since the BPL just preceded the world cup, it is expected that the performance of the Bangladeshi cricketers in BPL shall decide their selection in the World Cup squad. Any given team has to select 15 best players for ICC Twenty20 World Cup. Thus, once the values of performance score (P_i) for the entire set of available players is calculated, the next step is to select the optimal 15 players for the Twenty20 World Cup. The optimum team selected by us from the available players has equal number of players under each expertise as in the team chosen by the selectors of BCB. The BCB's team for Twenty20 World Cup had two openers, three specialist batsmen, two all-rounders, five pace bowlers, two spin bowlers, and one wicket keeper.

The data related to the performance of the players in BPL is considered as the basis of selection of the players for the optimum team. The information is collected from the website www.espncricinfo.com. To measure the performance of players, it is necessary that players' statistics for large number of games should be considered. The expectations regarding level of performance cannot be gauged fairly from only one match, and therefore, individual performances across a series of matches are required to provide a suitable frame of reference (Bracewell & Ruggiero, 2009). Thus, some selection criterion needs to be set up while considering the players for performance measurement. Only Bangladeshi cricketers who participated in the BPL of 2012 and satisfy the following criteria are considered for selection.

For batsmen	(a) Batted in at least three innings
	(b) Faced at least 60 balls
	(c) Batting average greater than equal to 10
For bowlers	(a) Bowled in at least three innings
	(b) Delivered at least 60 balls
	(c) Dismissed at least 3 batsman
For wicket keepers	(a) At least three matches played as wicket keeper
	(b) Dismissed at least 3 batsmen
	(c) Satisfied all three conditions of batsmen selection

Table 8.1 Performance related statistics and index values for middle-order and opening batsman

Player	Expertise	Team	Inns	Average	Strike rate	Avg. % Contri.	Performance index
Junaid Siddique	Opener	DR	11	26.44	147.82	11.54	0.3139
Imrul Kayes	Opener	SR	7	14.57	114.6	7.18	0.1483
Nazimuddin	Opener	DG	4	15.25	91.04	10.05	0.0879
Md. Ashraful	Batsman	DG	11	28.66	112.66	17.32	0.3006
Mahmudullah	Batsman	CK	9	25.71	118.42	14.00	0.2529
Mominul Haque	Batsman	BB	7	20.33	135.55	16.51	0.2517
Ziaur Rahman	Batsman	CK	7	25.8	137.23	13.00	0.2472
Alok Kapali	Batsman	SR	8	20.66	121.56	11.74	0.2167
Naeem Islam	Batsman	SR	6	25.25	114.77	12.42	0.1933
Sabbir Rahman	Batsman	DR	8	17.16	122.61	8.73	0.1911
Nazmul Hossain	Batsman	KRB	7	26.66	112.67	7.40	0.1778
Shahriar Nafees	Batsman	BB	3	19	67.85	8.84	0.0402

For all-rounders The conditions for both the batsmen and bowlers should be satisfied

There are 33 Bangladeshi players from BPL 2012, who had satisfied the above-mentioned criteria. The list includes 3 openers, 9 batsmen, 5 fast bowlers, 10 spinners, 2 all-rounders, and 4 wicket keepers. The performance related statistics of the players in the first season of BPL is shown in Tables 8.1, 8.2, 8.3 and 8.4.

The Twenty20 World Cup of 2012 just followed the BPL. So it was obvious that the selectors of Bangladesh must have considered the performance of the Bangladeshi cricketers in BPL 2012 for selecting the world cup squad of Bangladesh. As the world cup that followed and the BPL were of the same format, i.e., Twenty20, so the league must have provided the necessary inputs that the selectors look at, for selecting a balanced squad. The selectors named a fifteen member squad with one wicket keeper, two openers, three batsman, two spinners, five fast bowlers, and two all-rounders. Accordingly, from our Tables 8.1, 8.2, 8.3 and 8.4 we select the best players under different expertise matching the numbers as selected in the actual squad. In Tables 8.1, 8.2, 8.3 and 8.4, the selected players are differentiated from others using bold font. The actual selected squad and the squad as per our exercise (let us call it the Optimum Squad) are jointly placed in Table 8.5.

The team selected by the board has twelve players in common out of fifteen with the optimum team. Opener T. Iqbal was not considered for selection in the optimum squad as he played only two matches in BPL. He was suffering from groin injury.

Table 8.2 Performance related statistics and index values for spinners and fast bowlers

Player	Type	Team	Inn.	Bowl_Avg	Bowl_SR	Econ_rate	Perf. index
Elias Sunny	Spin	DG	11	14.88	13.29	6.72	0.6793
Abdur Razzak	Spin	KRB	11	19.36	19.64	5.92	0.6659
Enamul Haque Jr	Spin	CK	9	14.46	13.85	6.27	0.6481
Arafat Sunny	Spin	CK	10	19.36	19.00	6.11	0.6386
Suhrawadi Shuvo	Spin	BB	11	26.60	21.00	7.60	0.5772
Saqlain Sajib	Spin	DR	11	26.50	19.80	8.03	0.5674
Nazmul Islam	Spin	BB	9	62.25	51.00	7.32	0.4256
Monir Hossain	Spin	DR	6	38.00	36.00	6.33	0.4053
Mosharraf Hossain	Spin	DG	7	44.67	42.00	6.38	0.3795
Nabil Samad	Spin	SR	10	47.00	36.00	7.83	0.2973
Mashrafe Mortaza	Fast	DG	11	27.40	24.00	6.85	0.5883
Shafiul Islam	Fast	KRB	7	30.60	21.60	8.50	0.4375
Farhad Reza	Fast	CK	5	46.00	34.00	8.12	0.3022
Al-Amin Hossain	Fast	BB	6	49.67	30.00	9.93	0.2665
Abdul Hasan	Fast	SR	5	52.00	26.00	12.00	0.1799

However, considering the form of Iqbal and his service to Bangladesh cricket his selection was obvious. The first difference between the teams seems to be among a specialist batsman where Mominul Haque is included in the optimum team instead of Jahurul Islam as in the actual squad. The other difference is seen in the fifth fast bowler. Thus, overall it may be concluded that the selectors of BCB has given a very balanced team to the Bangladesh captain.

Table 8.3 Performance related statistics and index values for wicket keepers

Player	Team	Innings	Bat avg.	Strike rate	Avg % contri	Catch	Stumping	Bye runs	Perf. index
Mushfiqur Rahim	DR	9	39	125.8	17.64	4	2	5	0.4376
Anamul Haque	DG	10	25.14	107.31	15.12	2	1	1	0.3491
Jahurul Islam	CK	10	14.4	118.03	11.00	4	1	6	0.2998
Mithun Ali	BB	8	15.28	109.18	8.64	2	0	0	0.2255

Table 8.4 Performance related statistics and index values for all-rounders

Player	Team	Inn.	Bat avg.	Strike rate	Avg % contri	Bowl avg	Bowl SR	Econ. rate	Perf. index
Shakib Al Hasan	KRB	11	40	140	16.41	19.93	16.80	7.12	1
Nasir Hossain	KRB	7	30.33	119.73	16.80	47.00	36.00	7.83	0.6708

Team Names: *BB* Barisal Burners, *CK* Chittagong Kings, *DR* Duronto Rajshahi, *DG* Dhaka Gladiators, *KRB* Khulna Royal Bengals, *SR* Sylhet Royals

Table 8.5 Optimal squad and actual Bangladesh team for ICC Twenty20 World Cup 2012

Actual squad	Optimum squad
Wicket keeper	
M. Rahim	M. Rahim
Openers	
J. Siddique	J. Siddique
T. Iqbal	I. Kayes
Batsmen	
Md. Ashraful	Md. Ashraful
Mahmudullah	Mahmudullah
Jahurul Islam	Mominul Haque
Spin Bowlers	
E. Sunny	E. Sunny
A. Razzak	A. Razzak
Fast Bowlers	
A. Hasan	A. Hasan
Farhad Reza	Farhad Reza
M Mortaza	M Mortaza
Shafiul Islam	Shafiul Islam
Ziaur Rahman	Al-Amin Hossain
All-rounders	
S.A Hasan	S.A. Hasan
Nasir Hossain	Nasir Hossain

8.4 Applying Binary Integer Programming to Select a Fantasy Team

The method discussed above was a simple tool to quantify the performance of cricketers by composing their on-field act into a single index and then selecting the best players of different expertise as per the requirement. Let us make our model more effective and in a way more mathematical and closer to reality. How is an actual cricket team/squad selected? The captain of the team is named ahead of team selection. These days the captain is an invited member in most of the team selection meetings. The captain has a say in that meeting and his opinion is given a lot of importance. The squad is selected taking the best players of different expertise given that the captain is already a member of the team. The expertise of the captain also influences the other players to be selected. For example, if the captain of the team is an all-rounder, then fewer numbers of all-rounders are required. The captain being an obvious member of the playing XI, one slot (may be the only slot) for all-rounder is already engaged. Similarly, if the captain is a wicket keeper, then the only slot of a wicket keeper in the playing XI is covered. So, the team management is to decide about the other ten positions that are to be filled up. Those ten positions shall have best players from different expertise excluding the wicket keeper. Also at times the team management is in doubt of keeping a spinning all-rounder in the playing XI or shall take a specialized spinner. The selectors are at crossroads whether to increase the batting strength of the team by including an additional batsman or include a pace bowler. The method discussed in the subsequent sections shall provide us a solution to many such issues.

(i) Here, the team has already named his captain. The expertise of the captain is thus known. As captain is an evident member of the playing XI, so his expertise influences the team requirement.
(ii) The team management does not have fixed number of players of different expertise that are to be included into the playing XI, rather they want to include players so that all the departments are satisfied, the balance of the team is maintained and their chance of winning is optimized.

Here, we execute the entire exercise in a mathematical way. The exercise has two parts (a) to define a method of performance measurement of cricketers and (b) to frame an optimization problem with necessary objective function and constraints that can consider the expertise of the captain and accordingly remodel the constraints.

For (a), one can use the methods discussed in Sect. 8.2 above. But here we improvise on the technique discussed on 8.2 and we feel that the performance measurement technique presented in Sect. 8.4.1 shall be methodologically superior to those discussed in Sect. 8.2.

8.4.1 Performance Measurement

Batting average and strike rate are mostly used to measure the performance of the batsman, while bowling average, economy rate, and bowlers' strike rate are used to measure the performance of the bowlers. But it is widely recognized that such statistics have severe limitations in assessing the true abilities of a player's performance (Lewis, 2005). Further, Lewis (2005) mentioned that the traditional measures of performances do not allow combination of abilities of batting and bowling as they are based on incompatible scales. To overcome this limitation, the following performance measure is proposed.

The performance measure of the i^{th} player is given by,

$$S_i = w_1 S_{i1} + w_2 \delta_i \tag{8.9}$$

where δ_i and $S_{ik}, k = 1, 2, 3$ are as defined in Eq. (8.1). w_1 and w_2 represent the weights of batting and bowling/wicket keeping performance deduced from the entire data set. These are variance stabilizing functions restricting the undue dominance of S_{i1} or S_{i2}/S_{i3} in the composite index due to more variance

$$w_1 = \frac{C}{\sqrt{\operatorname{Var}_i(S_{i1})}} \quad \text{and} \quad w_2 = \frac{C}{\sqrt{\operatorname{Var}_i(\delta_i)}}$$

where $w_1 + w_2 = 1$ and C is a normalizing constant that follows:

$$C = \left[\frac{1}{\sqrt{\operatorname{Var}_i(S_{i1})}} + \frac{1}{\sqrt{\operatorname{Var}_i(\delta_i)}} \right]^{-1} \tag{8.10}$$

The weights shall stabilize the variance of the participating index and deny the undue dominance of any participating factor in the index (Iyenger & Sudarshan, 1982).

The set of equations in (8.9) ensure that the performance of a cricketer in more than one department is included in his performance measure. However, if the player is only a batsman, then $w_2 = 0$ and $w_1 = 1$ and reverse in case the player is only a bowler.

8.4.1.1 Batsman's Performance Measure (S_{i1})

All performance measures of the batsman take into consideration the number of runs scored by the batsman. The runs scored by a batsman, in a given match, depends on the bowling strength of the opposition, condition of the pitch, availability of resources of the batting team in terms of overs and wickets, etc. In addition to that, 50 runs scored by a batsman in a match, where the team total is 150, is more valuable compared to the same number of runs against the same opposition when the team scores 300 plus runs. Considering all these factors, Lemmer (2008) derived a technique that can convert the runs scored in a match to the adjusted runs based on the match condition and opposition's bowling strength. The adjusted runs scored by the i^{th} player in the j^{th} match is denoted by T_{ij} and is defined by,

$$T_{ij} = R_{ij}\left(\mathrm{SR}_{ij}/\mathrm{MSR}_j\right)^{0.5} \tag{8.11}$$

where R_{ij} is the runs scored by the i^{th} batsman in the j^{th} match.

$$\mathrm{SR}_{ij} = \text{Strike Rate of the } i^{\text{th}}\text{batsman in the } j^{\text{th}} \text{ match} = \frac{R_{ij}}{B_{ij}} \times 100 \tag{8.12}$$

where B_{ij} is the number of balls faced by the i^{th} batsman in the j^{th} match and strike rate of the j^{th} match (MSR_j) for all the batsmen of both teams

$$\mathrm{MSR}_j = \frac{\text{total no. of runs scored in the match}}{\text{total no. of balls bowled in the entire match}} \times 100 = \frac{R_j}{B_j} \times 100 \tag{8.13}$$

Thus, this measure compares the performance of a batsman in relation to his peers who also participated in the match. The data requirement for this measure includes individual as well as team performances from all the matches in which the i^{th} player has batted.

These adjusted runs are then used to define the batting performance measure S_{i1}. It often happens that in a match batsman may remain not out because the innings might have got terminated as all the overs got exhausted or the team batting second might have scored the runs necessary for victory or a batsman might have retired hurt or all his other team mates got out. If we consider runs scored by a batsman in different innings, some of which shall be complete innings and remaining few of them may be his not-out scores. This is clearly a case of right truncated data. If a batsman getting out is considered as the event of interest and adjusted runs scored by the batsman in an innings replaces the time lapsed in survival analysis then one can use the Kaplan–Meier estimate of the mean survival time (here mean adjusted runs) as the batting performance measure. So,

$$\text{BP}_i = \text{Kaplan-Meier estimate of the mean adjusted runs scored by the } i^{\text{th}} \text{ batsman}$$

The mean survival time is estimated using the adjusted runs scored by a batsman in his complete (when the batsman gets out) and the incomplete innings (when the batsman remains not out).

The batting performance score (BP_i) thus obtained is then standardized by the average value of BP across all batsmen,

$$S_{i1} = \frac{\text{BP}_i}{\underset{i}{\text{Avg}}(\text{BP}_i)} \tag{8.14}$$

8.4.1.2 Bowler's Performance Measure (S_{i2})

Lemmer (2002) proposed a bowling performance measure called the combined bowling rate (CBR) which is the harmonic mean of three traditional bowling statistics, viz. bowling average, economy rate, and bowling strike rate. If R be the total number of runs conceded by a bowler, W is the total number of wickets taken by a bowler and B is the total number of balls bowled by a bowler in a series of matches. Then, the traditional bowling statistics can now be defined as,

Bowling average $= \frac{R}{W}$,
Economy rate $= \frac{R}{B/6}$,
Bowling strike rate $= \frac{B}{W}$

To bring parity in the numerator of the above factors, a prerequisite of the harmonic mean, the bowling strike rate was adjusted by Lemmer (2002) as follows,

$$\text{Bowling strike rate} = \frac{B}{W} = \frac{B}{W} \times \frac{R}{R} = \frac{\text{RB}}{\text{RW}} = \frac{R}{\text{RW}/B} \tag{8.15}$$

Thus, the combined bowling rate (CBR) defined by Lemmer (2002) as

$$\text{CBR} = \frac{3}{\frac{1}{\text{bowling average}} + \frac{1}{\text{economy rate}} + \frac{1}{\text{bowling strike rate}}} = \frac{3R}{W + (B/6) + W \times \frac{R}{B}} \tag{8.16}$$

Later, Lemmer (2005) improved the CBR to an adjusted measure called CBR* which is more appropriate for quantifying bowling performance for small number of matches. The adjusted combined bowling rate (CBR*) for the i^{th} bowler is given by,

$$\text{CBR}_i^* = 3R_i'/\left[W_i^* + (B_i/6) + W_i^*(R_i'/B_i)\right] \tag{8.17}$$

where

B_i = number of balls bowled by the i^{th} bowler
W_i^* = sum of weights of the wickets taken by the i^{th} bowler
$R_i{}'$ = sum of adjusted runs (RA_{ij}) conceded by the i^{th} bowler in all the innings of the series/matches under consideration $= \sum_{j=1}^{n_i} \text{RA}_{ij}$

$$\text{RA}_{ij} = R_{ij}\left(\text{RPB}_{ij}/\text{RPBM}_j\right)^{0.5} \tag{8.18}$$

where $\text{RPB}_{ij} = \frac{\text{Runs conceded by the } i^{\text{th}} \text{ player in the } j^{\text{th}} \text{ match}}{\text{Balls bowled by the } i^{\text{th}} \text{ player in the } j^{\text{th}} \text{ match}}$

$$\text{RPBM}_j = \frac{\text{Total runs scored in the } j^{\text{th}} \text{ match}}{\text{Total no. of balls bowled by the } j^{\text{th}} \text{ match}}$$

The measure is noteworthy because of the following two inherent issues. The factor, $(\text{RPB}_{ij}/\text{RPBM}_j)$ considers the match situation in which the i^{th} bowler delivered and the factor W_i^* which refuses to give equal importance to all the wickets taken by the bowler but weights them differently based on their batting position. The detailed discussion and the different values of W_i^* is available in Lemmer (2005). The combined bowling rate has a negative dimension, i.e., lower the value the better is the bowler. So to bring parity with the batting performance, the CBR* is inverted and is standardized by the average value of inverse CBR* across all the bowlers, i.e.,

$$S_{i2} = \frac{1/\text{CBR}_i^*}{\underset{i}{\text{Avg}}\left(1/\text{CBR}_i^*\right)} \tag{8.19}$$

8.4.1.3 Wicket Keeper's Performance Measure (S_{i3})

For measuring the performance of a wicket keeper, two factors are considered. They are (i) dismissal rate (ii) bye runs conceded (rate). According to Narayanan (2010), dismissal rate of a wicket keeper is defined as the number of dismissals (stumping and catches) per match. Here, the term 'match' refers only to those matches where the player under consideration kept wicket for his team.

$$\text{Dismissal Rate}\left(D_i'\right) = \frac{\text{Total number of dismissals by the } i^{\text{th}} \text{ player}}{\text{No. of matches in which the } i^{\text{th}} \text{ player kept wickets}} \tag{8.20}$$

The rate in which bye runs were conceded is defined as,

$$\text{Byes Rate}\left(B_i'\right) = \frac{\text{Total bye runs conceded by the } i^{\text{th}} \text{ player}}{\text{No. of matches in which the } i^{\text{th}} \text{ player kept wickets}} \tag{8.21}$$

It is obvious that while 'dismissal rate' has a positive dimension, i.e., positively related to the skill of the player 'bye runs conceded' have a negative dimension. The lesser the byes rate better is the wicket keeper unlike that of the dismissal rate. Thus, instead of B_i' to bring parity between the two rates $(1/B_i')$ is considered. However, in order to combine these two measures, viz. D_i' and $(1/B_i')$ into a single measure, it is necessary to standardize them. The standardized values may be defined as follows:

$$D_i = \frac{D_i'}{\underset{i}{\text{Avg}}(D_i')} \quad \text{and} \quad B_i = \frac{1/B_i'}{\underset{i}{\text{Avg}}\left(1/B_i'\right)}$$

Thus, $D_i \times B_i$ can be considered as a performance measure for wicket keepers, but in order to ensure that dismissal measure and bye rate are comparable, scale adjustment of B_i is necessary by raising a real number α to the exponent of B_i such that standard deviation of D_i and that of B_i^{α} is exactly same. The value of α can be determined by any iterative method (see, Lemmer (2004) for details). So, $D_i \times B_i^{\alpha}$ is reached as a performance measure of the wicket keeper, but this measure considers both the factors, viz. dismissal measure and bye rate as equally important. But as a dismissal leads to loss of resources of the opponent team so it shall get relatively more importance compared to bye runs conceded. Thus, the factor $D_i \times B_i^{\alpha}$ is reformulated with a weighted product so that the relative importance of the two factors can be quantified. This leads to the definition of WK_i as,

$$\begin{aligned} WK_i &= D_i^{\beta} \times (B_i^{\alpha})^{1-\beta}, \quad 0 < \beta < 1 \\ &= \left(\frac{D_i'}{\text{Avg}\left(D_i'\right)}\right)^{\beta} \times \left(\left[\frac{1/B_i'}{\text{Avg}(1/B')_i}\right]^{\alpha}\right)^{1-\beta}, \quad 0 < \beta < 1 \end{aligned} \tag{8.22}$$

The value of β determines the relative importance of the factors and acts as a balance between the dismissal measure and bye rate. The number of bye runs conceded by a wicket keeper also depends on the quality of bowling and the activity is relatively less important than dismissals. Narayanan (2010) allocated 5 points to byes conceded and 40 to dismissals, making the later 8 times more important than saving bye runs.

The wicket keeping performance score (WK_i) thus obtained is then standardized by the average value of WK across all keepers, i.e.,

$$S_{i3} = \frac{\mathrm{WK}_i}{\underset{i}{\mathrm{Avg}}(WK_i)} \tag{8.23}$$

Now, for a given player (as often referred to as the i^{th} player) the values of S_{i1} and S_{i2} or S_{i3} are computed using (8.14), (8.19), and (8.23). The values are then replaced in (8.9) to get the performance score of the i^{th} player (i.e., S_i).

8.4.2 *Optimization Technique*

The optimization technique used for team selection is a binary (0-1) integer programming problem and the solution to the problem is attained using the Solver add-in available in Microsoft Excel. The procedure is discussed in details in Ragsdale (2007).

Suppose the final *XI* is to be composed in such a manner that there are at least four specialist batsman (including two openers), one wicket keeper, at least two fast bowlers, at least two spinners, and at least one all-rounder. The selection needs to have exactly six bowling options available in the playing XI, including the all-rounder(s). One wicket keeper–batsman and at least four specialist batsmen shall be accommodated in the team. Out of the wicket keeper, all-rounder(s) and specialist batsmen in the team, two batsmen should have the capability of opening the innings. In the entire process of selection, the captain gets automatically selected in the playing XI, so the actual decision is to be taken for the remaining ten positions of the team, considering the expertise of the captain himself. This is to be done to avoid over representation or under representation of a particular skill in the playing XI. Let a set of binary coefficients be defined. These coefficients are used in setting the objective function and the constraints.

$\theta_i = 1$ (0), if the i^{th} player is selected for the playing XI (Otherwise)
$b_i = 1$ (0), if the i^{th} player is an opening batsman (Otherwise)
$c_i = 1$ (0), if the i^{th} player is a specialist batsman but not an opener (Otherwise)
$d_i = 1$ (0), if the i^{th} player is a spinner (Otherwise)
$e_i = 1$ (0), if the i^{th} player is a fast bowler (Otherwise)
$f_i = 1$ (0), if the i^{th} player is a wicket keeper (Otherwise)
$g_i = 1$ (0), if the i^{th} player is an all-rounder (Otherwise)

Table 8.6 Constraints for different expertise of a player

Constraint on	$j=$	l'_j
Opener	1	(1, 0, 0, 0, 0)
Middle-order batsman	2	(0, 1, 0, 0, 0)
Spinner	3	(0, 0, 1, 0, 0)
Fast bowler	4	(0, 0, 0, 1, 0)
Wicket keeper	5	(0, 0, 0, 0, 1)
All-rounder	6	(1, 1, 1, 1, 0)

Since the captain is known and is already a member of the team, there are 10 more places to be filled up from a collection of k players (say) of different expertise. Therefore, the first constraint is,

$$\sum_{i=2}^{k} \theta_i = 10 \tag{8.24}$$

The subscript $i = 1$, indicates the captain and since the captain gets an obvious selection so i goes from 2 to k. The constraint in (8.24) ensures that there are exactly 10 players in the team excluding the captain. The other constraints can be defined, only after knowing the expertise of the captain. Here, attempt is made to design the remaining constraints in such a manner that the model can be generalized for any type of expertise of the captain. To do that two column vectors l and p are defined (c.f. Tables 8.6 and 8.7). A player in a cricket team may be either an opening batsman or middle-order batsman or spinner or fast bowler or wicket keeper–batsman or an all-rounder. To each of the expertise we attach a column vector l_j such that $l'_j = (x_1, x_2, x_3, x_4, x_5)$, where the x_i's are binary variables (Bhattacharjee & Saikia, 2016). The values of l'_j corresponding to the different expertise are given in Table 8.6.

Table 8.7 Constraints based on expertise of the captain

Expertise of the captain	p'	Expertise of the captain if all-rounder	p'
Opener	(1, 0, 0, 0, 0)	Fast bowler and opener	(1, 0, 0, 1, 0)
Middle-order batsman	(0, 1, 0, 0, 0)	Spinner and opener	(1, 0, 1, 0, 0)
Spinner	(0, 0, 1, 0, 0)	Fast bowler and middle-order batsman	(0, 1, 0, 1, 0)
Fast bowler	(0, 0, 0, 1, 0)	Spinner and middle-order batsman	(0, 1, 1, 0, 0)
Wicket keeper and opener	(1, 0, 0, 0, 1)		
Wicket keeper and Middle-order batsman	(0, 1, 0, 0, 1)		

Now, based on the expertise of the captain another column vector p is defined. Such that, $p' = (y_1, y_2, y_3, y_4, y_5)$. The values of y's are once again binary variables and is related to the expertise of the captain where the suffix 1, 2, 3, 4, and 5 represents opener, middle-order batsman, spinner, fast bowler, and wicket keeper, respectively. Table 8.7 explains how different expertise of the captain is notified for the column vector p.

It shall be noted that a captain may possess two different abilities at the same time. The vector p is defined, keeping in mind the dual ability that a captain might possess.

Accordingly, the following possible constraints are formulated in addition to (8.24) to select an optimal cricket team.

$$\sum_{i=2}^{k} \theta_i b_i = 2 - l_1' p \tag{8.25}$$

The constraint (8.25) ensures that the team has exactly two opening batsmen. In case, the captain is an opening batsman then the team needs one more opener otherwise two openers are to be selected. This issue is negotiated by $l_1'p$. When the captain is an opener, it may be noted that (c.f. Tables 8.6 and 8.7),

$$\begin{aligned} l_1' p &= (1,\ 0,\ 0,\ 0,\ 0)\,(1,\ 0,\ 0,\ 0,\ 0)' = 1, \\ &= 0 \quad \text{otherwise} \end{aligned}$$

The same would happen even if the captain is simultaneously a wicket keeper as well as an opening batsman or an all-rounder with ability to open the innings.

$$\sum_{i=2}^{k} \theta_i c_i \geq 2 - l_2' p \tag{8.26}$$

The constraint (8.26) ensures that the team has at least two middle-order batsmen. In case, the captain is a middle-order batsman then the team needs at least one more middle-order batsman otherwise at least two middle-order batsmen are to be selected. This issue is negotiated by $l_2'p$. When the captain is a middle-order batsman, then it may be noted that (c.f. Tables 8.6 and 8.7),

$$\begin{aligned} l_2' p &= (0,\ 1,\ 0,\ 0,\ 0)\,(0,\ 1,\ 0,\ 0,\ 0)' = 1, \\ &= 0 \quad \text{otherwise} \end{aligned}$$

The same would happen even if the captain is simultaneously a wicket keeper as well as a middle-order batsman or an all-rounder who is a sound middle-order batsman.

$$\sum_{i=2}^{k} \theta_i d_i \geq 2 - l_3' p \tag{8.27}$$

The constraint (8.27) ensures that the team has at least two spin bowlers. In case, the captain is a spinner then the team needs at least one more spinner otherwise at least two spinners are to be selected. This issue is negotiated by $l_3'p$. When the captain is a spinner, then it may be noted that (c.f. Tables 8.6 and 8.7),

$$\begin{aligned} l_3' p &= (0,\ 0,\ 1,\ 0,\ 0)\,(0,\ 0,\ 1,\ 0,\ 0)' = 1, \\ &= 0, \quad \text{otherwise} \end{aligned}$$

The same would happen even if the captain is an all-rounder who can spin the ball.

$$\sum_{i=2}^{k} \theta_i e_i \geq 2 - l_4' p \tag{8.28}$$

The constraint (8.28) ensures that the team has at least two fast bowlers. In case, the captain is a fast bowler then the team needs at least one more fast bowler otherwise at least two fast bowlers are to be selected. This issue is negotiated by $l_4'p$. When the captain is a fast bowler, then it may be noted that (c.f. Tables 8.6 and 8.7),

$$\begin{aligned} l_4' p &= (0,\ 0,\ 0,\ 1,\ 0)\,(0,\ 0,\ 0,\ 1,\ 0)' = 1, \\ &= 0, \quad \text{otherwise} \end{aligned}$$

The same would happen even if the captain is an all-rounder with fast bowling ability.

$$\sum_{i=2}^{k} \theta_i f_i \geq 1 - l_5' p \tag{8.29}$$

The constraint (8.29) ensures that the team has at least one wicket keeper. In case, the captain is a wicket keeper then the team generally does not need any other wicket keeper otherwise at least one wicket keeper is to be selected. This issue is negotiated by $l_5'p$. The vector p has two possible values, viz. $(1, 0, 0, 0, 1)'$ or $(0, 1, 0, 0, 1)'$. The former one is used when the wicket keeping captain is an opener and the later when the wicket keeping captain is a middle-order batsman. Thus, it may be noted that (c.f. Tables 8.6 and 8.7),

$$\begin{aligned} l_6' p &= (0,\ 0,\ 0,\ 0,\ 1)\,(1,\ 0,\ 0,\ 0,\ 1)' \quad \text{or} \quad (0,\ 0,\ 0,\ 0,\ 1)\,(0,\ 1,\ 0,\ 0,\ 1)' = 1, \\ &= 0, \quad \text{otherwise} \end{aligned}$$

$$\sum_{i=2}^{k} \theta_i g_i \geq 2 - l_6' p \tag{8.30}$$

The constraint (8.30) ensures that the team has at least one all-rounder. In case, the captain is an all-rounder then the team may or may not employ any other all-rounder otherwise at least one all-rounder needs to be selected to bring balance in the team. This issue is negotiated by $l_6'p$. The vector p related to all-rounder can assume any one of the following possible values, viz. $(1, 0, 0, 1, 0)'$ or $(1, 0, 1, 0, 0)'$ or $(0, 1, 0, 1, 0)'$ or $(0, 1, 1, 0, 0)'$. The vector p depends on whether the captain is an all-rounder by the virtue of being a fast bowler and opening batsman or spinner and opening batsman or fast bowler and middle-order batsman or spinner and middle-order batsman respectively (c.f. Table 8.6). Thus, it may be noted that (c.f. Tables 8.6 and 8.7),

$$\begin{aligned} l_6' p = &(1,\ 1,\ 1,\ 1,\ 0)\left[(1,\ 0,\ 0,\ 1,\ 0)' \text{ or } (1,\ 0,\ 1,\ 0,\ 0) \text{ or } (0,\ 1,\ 0,\ 1,\ 0)\right. \\ &\left. \text{or } (0,\ 1,\ 1,\ 0,\ 0)\right] = 2, \\ = &\ 1, \quad \text{otherwise} \end{aligned}$$

Thus, if the captain is an all-rounder the constraint may or may not select any other all-rounder, but if the captain is not an all-rounder then the model shall pick up at least one all-rounder in the optimum team.

Generally, most of the captains these days prefer to take the field with a sixth bowling option including the all-rounders especially in case of Twenty20 matches. In limited overs cricket, the maximum number of overs that can be bowled by a bowler is fixed. For example, in a 50-overs-a-side match, a bowler can bowl a maximum of 10 overs and in case of Twenty20 matches it is only four. Thus, it is mandatory that the fielding team needs to employ at least five bowlers (including all-rounders) in a complete innings. The constraints discussed earlier shall take care of this restriction. However, the authors feel that an optimum team shall have six bowling options (including the all-rounders). Accordingly, a constraint is proposed in (8.31).

$$\sum_{i=2}^{k} \theta_i (d_i + e_i + g_i) \geq 6 - (l_3' + l_4') p \tag{8.31}$$

(8.31) ensures that the team has six or more bowling options including the all-rounders. In case, the captain is an all-rounder or a bowler then the team needs to have at least five more bowling options. But if the captain is a batsman (opener or middle-order) or wicket keeper then not less than six bowling options including the all-rounders are necessary. This issue is negotiated by $(l_3' + l_4')p$. The vector p is related to the captain and can take any value laid down in Table 8.7, depending on the expertise of the captain. Values of the vectors $l_3'$and l_4' are provided in Table 8.6. It can be seen that the term $(l_3' + l_4')p$ results to 1 if the captain is an all-rounder (any expertise) or a bowler (any type either fast or spin) and 0 otherwise. However, this constraint is optional. If a captain is confident on the performance of his bowlers he may not pick up a sixth bowling option. But we feel that in these days of power cricket with lots of Twenty20 cricket around, a team needs to have an additional bowling option. This shall provide a protection to the captain in case one of the regular bowlers goes for lots of runs.

All these constraints from (8.24) to (8.31) shall be used while the optimization function is given in (8.32). The issue is to maximize Z, where

$$Z = \sum_{i=2}^{k} \theta_i S_i \tag{8.32}$$

8.4.2.1 Data Validation and Selection of a Fantasy Team

Often at the end of a tournament, be it cricket or football or any other team sports newspapers and sports magazine generally reports articles which contain a dream team. A dream team (sometimes called as a fantasy team) generally formed by veteran players of the game or reporters or some other concerned celebrity taking best players from all the participating teams satisfying all the expertise generally looked at for team formation. In order to validate the model, the data is collected from the eleventh season of Indian Premier League (IPL) played in 2018. The IPL is the first franchisee-based cricket tournament initiated by the Board of Control for Cricket in India (BCCI), where reputed international players team up with Indian players. In the eleventh season of IPL (IPL XI), eight teams participated. The teams were named after Indian cities or states (provinces) but were owned by franchisees. The teams played each other twice in a home and away basis. At the end of the league, the top four teams qualified for the play-offs. The play-offs comprised of four matches including the final. A total of 60 matches were played. We use the data from this tournament to select a fantasy team using the process discussed in Sects. 8.4.1 and 8.4.2. It was discussed earlier that for such selection it shall be clarified that the players to be considered must play some minimum level of cricket. Hence, the same criterion set in Sect. 8.3 is used here to ensure the entry of cricketers participating in IPL XI into the consideration set. The selection criteria mentioned in Sect. 8.3 provided us with a consideration set of players with the following expertise—19 openers, 24 batsmen, 32 fast bowlers, 17 spinners, 15 all-rounders, and 10 wicket keepers. Remember, here we are not restricted to Indian players only. Now from these players we are to select the optimum playing XI having players from all expertise. The performance measures of the players are provided in Tables 8.8, 8.9, 8.10, 8.11, 8.12 and 8.13.

Now once the performance index of all the players is computed, the next step is to perform the optimization. For that, at the very outset we need to decide who shall captain the team. Without any loss of generality, let us name the captain of the winning team of the tournament as the captain of our fantasy team. The XIth season

Table 8.8 Performance related statistics and index values for opening batsman in IPL XI

Player	Country	Team	Inns	Average adjusted runs	Performance index
A. Rayudu	India	CSK	16	40.64	1.5801
C. Gayle	West Indies	KXIP	9	40.41	1.5712
S. Yadav	India	MI	14	36.31	1.4117
S. Dhawan	India	SRH	16	32.51	1.2640
C. Lynn	Australia	KKR	16	30.93	1.2026
E. Lewis	West Indies	MI	13	29.16	1.1337
P. Shaw	India	DD	9	28.18	1.0956
R. Sharma	India	MI	14	26.89	1.0455
Fa Du Plessis	South Africa	CSK	6	26.7	1.0381
A. Rahane	India	RR	14	25.03	0.9732
A. Hales	England	SRH	6	23.87	0.9281
R. Tripathi	India	RR	12	23.63	0.9187
J. Roy	England	DD	5	23.07	0.8970
B. McCullum	New Zealand	RCB	6	21.4	0.8320
C. Munro	New Zealand	DD	3	21.16	0.8227
G. Gambhir	India	DD	5	14.83	0.5766
M. Vohra	India	RCB	4	14.82	0.5762
D. Short	Australia	RR	7	14.51	0.5642
M. Agarwal	India	KXIP	9	13.79	0.5362

of IPL was won by Chennai Super Kings (CSK), and the captain of the team was MS Dhoni. He is a wicket keeper and a middle-order batsman and accordingly the value of p' is (0, 1, 0, 0, 1). Now with the naming of MS Dhoni as the captain, the only slot of wicket keeper in the XI is filled up. Now for the remaining ten slots we need to select players of other expertise. The optimization model is designed to remodel itself given the expertise of the captain. If you remember at the beginning of Sect. 8.4.2, we decided what shall be composition of our team. It was to be composed of at least four specialist batsman (including two openers), one wicket keeper, at least two fast bowlers, at least one spinner, and at least one all-rounder. The selection needs to have exactly six bowling options available in the playing XI, including the all-rounder(s). The mathematical model, more specifically the binary integer programming problem, shall run successfully and give us the optimum team satisfying all the newly obtained constraints given the naming of MS Dhoni as the captain. The refreshed set of constrains are provided below. The objective function remains same as (8.32).

The first constraint is, $\sum_{i=2}^{k} \theta_i = 10$ where ($i = 1$ represents MS Dhoni). This constraint ensures that 11 players are selected in the team.

$\sum_{i=2}^{k} \theta_i b_i = 2$ # to ensure that team has exactly two opening batsmen

Table 8.9 Performance related statistics and index values for batsman (other than openers) in IPL XI

Player	Country	Team	Inns	Average adjusted runs	Performance index
K. Williamson	New Zealand	SRH	17	48.28	1.8771
A. B. de Villiers	South Africa	RCB	11	47.75	1.8565
V. Kohli	India	RCB	14	39.71	1.5439
S. Iyer	India	DD	13	33.88	1.3173
V. Shankar	India	DD	9	33.45	1.3005
S. Raina	India	CSK	15	31.91	1.2407
S. Samson	India	RR	15	29.71	1.1551
K. Nair	India	KXIP	9	29.56	1.1493
S. Gill	India	KKR	10	28	1.0886
M. Ali	England	RCB	3	26.44	1.0280
Y. Pathan	India	SRH	13	26.21	1.0191
D. Miller	South Africa	KXIP	3	24.86	0.9666
C. de Grandhomme	New Zealand	RCB	8	23.95	0.9312
W. Sundar	India	RCB	5	22.82	0.8872
N. Rana	India	KKR	15	21.32	0.8289
R. Uthappa	India	KKR	16	21.21	0.8247
M. Pandey	India	SRH	13	21.1	0.8204
K. Pollard	West Indies	MI	8	20.35	0.7912
M. Singh	India	RCB	12	18.63	0.7243
K. Gowtham	India	RR	13	17.29	0.6722
A. Finch	Australia	KXIP	8	17.01	0.6614
D. Hooda	India	SRH	8	16.09	0.6256
M. Stoinis	Australia	KXIP	6	14.43	0.5610
S. Billings	England	CSK	9	13.32	0.5179

$\sum_{i=2}^{k} \theta_i c_i \geq 1$ # to ensure that team has at least one more middle-order batsmen as MS Dhoni is already in the team who is simultaneously a middle-order batsman and wicket keeper

$\sum_{i=2}^{k} \theta_i d_i \geq 2$ # to ensure that team has at least two spin bowlers

$\sum_{i=2}^{k} \theta_i e_i \geq 2$ # to ensure that team has at least two fast bowlers

$\sum_{i=2}^{k} \theta_i f_i \geq 0$ # to ensure that team has at least zero number of wicket keeper as MS Dhoni is already in the team as captain

$\sum_{i=2}^{k} \theta_i g_i \geq 1$ # to ensure that team has at least one all-rounder

$\sum_{i=2}^{k} \theta_i (d_i + e_i + g_i) \geq 6$ # to ensure that team has at least six bowling options

Table 8.10 Performance related statistics and index values for spinners in IPL XI

Player	Country	Team	Wkt_wts	Adj_Runs	CBR	Perf. index
S. Lamichhane	Nepal	DD	6.63	75.19159	8.827413	1.4737
Mujeeb Ur Rahman	Afghanistan	KXIP	11.09	201.8784	9.855598	1.3200
I. S. Sodhi	New Zealand	RR	6.81	121.549	10.18334	1.2775
A. Mishra	India	DD	16.66	299.0335	10.42701	1.2477
Rashid Khan	Afghanistan	SH	23.14	419.3675	10.94719	1.1884
S. Gopal	India	RR	11.18	206.8894	11.3643	1.1448
Y. S. Chahal	India	RCB	13.64	304.8838	11.80164	1.1023
Kuldeep Yadav	India	KKR	21.86	399.737	11.80741	1.1018
M. M. Ali	England	RCB	4.25	91.88827	12.32847	1.0552
M. Markande	India	MI	13.88	320.0369	12.85182	1.0123
P. P. Chawla	India	KKR	15.97	369.6016	13.03761	0.9978
Imran Tahir	South Africa	CSK	8.48	188.1345	13.43413	0.9684
H. V. Patel	India	DD	7.79	176.9214	13.81627	0.9416
K. Gowtham	India	RR	10.8	299.6259	13.9831	0.9304
Harbhajan Singh	India	CSK	6.7	257.234	16.22698	0.8017
A. R. Patel	India	KXIP	2.75	191.3829	17.87303	0.7279
S. Nadeem	India	DD	4.15	171.8633	19.04316	0.6832

Binary integer programming problem can be used for solving this optimization problem. Based on the values of the performance measure and the constraints, the binary integer programming selects values either *0* or *1* for θ_i corresponding to the i^{th} player so that the objective function

$$Z = \sum_{i=2}^{k} \theta_i S_i$$

is maximized. $\theta_i = 1$ means selection of the i^{th} player in the team and 0 indicates that the i^{th} player out of the fantasy team. Though there are several softwares which are specialized toward performing optimization through mathematical programming, it can also be done successfully by applying the Solver add-in available in Microsoft Excel. A simple tutorial on binary integer programming using Solver add-in in Excel is available in Ragsdale (2007). The said optimization resulted into the selection of the following players for the fantasy team (Table 8.14).

Thus, we get our fantasy team based on the players participating in the IPL 2018. This team is the optimum team considering the performance of cricketers in that

Table 8.11 Performance related statistics and index values for fast bowlers in IPL XI

Player	Country	Team	Wkt_wts	Adj_Runs	CBR	Perf. index
L. Ngidi	South Africa	CSK	12.06	116.5117	7.4263	1.7518
D. T. Christian	Australia	DD	5.41	63.6140	8.6391	1.5059
A. S. Rajpoot	India	KXIP	13.43	161.9570	9.0901	1.4312
A. Tye	Australia	KXIP	25.57	389.3833	10.5047	1.2384
J. J. Bumrah	India	MI	15.87	259.4128	10.8199	1.2024
U. T. Yadav	India	RCB	23.79	393.7310	11.1098	1.1710
J. C. Archer	West Indies	RR	17.11	285.4694	11.1357	1.1683
M. J. McClenaghan	New Zealand	MI	15.25	276.9550	11.4054	1.1406
D. L. Chahar	India	CSK	13.55	267.9741	11.8971	1.0949
Sandeep Sharma	India	SH	13.98	300.6278	12.2042	1.0935
T. K. Curran	England	KKR	7.01	135.4843	12.4122	1.0660
T. A. Boult	New Zealand	DD	20.76	433.7987	12.7585	1.0481
B. Stanlake	Australia	SH	6.51	136.4025	12.8845	1.0197
S. Kaul	India	SH	23.92	526.2645	12.9719	1.0097
M. Prasidh Krishna	India	KKR	10.45	234.9779	13.2840	1.0029
S. N. Thakur	India	CSK	18.93	430.7955	13.6441	0.9793
B. Kumar	India	SH	10.66	314.5740	13.6905	0.9535
B. Laughlin	Australia	RR	10.25	229.2974	13.7845	0.9502
Mohammed Sir.aj	India	RCB	14.08	352.8409	14.0621	0.9438
C. R. Woakes	England	RCB	8.15	201.4544	14.5949	0.9251
Mustafizur Rahman	Bangladesh	MI	7.22	237.3279	15.7851	0.8914
T. G. Southee	New Zealand	RCB	5.40	237.0505	17.0308	0.8242
D. S. Kulkarni	India	RR	5.45	197.7189	17.2738	0.7639
Basil Thampi	India	SH	4.21	141.0134	17.5470	0.7531
L. E. Plunkett	England	DD	4.15	207.6522	17.8523	0.7414
Mohammed Shami	India	DD	3.86	155.0872	18.6204	0.7287
J. D. Unadkat	India	RR	11.18	502.4874	18.8161	0.6987
M. M. Sarmah	India	KXIP	8.77	358.6894	19.1781	0.6914
K. Khejroliya	India	RCB	2.36	111.5568	19.2905	0.6783
Shivam Mavi	India	KKR	6.21	289.1437	19.3200	0.6744
B. B. Sran	India	KXIP	4.73	253.9990	21.2660	0.6734
C. de Grandhomme	New Zealand	RCB	1.18	129.4804	21.7278	0.6117

Table 8.12 Performance related statistics and index values for wicket keepers in IPL XI

Player	Country	Team	Innings	Avg. Adj. Runs	Dismissal Rate	Bye rate	Perf. index
J. Buttler	England	RR	13	48.24	1.2115	0.8462	1.1728
M. S. Dhoni	India	CSK	15	53.57	1.0693	0.6250	1.1684
Q. de Kock	South Africa	RCB	8	22.82	1.6950	0.7500	1.1325
K. L. Rahul	India	KXIP	12	52.59	1.0693	0.5714	1.1203
D. Karthik	India	KKR	16	36.86	1.2169	0.6875	1.0588
S. Goswami	India	SRH	3	16.97	1.4983	0.1667	1.0277
R Pant	India	DD	14	52.05	0.6268	0.4286	0.9332
I Kishan	India	MI	12	25.05	0.9700	0.9286	0.8043
P Patel	India	RCB	6	29.77	0.7420	0.1667	0.7719
W. Saha	India	SRH	10	15.29	0.8682	0.1000	0.6953

Table 8.13 Performance related statistics and index values for all-rounders in IPL XI

Player	Country	Team	CBR	Avg. Adj Runs	Perf. index
K. H. Pandya	India	MI	9.7667	20.79	1.2921
D. J. Bravo	West Indies	CSK	16.5538	58.48	1.2916
S. R. Watson	Australia	CSK	15.9803	43.32	1.1403
S. P. Narine	West Indies	KKR	12.2984	26.12	1.1362
H. H. Pandya	India	MI	12.6021	26.79	1.1237
A. D. Russell	West Indies	KKR	14.6521	33.11	1.0818
S. A. Hasan	Bangladesh	SH	12.8007	19.61	1.0288
R. A. Jadeja	India	CSK	12.7063	14.2	0.9728
G. J. Maxwell	Australia	DD	13.5911	16.05	0.9412
R. Tewatia	India	DD	13.3975	15.01	0.9402
R. Ashwin	India	KXIP	14.4286	14.24	0.8765
B. Stokes	England	RR	15.3114	15.34	0.8480
B. Cutting	Australia	MI	21.3485	26.91	0.7904
Washington Sundar	India	RCB	19.7281	22.82	0.7831
C. H. Morris	South Africa	DD	17.6958	14.99	0.7534

Team Names: *CSK* Chennai Super Kings, *DD* Delhi Daredevils, *KKR* Kolkata knight Riders, *KXIP* Kings XI Punjab, *MI* Mumbai Indians, *RCB* Royal Challengers Bangalore, *RR* Rajasthan Royals, *SH* Sunrisers Hyderabad

Table 8.14 Fantasy team of IPL 2018 selected using binary integer programming

Name	Team	Country	S_i
Captain (wicket keeper)			
M. S. Dhoni	CSK	India	1.1684
Opener			
A. Rayudu	CSK	India	1.5801
C. Gayle	KXIP	West Indies	1.5712
Batsman			
K. Williamson	SRH	New Zealand	1.8771
A. B. de Villiers	RCB	South Africa	1.8565
Fast bowler			
L. Ngidi	CSK	South Africa	1.7518
D. T. Christian	DD	Australia	1.5059
Spinner			
S. Lamichhane	DD	Nepal	1.4737
Mujeeb Ur Rahman	KXIP	Afghanistan	1.3199
All-rounder			
K. H. Pandya	MI	India	1.2921
D. J. Bravo	CSK	West Indies	1.2916
Total (Z)			16.6884

season of IPL. No other combination of players, subject to the above-mentioned constraints can give a higher value of Z. The team contains three Indians, two each from West Indies and South Africa and one each from Afghanistan, Nepal, and Australia.

8.4.3 Conclusion

The two methods of team/squad selection discussed in this chapter are interesting exercises for students at the under graduate and postgraduate level. Such exercise shall train the students about how to collect data, how to clean and arrange them for computation and to actually compute them using different software. These are real life examples of data analysis. Mostly, students are taught about data analysis but how to collect data smartly is something for which proper training is also required. The exercise shall provide training in both.

Though these were examples from cricket, the same exercise can be modeled to select optimum teams for other team sports. One needs to identify measures of performance and requirement of players of different expertise in that sports for using tools of optimization. Though in most sports including cricket—the selection is based on the recent performance of players, but such models can perform the selection objectively—free from all bias. It is true that several other factors that cannot be quantified come into play while team selection is done, yet the objectively selected optimum teams shall provide guidelines to the selectors and support/oppose their arguments.

An optimistic onlooker can think in this way—the selectors are to select a squad of 15 players for a national level tournament. An optimum team of size 30, taking sufficient number of players from different expertise can be selected using the technique discussed in this chapter. Now this makes the work of the selectors easier. They are now to select 15 players out of the 30 players selected by the optimum model. This shall also eliminate the chance of biasness by the selectors to a great extent.

These models can be used dynamically to choose the exact playing XI out of the 15 players in the squad. It often happens in most sports that a squad of players is selected for a series of matches. At the beginning of each match, the team management (generally comprises the captain and the coach) need to identify the players from the squad who shall play that match. Fresh computation of the performance can be done at the end of each match and the optimum teams can be selected at the end of each match for the next match. Based on the strength of the opponent, playing conditions, etc., the requirement of the team may differ. The constraints in that case can be modeled accordingly to adapt to the changes.

References

Ahmed, F., Jindal, A., & Deb, K. (2012). Multi-objective optimization and decision making approaches to cricket team selection. *Applied Soft Computing*. Accessed October 17, 2012. http://dx.doi.org/10.1016/j.asoc.2012.07.031.

Barr, G. D. I., & Kantor, B. S. (2004). A criterion for comparing and selecting batsmen in limited overs cricket. *The Journal of the Operational Research Society, 55,* 1266–1274.

Bhattacharjee, D., & Saikia, H. (2016). An objective approach of balanced cricket team selection using binary integer programming method. *Opsearch, 53*(2), 225–247.

Boon, B. H., & Sierksma, G. (2003). Team formation: matching quality supply and quality demand. *European Journal of Operation Research, 148,* 277–292.

Bracewell, P. J., & Ruggiero, K. (2009). A parametric control chart for monitoring individual batting performance in cricket. *Journal of Quantitative Analysis in Sports, 5*(3), 1–21.

Brettenny, W. (2010) *Integer optimisation of the selection of a fantasy league cricket team.* M. Sc dissertation, Nelson Mandela Metropolitan University. Accessed December 21, 2010. http://dspace.nmmu.ac.za:8080/jspui/bitstream/10948/…/Warren%20Brettenny.pdf.

Gerber, H., & Sharp, G. (2006). Selecting a limited overs cricket squad using an integer programming model. *South African Journal for Research in Sport, Physical Education and Recreation, 28,* 81–90.

Iyenger, N. S., & Sudarshan, P. (1982). A method of classifying regions from multivariate data. *Economic and Political Weekly*, 2048–2052.

Kamble, A., Rao, R., Kale, A., & Samant, S. (2011). Selection of cricket players using analytical hierarchy process. *International Journal of Sports Science and Engineering, 5,* 207–212.

Lemmer, H. H. (2002). The combined bowling rate as a measure of bowling performance in cricket. *South African Journal for Research in Sport, Physical Education and Recreation, 24,* 37–44.

Lemmer, H. H. (2004). A measure for the batting performance of cricket players. *South African Journal for Research in Sport, Physical Education and Recreation, 26,* 55–64.

Lemmer, H. H. (2005). A method for the comparison of the bowling performances of bowlers in a match or a series of matches. *South African Journal for Research in Sport, Physical Education and Recreation, 27,* 91–103.

Lemmer, H. H. (2008). An analysis of players' performances in the first cricket Twenty20 world cup series. *South African Journal for Research in Sport, Physical Education and Recreation, 30,* 73–79.

Lemmer, H. H. (2011). Team selection after a short cricket series. *European Journal of Sport Science*, 1–7. Accessed February 28, 2012. https://doi.org/10.1080/17461391.2011.587895.

Lewis, A. (2005). Towards fairer measures of player performance in one-day cricket. *Journal of the Operational Research Society, 56,* 804–815.

Lourens, M. (2009). Integer optimization for the selection of a Twenty20 cricket team. M. Sc dissertation, Nelson Mandela Metropolitan University, Accessed on 21 June 2010. http://statssa.gov.za/isi2009/ScientificProgramme/IPMS/0738.pdf.

Narayanan, A. (2010, March 4, 7.13 p.m.). *Comment on* "*Test Wicketkeepers—an Analysis.*" The ESPNcricinfo Blog, 2008. http://blogs.cricinfo.com/itfigures/archives/wicketkeepers/.

Ragsdale, C. T. (2007). *Spreadsheet modeling and decision analysis*. USA: Thomson South-Western.