Probabilistic Match Importance in Professional Sports

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Michael de Lorenzo

B. Sci. (Stats) (Hons)

RMIT University

School of Science

College of Science, Engineering and Health

RMIT University

February, 2018
Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approach research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed. I acknowledge the support that I have for my research received through the provision of an Australian Government Research Training Program Scholarship.

Michael de Lorenzo
February, 2018
Acknowledgements

I wish to acknowledge the following people whose valuable assistance made this dissertation possible.

Dr. Stella Stylianou – Thank you for providing constant guidance and encouragement throughout my candidature. This dissertation would not have been possible without your incredible supervision.

Dr. Ian Grundy – Thank you for your constant input and guidance throughout my candidature. Your supervision helped shape this dissertation into what it is today and I am grateful for your guidance throughout my candidature.

RMIT Sports Statistics Research students – Thank you to both current and past sports statistics research students for your comradeship and helpful advice throughout my candidature. You guys helped make the difficult days all worth it.

Examiners – Thank you for your kind words and helpful suggestions which refined the overall quality of this dissertation.

The Australian Government Research Training Program Scholarship – Thank you for providing support and the opportunity to complete this dissertation.

Kevin, Dawn and Cathy – Thank you to my family for your constant encouragement and support throughout my candidature. None of this would have been possible without it and I am truly grateful.
Summary

Quantifying the importance of a match in professional sports can be beneficial in a variety of circumstances, including in the statistical modelling of match outcomes and attendance figures. Current literature on probabilistic match importance measures have overlooked key information, including the significance of a draw outcome in football (soccer), and the multiple end-of-season outcomes that a team can achieve in a season. The aim of this research was to develop a probabilistic measure of match importance that accounts for different end-of-season outcomes and is adaptable to both a two-result (win/loss) and a three-result (win/draw/loss) sport. By first furthering an existing probabilistic measure of match importance, a new model was developed that calculates the importance of achieving in different end-of-season outcomes using a Markov Chain model for both NBA basketball and Bundesliga football. Furthermore, the new model allows for the significance of a draw outcome to be quantified in the latter; which has not been fully achieved in past literature. The new model was compared to a complete Monte Carlo simulation model, where it was observed that the results between the two modelling techniques were similar. The results from the new model were then applied to an Elo ratings system and a stepwise regression model, where an effect on each model’s predictive performance was observed when matches were classified by their level of importance to the competing teams. The completion of this research helps further the knowledge on calculating and applying match importance within sports modelling.
Contents

Declaration .............................................................................................................................. i

Acknowledgements ............................................................................................................. ii

Summary ............................................................................................................................... iii

List of Tables ........................................................................................................................ ix

List of Figures ....................................................................................................................... xii

1. Introduction ...................................................................................................................... 1

   1.1. Bundesliga and the National Basketball Association .............................................. 2

   1.2. Literature review ....................................................................................................... 4

       1.2.1. Measuring match importance ......................................................................... 4

       1.2.2. Match importance in statistical modelling ................................................. 7

       1.2.3. In-play importance ....................................................................................... 9

       1.2.4. Performance analysis .................................................................................... 10

   1.3. Research questions and publications ..................................................................... 12

       1.3.1. Research questions ....................................................................................... 12

       1.3.2. Publications ................................................................................................ 14

2. History of featured sports .............................................................................................. 16

   2.1. Bundesliga ............................................................................................................ 17

       2.1.1. History .......................................................................................................... 17

       2.1.2. The game ...................................................................................................... 22

   2.2. National Basketball Association ............................................................................ 30
8.4. Discussion .............................................................................................................. 198
8.5. Summary ............................................................................................................... 203
9. Conclusions and future research ............................................................................ 204
  9.1. ParWins importance ............................................................................................ 205
  9.2. Overtake importance .......................................................................................... 206
  9.3. Simulation ........................................................................................................... 207
  9.4. Elo ratings ........................................................................................................... 208
  9.5. Regression analysis .............................................................................................. 209
  9.6. Summary ............................................................................................................. 211
Bibliography .................................................................................................................. 212
Appendix ....................................................................................................................... 219
  Appendix A: Overtake approach – Markov Chain VBA program for Bundesliga football ................................................................. 219
  Appendix B: Overtake approach – Markov Chain VBA program for NBA basketball ................................................................. 227
  Appendix C: Monte Carlo simulation VBA program for Bundesliga football ............................................................................. 233
  Appendix D: Monte Carlo simulation VBA program for NBA basketball ............................................................................. 250
  Appendix E: Binning cut-off values for Elo models ...................................................... 264
  Appendix F: NBA additional Elo models ................................................................. 266
  Appendix G: Frequency of variable inclusion for regression models ....................... 273
List of Tables

Table 2.1: List of Bundesliga teams, part one ................................................................. 25
Table 2.2: List of Bundesliga teams, part two ................................................................. 26
Table 2.3: Sample box-score for FC Bayern Munich, 2014/2015 ..................................... 29
Table 2.4: List of NBA teams .......................................................................................... 37
Table 4.1: NBA Eastern conference standings after 20th March, 2010 ............................. 63
Table 4.2: Positional importance calculation for Atlanta after game 68 ......................... 64
Table 4.3: Example of weighted-sum match importance approaches ............................. 65
Table 5.1: The change in deficit by match outcome dependent on team i and team j ......... 86
Table 5.2: Transition probabilities calculated from the first division Bundesliga constant match outcome probabilities ................................................................................................. 87
Table 5.3: Conditional transitional probabilities calculated from the first division Bundesliga match outcome probabilities ................................................................. 88
Table 5.4: Sample of the transition probability distribution for FC Bayern Munich’s round 31 match from 2008/2009 Bundesliga season ................................................................. 94
Table 5.5: Conditional probabilities for round 31 for FC Bayern Munich and VfL Wolfsburg from 2008/2009 Bundesliga season ................................................................. 95
Table 5.6: Win and draw importance for round 31 FC Bayern Munich and VfL Wolfsburg from the 2008/2009 Bundesliga season ................................................................. 96
Table 6.1: Bayesian conditional probabilities for Bundesliga football split by division and season period .............................................................................................................. 124
Table 6.2: Bayesian conditional probabilities for NBA basketball split by season period ... 125
Table 7.1: Bundesliga Elo ratings model parameters. overall correct prediction percentage and Brier scores .............................................................................................................. 144
Table 7.2: Bundesliga importance categories with required positions split by division ...... 145
Table 7.3: NBA Elo model parameters, overall correct prediction percentage and Brier Scores

Table 7.4: Bundesliga division one correct prediction percentage by Elo type, match classification, season period and importance category

Table 7.5: Bundesliga division two correct prediction percentage by Elo type, match classification, season period and importance category

Table 7.6: Brier score comparison for Bundesliga Division One ‘League Champion’ importance classifications split by season period and Elo favourite expected win probability bin

Table 7.7: Brier score comparison for Bundesliga Division Two ‘League Champion’ importance classifications split by season period and Elo favourite expected win probability bin

Table 7.8: NBA base Elo prediction percentage split by season period and match classification for ‘Conference Champion’ and ‘Top 4’ categories

Table 7.9: NBA base Elo prediction percentage split by season period and match classification for ‘Positions 5-8’ and ‘Positional Sum’ categories

Table 7.10: NBA base Elo prediction percentage split by season period and match classification for ‘Outcome Sum’ and ‘Total Sum’ categories

Table 7.11: Brier score comparison for NBA ‘League Champion’ for periods one to four split by Elo favourite expected win probability bin

Table 7.12: Brier score comparison for NBA ‘League Champion’ for periods five to eight split by Elo favourite expected win probability bin

Table 8.1: Description of independent variables for stepwise regression models

Table 8.2: Descriptive statistics of regression variables for both Bundesliga football and NBA basketball
Table 8.3: Forward stepwise coefficients for first division Bundesliga matches .................. 187
Table 8.4: Forward stepwise coefficients for second division Bundesliga matches ............ 191
Table 8.5: Forward stepwise coefficients for NBA matches ........................................ 194
List of Figures

Figure 2.1: Bundesliga logo ........................................................................................................ 22
Figure 2.2: Dimensions of a FIFA football field ........................................................................... 23
Figure 2.3: Bundesliga playing ball ............................................................................................ 24
Figure 2.4: Final Bundesliga standings for the 2011/2012 first division season ...................... 27
Figure 2.5: NBA logo .................................................................................................................... 34
Figure 2.6: Dimensions of an NBA basketball court ................................................................. 35
Figure 2.7: NBA playing ball ........................................................................................................ 36
Figure 2.8: Final NBA standings for the 2014/15 season .......................................................... 39
Figure 2.9: Sample NBA box-score ............................................................................................ 41
Figure 3.1: Screenshot of 2008/2009 Bundesliga data in Excel spreadsheet ....................... 44
Figure 3.2: Screenshot of 2005/2006 NBA data in Excel spreadsheet .................................. 46
Figure 4.1: Projected wins over-estimation for the 2005/2006 NBA Eastern conference ....... 56
Figure 4.2: Comparison between MaxWins function and its components from the 2005/2006
NBA Eastern conference ............................................................................................................. 57
Figure 4.3: Comparison between ProjWins and MaxWins for 2005/2006 NBA Eastern
conference positions .................................................................................................................. 59
Figure 4.4: Season average positional importance across all NBA seasons split by conference
.................................................................................................................................................. 68
Figure 4.5: Position-specific average importance distribution across all NBA seasons split by
conference .................................................................................................................................... 70
Figure 4.6: First position importance comparison for top two ranked teams from Eastern and
Western conferences during the 2013/2014 NBA season ........................................................ 71
Figure 4.7: Average positional importance across all Bundesliga seasons split by division... 73
Figure 4.8: First position importance for top two teams from 2006/2007 season for division one and two Bundesliga

Figure 5.1: Overtake average positional importance for first and second division Bundesliga

Figure 5.2: Average teams in contention and average positional importance for first division Bundesliga football

Figure 5.3: Average teams in contention and average positional importance for second division Bundesliga football

Figure 5.4: Win, draw and result importance for position one for VfB Stuttgart and Schalke 04 during 2006/2007 Bundesliga season

Figure 5.5: Win, draw and result importance for position one for Karlsruher SC and Hansa Rostock during 2006/2007 2. Bundesliga season

Figure 5.6: Overtake average positional importance for NBA Eastern and Western conferences

Figure 5.7: Average teams in contention and average positional importance for NBA Eastern Conference

Figure 5.8: Average teams in contention and average positional importance for NBA Western conference

Figure 5.9: Position one importance for top two teams in NBA Eastern and Western conferences from 2013/2014 NBA season

Figure 6.1: Comparison between Overtake approach and Monte Carlo simulation with two sets of match outcome probabilities for top-four first division Bundesliga positions

Figure 6.2: Comparison between Overtake approach and Monte Carlo simulation with two sets of match outcome probabilities for bottom-four first division Bundesliga positions
Figure 6.3: Comparison between *Overtake* approach and Monte Carlo simulation with two sets of match outcome probabilities for select second division Bundesliga positions ........130
Figure 6.4: Comparison between *Overtake* approach and Monte Carlo simulation with two sets of match outcome probabilities for top-four NBA basketball positions ......................131
Figure 7.1: Base Elo change in prediction percentage for periods two and three of outcome-specific importance categories for first division Bundesliga ................................................154
Figure 7.2: Base Elo change in prediction percentage for periods two and three of weighted-sum importance categories for first division Bundesliga ................................................156
Figure 7.3: Difference in correct prediction percentage for NBA Elo base ‘Conference Champion’ split by conference and season period .........................................................166
Figure 8.1: $R^2$ difference for first division Bundesliga matches classified by importance ....188
Figure 8.2: $R^2$ difference for second division Bundesliga matches classified by importance 192
Figure 8.3: $R^2$ difference for NBA matches classified by importance .................................196
Chapter 1

Introduction

In most sporting competitions, certain matches can be classified as ‘important’ if they affect a team’s chances of achieving some end-of-season outcome, such as finishing in first position in football (soccer), or qualifying for the playoffs in basketball. Determining the importance of these matches is beneficial with respect to decision making for a number of parties, including team coaches, stadium managers, sports researchers, and eager-eyed supporters. The quantification of match importance can provide these parties with an objective view of how critical a match is to a team; which is the primary aim of this dissertation.

There are numerous questions that can be asked when discussing the importance of a match within professional sports. How can the importance of a match be quantified within both a two-result and a three-result sport? Can this be accomplished using the same probabilistic approach? How can the significance of the draw outcome in professional football be accounted for within calculations? Can the importance of a match be calculated for individual finishing positions? What effect, if any, does the importance of a match have on the outcome of a sporting contest? This dissertation will aim to answer these questions subjectively through the use of statistics and probabilities.
Current literature on quantifying match importance using probabilistic approaches have overlooked key information, such as the different end-of-season outcomes within a sport and the potential significance of the draw outcome in a three-result (win/draw/loss) sport. Furthermore, a probabilistic measure that is adaptable to both a two-result and a three-result sport has yet to be established. Current research also has not explored the possible effect of match importance on a model’s predictive ability, with the importance of a match being primarily applied as an explanatory variable within a model. This leaves an unanswered question: what effect, if any, does the importance of a match have on a model’s predictive performance?

A key aim of this dissertation is to answer the aforementioned questions. This includes developing a probabilistic measure of match importance that is adaptable to both a two-result (win/loss) and a three-result (win/draw/loss) sport. While evaluating how the importance of a match varies by position throughout a season, this dissertation will also investigate the potential effect it may have on a statistical model’s predictive ability. As mentioned, match importance has long been applied within sports modelling but questions remain in regards to how it is best utilised. Can the importance of a match be used to predict the outcome of a match? What effect, if any, does match importance have on an existing predictive model? How can past research be built upon to further the knowledge of quantifying the importance of a match in professional sports?

1.1. Bundesliga and the National Basketball Association

The two sports that feature in this dissertation are football (soccer) and basketball, which are two of the most popular sports in the world. Football is represented by the Bundesliga, which is the national league in Germany; while basketball is represented by the National Basketball Association (NBA), which is the national league within North America.
The NBA was selected primarily due to its popularity and familiarity. Furthermore, research on basketball analytics is a growing field with a vast amount of data freely available. Since the majority of the current research focuses on assessing player and team performance, the importance of matches over the duration of a season, and the potential information gained from applying it within statistical models, has been overlooked thus far. This results in a need for research to help further the knowledge of quantifying the importance of a match in NBA basketball, while also assessing the effect it may have within a predictive model.

The German Bundesliga was selected due to some major differences from basketball, as well as the lack of research on the league within the match importance literature. One of the key distinctions between Bundesliga and the NBA is that the former has three match outcomes (win/draw/loss), while the latter only has two match outcomes (win/loss). Due to the low scoring nature of football, the draw outcome frequently occurs throughout a season, which should be accounted for when quantifying the importance of a match. While measures exist in past research on football, the significance of the draw outcome has been overlooked.

While the different number of match outcomes was a deciding factor in choosing two sports for this dissertation, another key reason was the end-of-season outcomes unique to each sport. In the NBA, the top eight teams from each conference at the conclusion of the season progress to the playoffs, where they compete in a knockout tournament to determine the champion. This differs from Bundesliga football, where the first-place team at the conclusion of the season is awarded the league championship. However, other positions in the Bundesliga standings also result in an end-of-season outcome, such as the top seven positions awarding teams with qualification to post-season European football tournaments. The end-of-season outcomes within both sports are detailed within Chapter 2 of this dissertation. With these factors in mind, this dissertation will aim to create and explore a
measure of match importance that is adaptable to both sports whilst accounting for the different end-of-season outcomes.

1.2. Literature review

This section details previous research on quantifying the importance of a match in professional sports. It includes defining and measuring match importance, the application of this within statistical modelling, the importance of scoring the next set of points during a live match, and its use in sports performance analysis. Although this dissertation focuses on football and basketball, research applicable from other professional sports is also included.

1.2.1. Measuring match importance

The importance of a match in professional sports has been defined and measured in a variety of ways in past research. Certain matches during a season have been subjectively defined as ‘important’ if it affects a team’s chances of winning the championship (Baumeister, 1995), or if the match has an effect on a team’s relegation or promotion opportunities (Goddard & Asimakopoulos, 2004). A match has also been classified as important according to a person’s perspective of a current situation, such as in Moreira et al. (2013), who described a volleyball match as being important according to the pressure felt by competing athletes; and Bojke (2007), who described matches as being called important according to the opinions of sports commentators. Russell (1983) established that there were three main determinants of match importance: the stage of the season, the league standing of the home team compared to the away team, and the current position in the standings (top, middle, or bottom). Subsequently, Bell et al. (2011) included team rivalry as a determinant of match importance.
A popular method of quantifying the importance of a match is to allocate a binary variable dependent on a team’s position within the standings (Forrest & Simmons, 2006; García & Rodríguez, 2002). Audas, Dobson, and Goddard (2002) and Dumangane, Rosati, and Volossovitch (2009) both applied this approach to determine match importance with respect to winning the league championship in football and handball, respectively. Goddard (2005) and Zuber et al. (2005) also applied a binary variable within regression models to predict the outcome of association football matches, but extended the measure to include matches that may affect a team’s promotion or relegation chances. This was also implemented within Forrest, Goddard, and Simmons (2005), who explored the effectiveness of forecasts based on the bookmaker odds in English football.

In a probabilistic sense, Schilling (1994) defined the importance of a team’s next match as the difference between two conditional probabilities: the probability that a team wins a best-of-seven match series in basketball given they win the next match, minus the probability that the team win the series given they lose the next match. This measure, known as the Schilling method, has been applied to other sports to measure the importance of matches over the duration of a season instead of a best-of-seven match series, including Belgian football (Goossens & Belien, 2010; Goossens, Beliën, & Spieksma, 2012) and Major League Baseball (Tainsky & Winfree, 2010). Bedford and Schembri (2006) applied the measure within Australian Rules football, where a cumulative binomial distribution was utilised to calculate the conditional probabilities with respect to finishing within the top eight.

Monte Carlo simulation is often applied to calculate the conditional probabilities for the Schilling (1994) definition of match importance (Lahvička, 2015). The procedure simulates a number of seasons to completion to determine the criticality of matches for a team. Scarf and Shi (2008) applied Monte Carlo simulation to determine the conditional probabilities for English Premier League (EPL) football matches, where a logistic regression model was
applied to predict the match outcomes. However, as noted in both Goossens et al. (2012) and Geenens (2014), the application of the Schilling method to date has failed to account for the draw outcome in football. Furthermore, this means that the importance of the draw outcome in football has yet to be defined within the literature.

Conversely, some measures of match importance have focused on alternative factors instead of the probability of an outcome occurring. Jennett (1984) applied an ad hoc measure of match importance to Scottish football to determine the required number of wins for a team to either win the league championship or avoid relegation. The match importance for finishing in first position, or to avoid relegation, was calculated as the reciprocal of the number of matches required to complete either objective. This method has been applied to research on match attendance in Australian Rules football (Borland & Lye, 1992), and has been extended by Dobson and Goddard (1992) to include a ‘glory’ effect if a team had already secured the league championship with a number of matches remaining in the season.

While the majority of the aforementioned measures evaluate match importance for particular matches, Geenens (2014) sought to explore the decisiveness of a match based on its impact on an entire football tournament. A match was considered decisive if its result would affect the entropy of the tournament, where the entropy was defined as how uncertain the outcome of the tournament was. If a match lowered the entropy then it was considered decisive, whereas high entropy corresponded with a high tournament uncertainty. This approach was subsequently applied within an unpublished research paper by Corona, Tena, and Wiper (2015) to predict the outcomes of football matches, with the authors concluding that the identification of key matches depends on the type of predictive model being used.

Alternative measures of match importance include a ‘win percentage’ model conceived for Major League Baseball by Lei and Humphreys (2013). The method evaluates the
difference in win percentage of the division leader and the current team of interest to determine the importance of the latter’s next match, which was calculated with respect to finishing in first position within the division. Other approaches have required fewer calculations, with Jamieson (2010) defining match importance in multiple sports according to whether or not a match was being played within a post-season tournament, such as the finals or playoffs.

1.2.2. Match importance in statistical modelling

Match importance has been applied as a component within a model to predict the outcome of a sporting contest, determine the effect on attendance figures, and explore the competitive balance of sports leagues. When evaluating any of these areas, the importance of a match is often included as an explanatory variable within a statistical model, such as a Poisson regression model (Goddard, 2005), a probit model (Forrest et al., 2005), or a team ratings system (Stekler, Sendor, & Verlander, 2010).

Using their ‘win percentage’ model, Lei and Humphreys (2013) sought to predict the outcome of Major League Baseball matches, believing that “it is possible GI (game importance) reflects the incentives that teams have to perform well in games” (page 9). Using a generalised estimation equation, it was found that the home team is more likely to win when the match is of high importance to them compared to the visiting team. Using the approach by Jennett (1984), Morley and Thomas (2005) found through logistic regression that the match importance for both the home and away teams were significant in predicting the outcomes of English County one-day cricket matches.

By including the match importance as an explanatory variable, Goddard and Asimakopoulos (2004) applied an ordered probit regression model to forecast match results in English League football. It was determined that the model could be applied to obtain a
positive return in the fixed-odds betting market, while also concluding that the match importance had a significant relationship to a team’s performance. Goddard (2005) compared the forecasting performance of a bivariate Poisson regression model to an ordered probit model with respect to predicting match outcomes in association football. Match importance was included as an explanatory variable within both models, where it was defined as a binary value that equaled one if a team was considered in contention for the league championship, promotion, or relegation. It was concluded that there was little difference between the two models in terms of forecasting performance, while the binary match importance variable was found to be significant within both models.

Jennett (1984) applied match importance to assess attendance figures in Scottish football, concluding that crowd numbers will increase when a match is of high importance with respect to winning the league championship. Borland and Lye (1992) found that the importance of a match had a positive effect on crowd size in Australian Rules football, while Paul (2003), defining match importance based on team rivalry, reported similar results within the National Ice Hockey League (NHL). Using multiple linear regression, Baimbridge (1997) found that match significance, measured according to the probability of winning the championship, had a positive contribution to attendance figures at the 1996 European Football Championships. Similar research has also been completed within Spanish football (García & Rodríguez, 2002) and international rugby union (Owen & Weatherston, 2004), while other studies have evaluated television viewership (Baimbridge, Cameron, & Dawson, 1996), fan interest in teams (Tainsky & Stodolska, 2010), and changes in the structure of sports leagues (Cairns, 1987; Goossens et al., 2012).

Further research on match importance has focused on the reaction of a team’s stock price to certain match results (Palomino, Renneboog, & Zhang, 2009). Using two definitions of match importance, rivalry and proximity to the conclusion of the season, Bell et al. (2011)
reported that the outcome of important matches had a modest impact on the share price of English football teams. Distinguishing match importance according to the stage of a football tournament, Hanke and Kirchler (2012) explored the stock prices of the sponsors located on a team’s jersey. It was concluded that matches of high importance had a greater impact on a sponsors’ stock prices compared to matches of low importance.

1.2.3. In-play importance

Research on match importance in sports has also been evaluated within live gameplay, where the significance of scoring next is assessed with respect to winning the match. A common approach is to determine the win probability for both teams throughout a contest, and then evaluate the consequence of the next score in the match on each team’s chances of winning. Examples of this include using a Poisson distribution model in football (Vecer, Ichiba, & Laudanovic, 2007), using in-play match statistics in tennis (Barnett & Clarke, 2005), and applying a Brownian motion model in NBA basketball (Stern, 1994).

Preceding the work completed by Schilling (1994), Morris (1977) defined the importance of a point in tennis as the difference between success probabilities conditional on which player scores next in the current game. This approach was further explored by Croucher (1986), who presents formula for calculating the conditional probability of a player winning given any score line, where each score is ranked from highest to lowest importance. O'Donoghue (2001) also evaluated the importance of points using the conditional probability approach, concluding that the most critical score lines in singles tennis are 30-40, 30-30 and deuce.

In recent research, Ladds and Bedford (2010) evaluated the importance of points in badminton in order to create a serving strategy. Using the Morris (1977) definition, it was concluded that there were certain times during a match when a player should attempt a long
serve or a short serve, which was dependent on how critical it was to win the next point. González-Díaz, Gossner, and Rogers (2012) also applied the conditional probability approach within tennis to evaluate a player’s ability to adapt to critical moments of a match, concluding that there was a significant relationship between a player’s ability to cope with important moments and their overall career success.

Goldman and Rao (2012) assessed the importance of points during a basketball match using a standard normal distribution function, which required model inputs on the number of possessions remaining, and the mean and variance of the points per possession for each team. The standard normal distribution function then calculated the probability that the home team would win the match, given the number of possessions remaining and the current score line. The results were then applied to assess the free-throw shooting and offensive rebounding ability of the home and away teams, where it was concluded that the home team was significantly better at free-throw shooting during the final eight minutes of the match. Other point importance approaches include Cervone et al. (2014), who applied player-tracking to assign a point value to each moment of a team’s possession during a basketball match to determine the optimal time for a team to shoot; and Quarrie and Hopkins (2014), who defined kick importance in international rugby matches as a function of the time remaining and the score difference between the two competing teams.

1.2.4. Performance analysis

While it has been discussed that match importance has been applied in statistical modelling of attendance figures and match outcomes, it has also been utilised to evaluate both team and player performance. This research has focused on determining if the match or point importance has an effect on a team or player’s ability to perform. For example, Dosil (2005) found that the in-game score line had a considerable impact on a football player’s
psychological responses, while Hoedaya and Anshel (2003) found that professional athletes will either accept a situation or seek social support when playing an important match.

In team performance analytics, Audas et al. (2002) explored the possible effect that managerial change has on a team’s performance in English football. While defining match importance using a binary value, it was concluded that football managers were replaced in-season when their team was facing a high degree of uncertainty with respect to avoiding relegation. Goldman and Rao (2013) assessed the effect that the score line has on a team’s willingness to attempt a three-point shot in NBA basketball, concluding that trailing teams will take the risk of shooting more three-point shots as the deficit increases. Using a cumulative binomial distribution function to calculate the conditional probabilities within the Schilling match importance definition, Bedford and Schembri (2006) found that teams who finish within the top two or bottom three in Australian Rules football will play their most important matches at the start of the season.

Bradley and Noakes (2013) evaluated the effect that match importance has on a player’s running performance in professional football. By defining match importance according to whether or not a team was in contention for the championship, promotion, or relegation, it was concluded that players exerted more energy at the start of critical matches before being unable to maintain their output for the remainder of the contest. Pollard (2004), using the point importance definition by Morris (1977), found evidence that some players have the ability to increase their performance during important moments of a tennis match. Further research on the effect of importance on player performance has been explored within tennis (Knight & O’Donoghue, 2011) and Australian Rules football (O’Shaughnessy, 2006).

While it has been discussed that match importance has been assessed throughout past literature, there are still some areas that require work. This includes creating a probabilistic
measure of match importance that is adaptable to both a two-result sport and a three-result sport, where the draw outcome can be accounted for within the latter. Furthermore, investigating the possible effect of match importance on a predictive model’s performance has yet to be completed. This dissertation will aim to add to the knowledge of quantifying match importance by developing an adaptable measure and exploring its application within sports modelling; which is completed by building upon and further developing an existing probabilistic measure of match importance.

1.3. Research questions and publications

This section contains the research questions that this dissertation aims to answer. The questions are sorted according to their relevant chapter. Publications, both peer-reviewed journal and full-refereed conference papers, are also included along with the chapter in which they appear.

1.3.1. Research questions

Chapter 4: ParWins importance

(i) How do we define ‘importance’ in the context of sports matches?
(ii) Can the ParWins calculations be improved, particularly in late season scenarios?
(iii) How does the average match importance by position vary throughout a season?
(iv) To what extent can the existing ParWins measure be successfully adapted into two-result and three-result sports?

Chapter 5: Overtake importance

(i) How can the draw outcome in football be accounted for in probabilistic match importance calculations?
(ii) To what extent can change in focal model inputs for the new *Overtake* approach improve the average match importance results for both NBA basketball and Bundesliga football?

(iii) When during the season can the draw outcome be beneficial to a football team?

**Chapter 6: Simulation**

(i) To what extent can the *Overtake* approach provide similar results to a complete Monte Carlo simulation procedure with a reduction in computational complexity, and in what circumstances does it fail?

(ii) How does a variation of match outcome probabilities affect the average match importance distributions?

(iii) What are the key advantages/disadvantages of the *Overtake* approach over the Monte Carlo simulation model for calculating the importance of a match within both sports?

**Chapter 7: Elo ratings**

(i) To what extent does match importance have an effect on the predictive accuracy of an Elo ratings system?

(ii) How do the results differ by season period in both NBA basketball and Bundesliga football?

(iii) In terms of overall effect on the change in correct prediction percentage, how do different summations of the positional importance compare?
Chapter 8: Regression analysis

(i) To what extent is the match importance generated from the Overtake approach a significant predictor of match outcomes in both NBA basketball and Bundesliga football?

(ii) To what degree can the linear regression model results be affected by binning matches by their importance?

(iii) Does binning matches by their level of importance have a greater effect within NBA basketball or Bundesliga football, or are the results consistent across the sports?

1.3.2. Publications

Chapter 4: ParWins importance


Chapter 5: Overtake importance

Chapter 6: Simulation

Chapter 2

History of featured sports

This section provides a brief history of both Bundesliga football and the National Basketball Association (NBA). Section 2.1 focuses on Bundesliga football while Section 2.2 focuses on NBA basketball. Within both sections, the origin of each sport, as well as the development and evolution into modern times, is detailed. This includes team rivalries, changes to the league structure, and the current end-of-season outcomes available to teams. Key references for further reading are also provided where applicable.
2.1. Bundesliga

This section details the history of Bundesliga football, including the origins of German football and the development of the national league. This section also covers the basic rules of professional football and the key statistics that are recording during a match. The information presented in this section was summarised from Hesse (2003) and Donnelly (2013).

2.1.1. History

In the 18th and early 19th centuries, a primitive, rule-less version of football was played between villages consisting of an undefined number of players. In 1863, the English Football Association (FA) was formed to establish standard rules, including the number of players per side, the size of the ball and the dimensions of the playing field. The structured game gained popularity quickly in England and was soon introduced into German schools in 1874, where the game was only played by privileged school students.

2.1.1.1. Pre 1930’s

The game slowly began to grow in popularity after its introduction into Germany in 1874, with teachers and past students establishing amateur football clubs outside of school. These clubs were created as part of existing gymnastic societies, where matches were often played on cow paddocks or rubbish tips. However, at the time gymnastics was the most popular sport in Germany and many gymnasts disliked the new sport, which caused football players to break away and create their own clubs specifically for football. The earliest German football clubs were formed in the 1890’s and included BFC Germania 1888 and Hamburger SV.
A number of semi-professional competitions began to emerge in regional areas during the 1890’s, including the German Football and Cricket Association of Berlin, and the South German Football Association. Soon after, the German people decided to form a governing body for the sport, which was called the Deutscher Fußball-Bund (DFB). The DFB decided to hold a national championship in 1903, which would be contested by the victors of the smaller regional competitions. The first championship was contested between DFC Prag and VfB Leipzig, where the latter won 7-2.

Despite the national championship being established in 1903, the sport struggled to gain widespread popularity, with media support virtually non-existent. It wasn’t until the Crown Prince attended a match that the sport gained a considerable following, as the sport was now seen to be ‘socially acceptable’. This new found popularity caused the attendance figures for the national championship to increase year-to-year, eventually leading football to become the most popular sport in Germany. The sport was particularly popular to industry workers, who began competing in the West German Football Association.

After war broke out across Europe, many football clubs lost players as it was their duty to enlist. Competitive games continued, however, until the latter stages of the war, where football fields were converted into potato patches to help with the food shortage. It has also been debated that a Christmas truce was called during the war, during which German soldiers played football against British soldiers (Barajas, 2014). After the war, people used football as a way of distracting themselves from the hardships they were now faced with. The first post-war national championship was held in 1920 in front of more than 30,000 spectators, with FC Nuremberg defeating SpVgg Fürth 2-0.

During the mid-1920’s, a push was made to make football a professional sport in Germany. At the time, the DFB considered football to be an amateur sport and refused to
allow players to receive match payments. Soon, players started receiving illegal payments from club officials, who sought to attract the best players to their football clubs. Opinion were split on whether making the sport professional was the right decision, with the poorer competitions being in favour of player payments due to the Great Depression. The matter remained unresolved at the conclusion of the decade.

2.1.1.2. 1930’s – 1960’s

In the early 1930’s, a number of regional leagues were re-organised into sixteen football organisations, where the winner of each league would advance to compete for the national championship. However, the breakout of the Second World War led to a temporary halt of play as football players were drafted into the German army. Matches recommenced in 1940 as a method of distracting the German people from the war, with the 1940 national championship being played in front of 100,000 spectators. However, travelling to matches soon became too dangerous for some teams, while others struggled to field a team altogether.

Following the conclusion of the war, Germany was divided into several occupied zones. This separation made it difficult to re-establish an organised football league, especially given that planned events were banned in most occupied zones. Eventually, some zones began allowing the re-formation of old football clubs, but under the condition that the teams change their names. A new top tier league was also established in West Germany called the Oberligen, which consisted of teams from five regions: the North, the South, the Southwest, the West and Berlin.

In the early 1950’s, a West German national team was assembled to compete in the 1954 FIFA World Cup, which was organised every four years by the world football’s governing body, the Fédération Internationale de Football Association (FIFA). The team would surprise many by winning the World Cup, coming from behind to defeat Hungary in
the final. This success led many to argue that Germany deserved their own professional national league, regardless of the occupied zones.

In 1962, the DFB announced a new national football league to replace the five-division Oberligen. The league would be called the Bundesliga and would feature sixteen teams chosen from the Oberligen divisions. Teams were chosen based on the number of match points accumulated over the past decade, causing controversy as some successful teams were not selected. The DFB also introduced a second tier league called the Regionalliga, where Bundesliga teams could be relegated to if they finished last during league play. The Bundesliga officially commenced in 1963, with SV Werder Bremen defeating Borussia Dortmund in the first ever match. During the first seven years of the newly formed league, seven different teams claimed the Bundesliga championship.

When the league commenced, the DFB placed a maximum salary on each player in an attempt to keep the competition even. However, illegal player payments resurfaced again shortly after the formation of the new league, with Hertha BSC being found guilty and automatically relegated to the Regionalliga. Thirteen other teams were also found guilty but escaped punishment. During this same period of time, the DFB decided to expand the number of teams within the Bundesliga from sixteen to eighteen.

2.1.1.3. 1970’s – Present day

Illegal player payments continued into the 1970’s, with tape recordings emerging of players agreeing to accept money in return for performing poorly during particular matches. Despite initially ignoring these allegations, the DFB launched an investigation and found fifty players, six club officials and two coaches guilty, with penalties ranging from financial fines to life bans. The DFB concluded that the rule relating to maximum salary was to blame for
the bribery scandal, and that the only way to proceed was to remove the rule and allow teams to pay players whatever they wished.

During the 1970’s, the league saw the start of a new rivalry between two teams, FC Bayern Munich and Borussia Mönchengladbach. The two teams spent a majority of the 1960’s developing their young players and were ready to compete for the Bundesliga championship by the end of the decade. Borussia Mönchengladbach broke through first, winning back-to-back Bundesliga titles in 1970 and 1971. FC Bayern Munich then went on to win the next three championships before Borussia Mönchengladbach won the next three after. During this same period, the DFB introduced a new second division called the “2 Bundesliga”, with the Regionalliga being demoted to the third division.

In the 1980’s, FC Bayern Munich established itself as the league’s most dominant team, winning seven championships in the decade alone. The league also saw an increase in the number of players departing for other international football competitions. However, the fall of the Berlin Wall in 1989 allowed many of the Bundesliga teams to poach some of the best young players from the secluded East Germany, while two East German teams were also selected to compete in the Bundesliga. The new teams and players led to an increase in the popularity of the league, with Borussia Dortmund becoming the first team in the league’s history to average 60,000 spectators per match in 1991.

At the turn of the century, FC Bayern Munich re-established itself as the premier Bundesliga team, winning three titles in a row from 1999 to 2001. The team would win a further ten titles between 2002 and 2017 for a total of 27 Bundesliga championships. The league itself has also continued to grow, with the Bundesliga consistently ranked as one of the highest attended sports leagues in the world (sportingintelligence.com, 2017). The popularity of the league also saw an increase in talent within the lower leagues, with the “3
Liga” replacing the Regionalliga as the third tier division before the commencement of the 2008/2009 season.

2.1.2. The game

This section covers the current structure of Bundesliga football to provide an understanding of the league. Topics of interest include the teams that feature in this dissertation, details of the scoring system, the rule and objectives, the playing field and positions, the fixture, and the structure of the league standings.

![Bundesliga logo](image)

Figure 2.1: Bundesliga logo

2.1.2.1. Field and playing positions

Football is played on a rectangular grass field where the objective is to score a ball through the goals situated at both ends. Figure 2.2 details a standard FIFA playing field, which includes two penalty boxes and four flag markers to indicate the corners of the field.
There are a total of eleven players for each team on the field during a match. The players are typically divided into four positions: defenders, midfielders, forwards and a sole goalkeeper. The number of players within each position, excluding the sole goalkeeper, depends on the formation of the team. For example, a 4-4-2 formation would mean a team has four defenders, four midfielders and two forwards. Each team has seven players on the bench and are allowed to make substitutions to their line-up during a match, with a maximum of three changes allowed per side in Bundesliga football.

2.1.2.2. Rules and objectives

Football is played over a period of 90 minutes where the objective is to score more goals than your opponent. Goals are scored by putting the ball into your attacking goal using your feet, head or torso; the use of hands is prohibited. In Bundesliga football, if the scores are level at the end of the 90 minutes then the match is called a draw.

During a contest, no player, excluding the goalkeeper, is permitted to touch the ball with their hands. If this occurs, a foul is called and the opposition receives a free kick. Fouls are assigned to teams by the sole field referee, who can also allocate a yellow card (warning)
or a red card (ejection) for unsportsmanlike conduct. If a foul occurs within the penalty box, then the non-offending team receives a free kick directly in front of the goals.

All players on the field, excluding the goalkeepers, must remain onside throughout the match. A player is deemed to be offside if they are closer than the last defender to their attacking goal. However, the attacking player can receive the ball behind the defender if their starting position is in front or equal to the defender when a pass is made. A sideline official either side of the field tracks the last defender to determine if an offside infringement has occurred. The sideline officials are also responsible for determining when the ball goes out-of-bounds and which team receives the ball to throw back into play.

![Bundesliga playing ball](image)

Figure 2.3: Bundesliga playing ball

### 2.1.2.3. Teams

There are a total of 18 teams in both the first and second division Bundesliga. The competing teams within each division change year-to-year depending on which teams are promoted and relegated. This section lists only the teams that may feature in the forthcoming chapters of this dissertation. Information on these teams is provided in Tables 2.1 and 2.2, including each team’s official name, location and the total number of Bundesliga titles won as at the completion of the 2016/2017 season.
<table>
<thead>
<tr>
<th>Logo</th>
<th>Team Name</th>
<th>Location</th>
<th>Year Formed</th>
<th>Bundesliga Titles</th>
<th>Logo</th>
<th>Team Name</th>
<th>Location</th>
<th>Year Formed</th>
<th>Bundesliga Titles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. FC Heidenheim</td>
<td>Heidenheim</td>
<td>1846</td>
<td>0</td>
<td></td>
<td>FC Augsburg</td>
<td>Augsburg</td>
<td>1907</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1. FC Kaiserslautern</td>
<td>Kaiserslautern</td>
<td>1900</td>
<td>2</td>
<td></td>
<td>FC Bayern Munich</td>
<td>Munich</td>
<td>1900</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>1. FC Köln</td>
<td>Cologne</td>
<td>1948</td>
<td>2</td>
<td></td>
<td>FC Carl Zeiss Jena</td>
<td>Jena</td>
<td>1903</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1. FC Nürnberg</td>
<td>Nuremberg</td>
<td>1900</td>
<td>1</td>
<td></td>
<td>FC Energie Cottbus</td>
<td>Cottbus</td>
<td>1966</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1. FC Union Berlin</td>
<td>Berlin</td>
<td>1966</td>
<td>0</td>
<td></td>
<td>FC Erzgebirge Aue</td>
<td>Aue</td>
<td>1949</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1. FSV Mainz 05</td>
<td>Mainz</td>
<td>1905</td>
<td>0</td>
<td></td>
<td>FC Hansa Rostock</td>
<td>Rostock</td>
<td>1965</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Alemannia Aachen</td>
<td>Aachen</td>
<td>1900</td>
<td>0</td>
<td></td>
<td>FC Ingolstadt 04</td>
<td>Ingolstadt</td>
<td>2004</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Arminia Bielefeld</td>
<td>Bielefeld</td>
<td>1905</td>
<td>0</td>
<td></td>
<td>FC Schalke 04</td>
<td>Gelsenkirchen</td>
<td>1904</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Bayer Leverkusen</td>
<td>Leverkusen</td>
<td>1904</td>
<td>0</td>
<td></td>
<td>FC St. Pauli</td>
<td>Hamburg</td>
<td>1910</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Borussia Dortmund</td>
<td>Dortmund</td>
<td>1909</td>
<td>5</td>
<td></td>
<td>Fortuna Düsseldorf</td>
<td>Düsseldorf</td>
<td>1895</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Borussia Mönchengladbach</td>
<td>Mönchengladbach</td>
<td>1900</td>
<td>5</td>
<td>FSV Frankfurt</td>
<td>Frankfurt</td>
<td>1899</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dynamo Dresden</td>
<td>Dresden</td>
<td>1953</td>
<td>0</td>
<td></td>
<td>Hamburger SV</td>
<td>Hamburg</td>
<td>1887</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Eintracht Braunschweig</td>
<td>Braunschweig</td>
<td>1895</td>
<td>1</td>
<td></td>
<td>Hannover 96</td>
<td>Hanover</td>
<td>1896</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Eintracht Frankfurt</td>
<td>Frankfurt</td>
<td>1899</td>
<td>0</td>
<td></td>
<td>Hertha BSC</td>
<td>Berlin</td>
<td>1892</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2.1: List of Bundesliga teams, part one
<table>
<thead>
<tr>
<th>Logo</th>
<th>Team Name</th>
<th>Location</th>
<th>Year Formed</th>
<th>Bundesliga Titles</th>
<th>Logo</th>
<th>Team Name</th>
<th>Location</th>
<th>Year Formed</th>
<th>Bundesliga Titles</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>Karlsruher SC</td>
<td>Karlsruhe</td>
<td>1894</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>SV Sandhausen</td>
<td>Sandhausen</td>
<td>1916</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>Kickers Offenbach</td>
<td>Offenbach an Main</td>
<td>1901</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>SV Wacker Burghausen</td>
<td>Burghausen</td>
<td>1930</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>MSV Duisburg</td>
<td>Duisberg</td>
<td>1902</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>SV Wehen Wiesbaden</td>
<td>Wiesbaden</td>
<td>1926</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>RB Leipzig</td>
<td>Leipzig</td>
<td>2009</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>SV Werder Bremen</td>
<td>Bremen</td>
<td>1899</td>
<td>4</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>Rot Weiss Ahlen</td>
<td>Ahlen</td>
<td>1996</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>TSG 1899 Hoffenheim</td>
<td>Hoffenheim</td>
<td>1899</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>Rot Weiss Essen</td>
<td>Essen</td>
<td>1907</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>TSV 1860 München</td>
<td>Munich</td>
<td>1860</td>
<td>1</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>Rot-Weiß Oberhausen</td>
<td>Oberhausen</td>
<td>1904</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>TuS Koblenz</td>
<td>Koblenz</td>
<td>1911</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>SC Freiburg</td>
<td>Freiburg</td>
<td>1904</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>VfB Stuttgart</td>
<td>Stuttgart</td>
<td>1893</td>
<td>3</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>SC Paderborn 07</td>
<td>Paderborn</td>
<td>1907</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>VfL Bochum</td>
<td>Bochum</td>
<td>1938</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>SpVgg Greuther Fürth</td>
<td>Fürth</td>
<td>1903</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>VfL Osnabrück</td>
<td>Osnabrück</td>
<td>1899</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>SpVgg Unterhaching</td>
<td>Unterhaching</td>
<td>1925</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>VfL Wolfsburg</td>
<td>Wolfsburg</td>
<td>1945</td>
<td>1</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>SSV Jahn Regensburg</td>
<td>Regensburg</td>
<td>1907</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>VfR Aalen</td>
<td>Aalen</td>
<td>1921</td>
<td>0</td>
</tr>
<tr>
<td><img src="image" alt="Logo" /></td>
<td>SV Darmstadt 98</td>
<td>Darmstadt</td>
<td>1898</td>
<td>0</td>
<td><img src="image" alt="Logo" /></td>
<td>VIR Aalen</td>
<td>Aalen</td>
<td>1921</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2.2: List of Bundesliga teams, part two
2.1.2.4. Fixture

Bundesliga football is played over the course of ten months, commencing in August and finishing in May the following year. The season is split into 34 rounds (match days) where each team plays each other twice, once at home and once away. This occurs in both the first division and the second division. Matches can sometimes be rescheduled depending on the fixture for European football competitions, including the UEFA Champions League and Europa League.

2.1.2.5. Bundesliga standings

The standings, or league table, are determined by the total number of match points accumulated by each team throughout the season. In Bundesliga football, a win is worth three points, a draw is worth one point, and a loss is worth zero points. In the event that two or more teams have the same number of match points, the total goal difference and total number of goals scored is applied as a tie-breaker. An example of the final standings is presented in Figure 2.4.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Played</th>
<th>Wins</th>
<th>Draws</th>
<th>Loses</th>
<th>Goals For</th>
<th>Goal Diff.</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Borussia Dortmund</td>
<td>34</td>
<td>25</td>
<td>6</td>
<td>3</td>
<td>80</td>
<td>55</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td>FC Bayern Munich</td>
<td>34</td>
<td>23</td>
<td>4</td>
<td>7</td>
<td>77</td>
<td>55</td>
<td>73</td>
</tr>
<tr>
<td>3</td>
<td>FC Schalke 04</td>
<td>34</td>
<td>20</td>
<td>4</td>
<td>10</td>
<td>74</td>
<td>30</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>Borussia M’gladbach</td>
<td>34</td>
<td>17</td>
<td>9</td>
<td>8</td>
<td>49</td>
<td>25</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>Bayer Leverkusen</td>
<td>34</td>
<td>15</td>
<td>9</td>
<td>10</td>
<td>52</td>
<td>8</td>
<td>54</td>
</tr>
<tr>
<td>6</td>
<td>VfB Stuttgart</td>
<td>34</td>
<td>13</td>
<td>8</td>
<td>11</td>
<td>63</td>
<td>17</td>
<td>53</td>
</tr>
<tr>
<td>7</td>
<td>Hannover 96</td>
<td>34</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td>41</td>
<td>-4</td>
<td>48</td>
</tr>
<tr>
<td>8</td>
<td>VfL Wolfsburg</td>
<td>34</td>
<td>13</td>
<td>6</td>
<td>16</td>
<td>47</td>
<td>-13</td>
<td>44</td>
</tr>
<tr>
<td>9</td>
<td>SV Werder Bremen</td>
<td>34</td>
<td>11</td>
<td>9</td>
<td>14</td>
<td>49</td>
<td>-9</td>
<td>42</td>
</tr>
<tr>
<td>10</td>
<td>1. FC Nürnberg</td>
<td>34</td>
<td>12</td>
<td>6</td>
<td>16</td>
<td>38</td>
<td>-11</td>
<td>42</td>
</tr>
<tr>
<td>11</td>
<td>TSG 1899 Hoffenheim</td>
<td>34</td>
<td>10</td>
<td>11</td>
<td>13</td>
<td>41</td>
<td>-6</td>
<td>41</td>
</tr>
<tr>
<td>12</td>
<td>SC Freiburg</td>
<td>34</td>
<td>10</td>
<td>10</td>
<td>14</td>
<td>45</td>
<td>-16</td>
<td>40</td>
</tr>
<tr>
<td>13</td>
<td>1. FSV Mainz 05</td>
<td>34</td>
<td>9</td>
<td>12</td>
<td>13</td>
<td>47</td>
<td>-4</td>
<td>39</td>
</tr>
<tr>
<td>14</td>
<td>FC Augsburg</td>
<td>34</td>
<td>8</td>
<td>14</td>
<td>12</td>
<td>36</td>
<td>-13</td>
<td>38</td>
</tr>
<tr>
<td>15</td>
<td>Hamburger SV</td>
<td>34</td>
<td>8</td>
<td>12</td>
<td>14</td>
<td>35</td>
<td>-22</td>
<td>36</td>
</tr>
<tr>
<td>16</td>
<td>Hertha BSC</td>
<td>34</td>
<td>7</td>
<td>10</td>
<td>17</td>
<td>38</td>
<td>-26</td>
<td>31</td>
</tr>
<tr>
<td>17</td>
<td>1. FC Köln</td>
<td>34</td>
<td>8</td>
<td>6</td>
<td>20</td>
<td>39</td>
<td>-36</td>
<td>30</td>
</tr>
<tr>
<td>18</td>
<td>1. FC Kaiserslautern</td>
<td>34</td>
<td>4</td>
<td>11</td>
<td>19</td>
<td>24</td>
<td>-30</td>
<td>23</td>
</tr>
</tbody>
</table>

Figure 2.4: Final Bundesliga standings for the 2011/2012 first division season
In the first division, there are no playoffs or finals at the end of the season. Instead, the first-place team is awarded the Bundesliga championship and receives automatic qualification for the group stage of the UEFA Champions League. The teams who finish in second and third place also receive qualification for the competition, while the fourth-place team qualifies for the competition’s playoff round. The two teams who finish in positions five and six in the standings qualify for the group stage and playoff round, respectively, of the UEFA Europa League. Finally, the seventh-place team can also qualify for the Europa League depending on the outcome of the DFB-Pokal (German Cup), which is an annual knockout tournament that occurs throughout the Bundesliga season. If the winner of the DFB-Pokal also finishes within the top six division one positions, then the team ranked in seventh place qualifies for the Europa League. If the DFB-Pokal winner is not within the top six, then they receive the Europa League qualification and the seventh-place team does not qualify.

The bottom two teams from the first division at the end of the season are automatically relegated down into the second division for the next season. The top two teams from the second division replace those teams. The third-last team from the first division then faces a relegation playoff series against the third-place team from the second division, with the victor being awarded the final division one spot for the next season. The same scenario occurs for the final positions of the second division, where teams can be relegated down into the third league.

2.1.2.6. Player recruitment

There are different methods for recruiting players in the German Bundesliga. German-born football players who are highly talented are typically recruited at a young age and signed to professional contracts. This process allows for the Bundesliga teams to develop their skills earlier in preparation for playing at the highest competition level. International
players can also be signed via a transfer or loan, where a number of transfer periods occur throughout a season. If a player is being signed via a transfer, then the Bundesliga team must pay an agreed upon transfer fee to the player’s original club. This transfer fee is not applicable with players who are signed via a loan, where the player will return to their original club after the agreed upon loan period has expired. As there is no salary cap in Bundesliga football, teams are permitted to pay their players whatever amount they desire.

2.1.2.7. Football statistics

Statistics are recorded throughout a football match and are presented in the form of a box-score or match summary. Table 2.3 presents a sample box-score for FC Bayern Munich and includes the following statistics: shots (SH), shots on goal (SG), goals scored (G), goal assists (GA), offsides (OF), fouls drawn (FD), fouls committed (FC), goals saved (SV), yellow cards (YC) and red cards (RC). Advanced statistics, such as the total distance run and the number of tackles won, are collected and published on the official Bundesliga website (www.bundesliga.com).

<table>
<thead>
<tr>
<th>POS</th>
<th>No</th>
<th>Name</th>
<th>SH</th>
<th>SG</th>
<th>G</th>
<th>A</th>
<th>OF</th>
<th>FD</th>
<th>FC</th>
<th>SV</th>
<th>YC</th>
<th>RC</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>1</td>
<td>Manuel Neuer</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>Dante</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>D</td>
<td>13</td>
<td>Rafinha</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>D</td>
<td>17</td>
<td>Jérôme Boateng</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D</td>
<td>18</td>
<td>Juan Bernat</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>M</td>
<td>3</td>
<td>Xabi Alonso</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>M</td>
<td>21</td>
<td>Philipp Lahm</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M</td>
<td>31</td>
<td>Bastian Schweinsteiger</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F</td>
<td>9</td>
<td>Robert Lewandowski</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F</td>
<td>19</td>
<td>Mario Götze</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F</td>
<td>25</td>
<td>Thomas Müller</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2.3: Sample box-score for FC Bayern Munich, 2014/2015
2.2. National Basketball Association

This section provides a brief history of the National Basketball Association (NBA). It includes details about the sport’s origins within North America and its evolution into modern times. The basic rules of basketball and the various match statistics that are collected are also detailed. The information presented within this section was summarised from Steback (2014) and the official NBA History website (www.nba.com/history).

2.2.1. History

The game of basketball was first conceived by Dr. James Naismith in 1891 as a method for keeping high school students active during the winter months. The game originally consisted of a peach basket hanging above a gymnasium with two teams competing to put a ball into it. Soon, a second basket was added, meaning that both teams had to score in their own basket rather than competing to score in just one. The game originally consisted of nine players on the court per side but was soon reduced to just five due to congestion.

2.2.1.1. Pre 1940’s

The new game grew in popularity towards the turn of the century, where it was frequently played by members of the United States Army. Matches were often played in high school gymnasiums or ballrooms as no permanent courts had been established yet. The game eventually grew in popularity within universities, where the first ever match was played in 1893. During this same time, the peach baskets were replaced by metal hoops to allow for the basketball to easily fall to the ground.

While the game rapidly grew in popularity across North America, a single competition had yet to be established. Smaller leagues were also established with mixed results, including the Metropolitan Basketball League and the American Basketball League. The number of
competing teams within each league ranged from six to forty, with teams often dropping out due to a lack of players. Often, teams chose to travel around the country to compete for cash rewards against other teams instead of playing in tournaments for a championship.

2.2.1.2. 1940’s – 1970’s

The 1940’s saw the creation of a new competitive league called the Basketball Association of America (BAA). The new league aimed to establish professional basketball teams in major U.S. cities, with matches to be played at iconic sporting stadiums in order to attract large crowds. The first season commenced in 1946 with eleven teams divided into two divisions, East and West. Each team would play a total of 60 matches with the best three teams from each division advancing to the playoffs to compete for the BAA championship; where the first ever title was won by the Philadelphia Warriors.

At the conclusion of the 1949 season, the BAA and its rival, the National Basketball League (NBL), agreed on a merger to create a single national competition called the National Basketball Association (NBA). The league was to originally consist of seventeen teams but only eight teams featured in the first NBA season due to financial difficulties. The NBA also began playing matches over 48 minutes instead of 40, while also allowing players to commit six personal fouls before disqualified from the match instead of five.

In the 1950’s, the Minneapolis Lakers emerged as the NBA’s first dominant team, winning four of the first five championships. During this time, the league introduced a 24-second shot clock, which meant that teams had 24 seconds per possession to attempt a shot or the ball was turned over to the opposition. The league also introduced an All-Star game, which would be an annual match consisting of the best players in the league. The idea behind the match was to increase crowd numbers and help grow the popularity of the sport.
The start of the 1960’s was dominated by the Boston Celtics, who won eight straight NBA championships from 1959 to 1966. During this time, many teams relocated to different cities while new teams entered the league, including the Chicago Bulls, Seattle Supersonics and Phoenix Suns. This resulted in a total of fourteen NBA teams by the conclusion of 1968. Despite the introduction of new teams, the Boston Celtics won nine of the ten championships during the decade.

In 1967, a rival league called the American Basketball Association (ABA) was formed. The new league established teams in cities that did not already have an NBA team, including Kentucky and Indiana. It also offered alternative rules to the NBA, including a 30-second shot clock and a three-point line. Initially popular due to these alternative rules, the league ran into trouble during the 1970’s due to a lack of television exposure and financial difficulties. In 1976, the NBA and the ABA agreed on a merger which saw four ABA teams join the NBA - including teams from Denver, Indiana, New York and San Antonio. By the mid-1970’s, the NBA had expanded to 22 teams across North America, with the championship being shared amongst several teams during the decade. The NBA also introduced the three-point line, which had initially been popular in the ABA.

**2.2.1.3. 1980’s – Present day**

In the 1980’s, the Los Angeles Lakers and the Boston Celtics established themselves as the two premier basketball teams in the competition, sharing a total of eight championships during the decade. The NBA also continued to expand, with four new teams introduced during the decade. This included teams from Charlotte, Miami, Minnesota and Orlando. Two Canadian teams also joined the league in the 1990’s, with one team being based in Toronto and the other in Vancouver. This marked the first time that the NBA had expanded into a country other than the United States.
The late 1990’s saw a dispute erupt between the team owners and the National Basketball Players Association (NBAPA). The owners of the teams were recording significant financial losses and blamed the matter on player salaries, which was paid using a percentage of the basketball-related income. Both parties disagreed on what value this percentage should be and both failed to sign a new Collective Bargaining Agreement (CBA), causing a lockout prior to the 1998/99 season. A new CBA was eventually agreed on and the season commenced with teams playing only 50 games each instead of the regular 82.

At the turn of the century, the NBA consisted of 30 teams split evenly into two conferences, the East and the West. A number of teams also relocated during this time, including Seattle moving to Oklahoma City and the New Jersey Nets moving to Brooklyn. The team located in Charlotte also relocated to New Orleans, while the final team to enter the league became the new Charlotte team. The league also experienced another NBA lockout in 2011 due to another dispute about the CBA between players and owners. The lockout resulted in the 2011/12 season consisting of each team playing just 66 matches.

The NBA championship during the 2000’s was shared amongst teams, with the Los Angeles Lakers securing five titles during this time. Former ABA and new expansion teams also secured their first ever NBA championship, including the San Antonio Spurs, the Dallas Mavericks and the Miami Heat. Prior to the commencement of the 2017/2018 season, the Boston Celtics have won the most NBA championships with 17, while the next best if the Los Angeles Lakers with 16 championships.
2.2.2. The game

This section details the current format of the NBA. Topics of interest include the current teams competing in the league, the scoring system, rules and objectives, the playing court and positions, and the structure of the league standings.

![NBA logo](image)

Figure 2.5: NBA logo

2.2.2.1. Court and playing positions

Basketball is played on a rectangular court made of wooden floorboards. The court consists of lines that represent the playing boundary, where foul shots occur, and where a three point shot can be attempted. The dimensions of an NBA court are presented in Figure 2.6.

Each team is allowed a total of five players on the court at any given time. The five players are typically broken down into five positions: point guard, shooting guard, small forward, power forward, and center. Each NBA team can have a total of eight players on the bench, with an additional two players on the ‘inactive list’. This results in each team having a
maximum of 15 players on their roster. Players can rotate on and off the court whenever the

clock has stopped, and there are no restrictions on the number of rotations a team can make.

![Image of NBA basketball court]

Source: www.sportscourtdimensions.com/basketball/

Figure 2.6: Dimensions of an NBA basketball court

2.2.2.2. Rules and objectives

An NBA match is played over the course of 48 minutes, which are split evenly into four

quarters. The primary objective is to outscore the opposition by the end of the 48 minutes.

Points are scored by shooting the basketball into the metal hoops located at either end of the
court. All shots attempted inside the arc are worth two points while all shots attempted
outside the arc are worth three. If the scores are tied at the end of the 48 minutes then an
additional five-minute (overtime) period is played. Additional overtime periods can be played
until a single victor is determined.
During a match, a player can only move with the ball if they are bouncing (dribbling) the ball. When a team is in possession of the ball, they must attempt a shot on goal within 24 seconds or the ball is turned over to their opponent. The main game clock stops every time the ball goes out-of-bounds, or when a foul is called. Players are given a foul if they commit an illegal offence during a match, such as holding or bumping another player. Excessive physical contact or unsportsmanlike behaviour can result in a technical foul, where a player is ejected from the match if they receive two such fouls within the same match. A foul that occurs when a player is attempting a shot on goal results in the player receiving free throw shots, which are taken from the free throw line located directly in front of the team’s goals. Each free throw is worth one point. The standard ball used during an NBA match is presented in Figure 2.7.

![Figure 2.7: NBA playing ball](image)

### 2.2.2.3. Teams

As of 2018, the NBA consists of 29 teams from the United States of America and one team from Canada. The 30 teams are split evenly into two conferences, the East and the West. The teams are presented in Table 2.4, along with the year in which they debuted in the NBA and the number of NBA championships won at the completion of the 2016/17 season.
<table>
<thead>
<tr>
<th>Logo</th>
<th>Team Name</th>
<th>Location</th>
<th>NBA Debut</th>
<th>NBA Titles</th>
<th>Logo</th>
<th>Team Name</th>
<th>Location</th>
<th>NBA Debut</th>
<th>NBA Titles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Atlanta Hawks</td>
<td>Atlanta, GA</td>
<td>1946</td>
<td>1</td>
<td></td>
<td>Dallas Mavericks</td>
<td>Dallas, TX</td>
<td>1980</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Boston Celtics</td>
<td>Boston, MA</td>
<td>1946</td>
<td>17</td>
<td></td>
<td>Denver Nuggets</td>
<td>Denver, CO</td>
<td>1976</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Brooklyn Nets</td>
<td>Brooklyn, NY</td>
<td>1976</td>
<td>0</td>
<td></td>
<td>Golden State Warriors</td>
<td>Oakland, CA</td>
<td>1946</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Charlotte Hornets</td>
<td>Charlotte, NC</td>
<td>1988</td>
<td>0</td>
<td></td>
<td>Houston Rockets</td>
<td>Houston, TX</td>
<td>1967</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Chicago Bulls</td>
<td>Chicago, IL</td>
<td>1966</td>
<td>6</td>
<td></td>
<td>Los Angeles Clippers</td>
<td>Los Angeles, CA</td>
<td>1970</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Cleveland Cavaliers</td>
<td>Cleveland, OH</td>
<td>1970</td>
<td>1</td>
<td></td>
<td>Los Angeles Lakers</td>
<td>Los Angeles, CA</td>
<td>1947</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Detroit Pistons</td>
<td>Detroit, MI</td>
<td>1941</td>
<td>3</td>
<td></td>
<td>Memphis Grizzlies</td>
<td>Memphis, TN</td>
<td>1995</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Indiana Pacers</td>
<td>Indianapolis, IN</td>
<td>1976</td>
<td>0</td>
<td></td>
<td>Minnesota Timberwolves</td>
<td>Minneapolis, MS</td>
<td>1989</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Miami Heat</td>
<td>Miami, FL</td>
<td>1988</td>
<td>3</td>
<td></td>
<td>New Orleans Pelicans</td>
<td>New Orleans, LA</td>
<td>2002</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Milwaukee Bucks</td>
<td>Milwaukee, WI</td>
<td>1968</td>
<td>1</td>
<td></td>
<td>Oklahoma City Thunder</td>
<td>Oklahoma City, OK</td>
<td>1967</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>New York Knicks</td>
<td>New York City, NY</td>
<td>1946</td>
<td>2</td>
<td></td>
<td>Phoenix Suns</td>
<td>Phoenix, AZ</td>
<td>1968</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Orlando Magic</td>
<td>Orlando, FL</td>
<td>1989</td>
<td>0</td>
<td></td>
<td>Portland Trailblazers</td>
<td>Portland, OR</td>
<td>1970</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Philadelphia 76ers</td>
<td>Philadelphia, PA</td>
<td>1946</td>
<td>3</td>
<td></td>
<td>Sacramento Kings</td>
<td>Sacramento, CA</td>
<td>1945</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Toronto Raptors</td>
<td>Toronto, Canada</td>
<td>1995</td>
<td>0</td>
<td></td>
<td>San Antonio Spurs</td>
<td>San Antonio, TX</td>
<td>1976</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Washington Wizards</td>
<td>Washington D.C.</td>
<td>1961</td>
<td>1</td>
<td></td>
<td>Utah Jazz</td>
<td>Salt Lake City, UT</td>
<td>1974</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2.4: List of NBA teams
2.2.2.4. Fixtures

As mentioned, the 30 NBA teams are split evenly into two conferences, the East and the West. Each conference is then split into three divisions consisting of five teams each. During an NBA season, each team will play a total of 82 matches, with 16 of these matches being played against teams in their division. Teams will also play two matches against each team from the opposing conference, with the remaining matches being played against teams from their own conference. There is also no round structure within the NBA, where matches are scheduled throughout a week on any given day. An NBA season typically runs from late October through to the completion of the NBA Finals in June the following year.

2.2.2.5. NBA standings

The league standings are generated for both the Eastern and Western conferences after the completion of each day of play. Standings are also created for each division as a method for determining tie-breakers within the main conference standings. The order of the conference standings is primarily determined by the win percentage of each team up until that day of the season. If two or more teams have the same win percentage, then the following tie-breaker criteria is assessed: division leader being ranked higher than non-division leader, the head-to-head win percentage between the teams, the win percentage against teams within their own conference, and win percentage against teams within their own division. Up until the start of the 2015/16 season, the three division leaders were guaranteed a top four position regardless of their win percentage. An example of the league standings showing this rule is presented in Figure 2.8.
Figure 2.8: Final NBA standings for the 2014/15 season

The top eight teams from each conference at the end of the season progress to the NBA Playoffs, where they compete in a knockout tournament to decide the NBA champion. There are four rounds of the playoffs, with the first three rounds being played between teams of the same conference. Each round of the playoffs requires a team to win a best-of-seven match series against the opposing team, where the loser is knocked out of the tournament. Teams
can gain a home court advantage in each round of the playoffs by finishing higher in the standings. For example, finishing in any of the top four positions guarantees a team at least one round with a home court advantage, while finishing in the top two guarantees a team at least two rounds of home advantage. Finishing first in a conference guarantees a team home advantage through the first three rounds, where the home advantage within the final round, called the NBA Finals, is decided by the best overall win percentage across the two remaining teams.

2.2.2.6. **Player recruitment**

In the NBA, there are numerous ways in which a player can be recruited to a team. The first method is through the NBA draft, which is held each year after the conclusion of the season. The draft allows the worst performed teams from the previous season the first opportunity to select the best young players from the U.S. college system. The draft consists of two rounds with each team selecting twice. Players chosen in the first round receive an automatic two year contract with the team that selects them, while players taken in the second round do not receive a guarantee of a playing contract.

Players in the NBA can also be traded throughout the season or during the off-season. In order for a trade to be accepted, the trade amount for both parties must be relatively equal. For example, a player worth $500,000 per season cannot be traded for a player worth $20 million per season, unless the team receiving the latter contract has the salary space. Each team has a salary cap on the amount they can pay their 15 players, but are allowed to go over this mark if they pay a luxury tax. Teams can also release (waive) players whenever they wish, but their contract amount will still count towards the team’s salary cap unless another team signs the released player.
The NBA also has a free agency rule, which allows players to freely join a team without requiring a trade. There are two types of free agents: restricted and unrestricted. Restricted free agents mean that they can sign with any team but their original team has the option to match the offer and retain them for the length of the contract. Unrestricted free agents, however, can freely move to whichever team they wish, with their original team being unable to stop their move.

2.2.2.7. NBA statistics

The NBA collects various statistics during each match, which are presented in the form of a box-score at the conclusion of the match. The statistics collected include: field goals made/attempted (FGM/FGA), three point field goals made/attempted (3PM/3PA), offensive rebounds (OFF), defensive rebounds (DEF), total rebounds (TOT), assists (AST), personal fouls (PF), steals (ST), turnovers (TO), block shots (BS), blocks against (BA), and the total points scored by the player (PTS). An example of a box-score from the official NBA website (nba.com) is presented in Figure 2.9.

![Figure 2.9: Sample NBA box-score](image-url)
Chapter 3

Methods

This chapter details the statistical methods and concepts that appear throughout this dissertation. It includes details on the collection and manipulation of the data that was completed for both Bundesliga football and NBA basketball; information on featured probability distributions; and explanations on statistical modelling techniques that are applied within this dissertation. Key references for further information are also included where applicable.
3.1. **Data**

This section details the collection and manipulation of the sports data that is assessed throughout this dissertation.

### 3.1.1. Bundesliga

For Bundesliga football, the primary data for the upcoming match importance calculations consists of the results of matches played between 2006 and 2015. The match results were collected from [www.football-data.co.uk](http://www.football-data.co.uk) for both the first and second divisions, totalling nine seasons per division. All seasons contained the complete 306 matches, where the 18 teams play each other twice over 34 rounds, for a total of 2,754 matches per division. This period of matches was selected as it aligns with the commencement of this dissertation, and it provides a sufficiently large sample size to explore the upcoming match importance calculations in Bundesliga football. Note that non-league matches, such as Champions League or DFB-Pokal matches, are not included as the focus of this dissertation is the league matches.

The secondary data consists of match results from the 2004/2005 and 2005/2006 seasons for both the first and second divisions. The secondary data will be used to optimise the model parameters for the Elo ratings systems which will be assessed in Chapter 7. While a greater number of seasons could be used to optimise the parameters, this two-year period is sufficient as it still contains a large sample of Bundesliga matches to configure the Elo ratings. The additional seasons each consist of the completed 306 matches, where the 18 teams play each other twice over 34 rounds, totalling 612 matches per division.

The match results for the Bundesliga were uploaded into a Microsoft Excel spread sheet and divided by season and division. The results for each team after each round of the season
was then extracted and entered into a results matrix, where each result for a team was represented by the number of match points earned in the match (3 = win, 1 = draw, 0 = loss). The total number of goals scored within each match was also entered into a separate matrix, which allowed for the total goal difference and total goals scored after each round to be calculated. The league standings after each round of the season were then reconstructed using the total match points accumulated and the tie-breaker criteria outlined in Chapter 2. A screenshot of the data from the 2008/2009 Bundesliga season is presented in Figure 3.1.

Figure 3.1: Screenshot of 2008/2009 Bundesliga data in Excel spreadsheet

3.1.2. NBA

For NBA basketball, the primary data for the upcoming match importance calculations consists of the results of matches played between 2005 and 2015. The match results for the ten seasons played during this period of time were collected from the publically available website www.basketball-reference.com. Eight of the seasons consisted of the complete 1,230
matches, where all 30 teams play 82 matches, while the 2012/2013 season comprised of only 1,229 matches due to one cancellation. The final season (2011/2012) contained only 990 matches due to the 2011 NBA lockout (see Chapter 2). This period of matches was selected as it aligns with the commencement of this dissertation, and it provides a sufficient sample size to explore the upcoming match importance calculations in NBA basketball.

The secondary data consists of additional match results from the 2003/2004 and 2004/2005 seasons. This data was collected to assist with optimisation of model parameters for the Elo ratings systems presented in Chapter 7. Like Bundesliga football, this two-year period was selected as it provides a sufficient sample size to configure the Elo ratings in the NBA. One of the seasons contained the complete 1,230 matches while the other contained only 1,189 matches due to only 29 teams competing in the league at the time. The 2011/2012 NBA season is also included within the secondary data set due to its shortened nature, resulting in a total of 11,069 matches for the primary data and 3,409 for the secondary data. The average length of the season within the primary data was approximately 163 days, while the average length for the secondary data was 164 days.

The results for each season were uploaded into a Microsoft Excel spreadsheet where they were extracted into matrices that represent the total number of wins for each team after the completion of each day of play. These matrices also allowed the total number of matches played up until a certain date to be determined for each team. The conference standings were then reconstructed after the completion of each day using the NBA tie-breaker rules outlined in Chapter 2. A screenshot of the data from the 2005/2006 NBA season is presented in Figure 3.2.
3.2. **VBA programming**

Visual Basic for Applications (VBA) is the in-built programming language of Microsoft Excel. A user can write a VBA program to automate processes within Microsoft Excel that would usually require a great deal of runtime, such as advanced mathematical calculations or generating analysis reports. In this dissertation, multiple VBA programs were built to automate the statistical calculations that are required to complete the research. The details of the content pertaining to each program are provided within the relevant chapters. For further information on VBA coding, see Walkenbach (2010).

3.3. **Binomial distribution**

The binomial distribution is the probability distribution for the number of successes \( x \) observed in a specific amount of independent trials \( n \); where each experiment results in either a success of failure (Hogg & Tanis, 2010). Each experiment is called a Bernoulli trial,
where the probability of success \((p)\) is held constant between trials. Therefore, a binomial distribution can be viewed as a sequence of Bernoulli trials.

To determine the probability of observing exactly \(x\) number of successes, the probability mass function (p.m.f) of the binomial distribution is applied. This is presented in (3.1).

\[
b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \ x = 0, 1, 2, \ldots n\quad (3.1)
\]

To determine the probability that a continuous random variable \((X)\) takes a value less than or equal to \(x\), the cumulative distribution function (c.d.f) of the binomial distribution is applied. This is presented in (3.2) and is applied throughout this dissertation.

\[
B(x; n, p) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i}, x = 0, 1, 2, \ldots n\quad (3.2)
\]

### 3.4. Markov Chain

A Markov Chain is a stochastic model that can make predictions about the future of a process based on the information contained within the current state \(X_n\) \((n = \{0, 1, 2\ldots\})\). The conditional probability of some future state \(X_{n+1}\) is independent on past states \(X_0, X_1, \ldots, X_{n-1}\) and only dependents on the current state (Ross, 2014). The probability that a process will transition from some state \(i\) to state \(j\) is denoted by \(P_{ij}\):

\[
P_{ij} = P\{X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\}\quad (3.3)
\]

Where,

\[
\sum_{j=0}^{\infty} P_{ij} = 1, i = 0, 1, \ldots
\]
3.5. Monte Carlo simulation

Monte Carlo simulation is the process of drawing samples from a specified probability distribution to approximate expectations (Gilks, Richardson, & Spiegelhalter, 1995). Completed using a computer algorithm, the procedure is typically applied to forecast the final result of a mathematical model. In sports modelling, Monte Carlo simulation is frequently applied to forecast the final order of teams in the standings. For further information, see Gilks et al. (1995) and Modica and Poggiolini (2012).

3.6. Elo ratings system

The Elo ratings system is a statistical model that was originally developed to assess the strength of chess players (Elo, 1978). However, in recent literature the model has been applied to professional sports to rate the overall strength of teams, including Australian Rules football (Ryall & Bedford, 2010) and European football (Leitner, Zeileis, & Hornik, 2010). FIFA also applies the Elo ratings to officially rank women’s football teams, where an adjustment to include a home advantage is included (Fifa.com, 2018). The team rating for both the home and away team is based on the previous results up until the match of interest. The Elo rating for the home team ($H$) takes the following form:

$$R_t^H = R_{t-1}^H + k(O^H - E^H) \quad (3.4)$$

Where,

$R_t^H$ = the new rating for the home team after time $t$

$R_{t-1}^H$ = the old rating for the home team at time $t-1$
When \( t = 0 \), the initial rating for a team is typically set to 1,500. The observed result for the home team \((O^H)\) is the end result of the match of interest, which, for football, is presented in (3.5).

\[
O^H = \begin{cases} 
1, & \text{if home team wins} \\
0.5, & \text{if match drawn} \\
0, & \text{otherwise}
\end{cases} \tag{3.5}
\]

The expected result for the home team \((E^H)\) is the score that the team should achieve in the match of interest. This is calculated using the logistic curve presented in (3.6).

\[
E^H = \frac{1}{1 + c \left( \frac{R_t - 1 - R_t^H - h}{d} \right)} \tag{3.6}
\]

There are four model parameters within the Elo ratings system that require a base setting. The parameter \( c \) must be 10 to follow Elo (1978)’s derivation. According to Elo (1978), \( d \) is typically chosen to be 400, forcing the team ratings to become distributed with a standard deviation of approximately 200. The next parameter, \( k \), determines the change in the ratings, which requires careful consideration since extreme values will cause the ratings to move too quickly. The final parameter, \( h \), reflects the advantage gained by the home team.

The parameters \( k \) and \( h \) can be determined through optimisation of test data prior to the model’s application to future matches. In this dissertation, maximum log-likelihood is applied to optimise the results of matches contained within the secondary data (see Section 3.1), with the procedure presented in (3.7) (Devore, 2008).

\[
\sum_{i=1}^{n} O^H \left( \log E^H \right) + (1 - O^H) \left( \log(1 - E^H) \right) \tag{3.7}
\]

### 3.7. Linear regression

Regression analysis is the statistical process for investigating the relationship between a dependent (response) variable, \( y \), and an independent (predictor/explanatory) variable, \( x \).
When there are multiple independent variables, and the relationship is linear, the process is referred to as a multiple linear regression analysis. The standard model for \( k \) independent variables is presented in (3.8). Note that \( \epsilon \) is the error term that is assumed to be normally distributed. The term signifies the amount in which the model may differ.

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon
\]  

(3.8)

The coefficients for the various independent variables are typically estimated using the least squares technique, which seeks to minimise the sum of the squares of the residuals. The coefficient of determination, \( R^2 \), is assessed to determine the suitability of the model, with values close to 1 indicating a strong correlation between the independent variables and the dependent variable. To determine which of the independent variables is significant in determining the dependent variable, a model utility \( F \) test is applied. If the probability value (\( p \)-value) for a specific independent variable is less than the significance level (\( \alpha \)), then the variable is a significant predictor of the dependent variable.

When assessing a linear model with a large number of independent variables, it is sometimes best to reduce the number of variables to only those that are significant. This can be completed using stepwise regression, where variables are chosen either using forward selection, backward elimination, or a combination of the two procedures. Forward selection enters variables one-by-one into the model based on their \( p \)-value, while backward elimination removes non-significant variables from the complete model until an optimised model is established. For further information, see (Devore, 2008).
Chapter 4

ParWins importance

In this chapter, a method for quantifying the importance of a match with respect to achieving different end-of-season outcomes is presented. This method is an extension of an existing ‘projected wins’ method and is applied to both a two-result (NBA basketball) and a three-result (Bundesliga football) sport. The chapter is broken down into the following sections: Section 4.1 provides a brief introduction on measuring match importance. Section 4.2 details the methodology for applying the measure to both NBA basketball and Bundesliga football. Section 4.3 provides an evaluation of how the importance for each position changes throughout a season, while Section 4.4 provides a critical discussion of the findings, including drawbacks of applying the measure and future improvements. Finally, Section 4.5 concludes and summarises the chapter.
4.1. Introduction

As discussed extensively in Chapter 1, numerous methods exist for quantifying the importance of a match in professional sports. Measures have ranged from a ‘percentage behind’ model for Major League Baseball (Lei & Humphreys, 2013) to a probabilistic approach for a best-of-seven match series in NBA basketball (Schilling, 1994). The latter defines match importance as the difference between success probabilities conditional on the outcome of the next match in the series. This definition is the most commonly accepted in the literature (Goossens & Belien, 2010), and has been applied within other professional sports, including Australian Rules football (Schembri & Bedford, 2010).

While the probabilistic definition has been applied within other professional sports, an extensive evaluation of match importance for different end-of-season outcomes over the duration of a season has not been completed. Furthermore, the definition’s application within both NBA basketball and Bundesliga football has not been assessed. To complete this, the probabilistic approach conceived by Bedford and Schembri (2006) will be applied within both sports. The method follows the work completed by Morris (1977) and Schilling (1994) by defining the importance of a match in Australian Rules football as the difference between two conditional probabilities: the probability that a team achieves its target position (i.e. finish within the top eight) given they win their next match, minus the probability that a team achieves its target position given they lose their next match. A cumulative binomial distribution function is applied to calculate the probabilities with respect to a team achieving a projected wins total.

As described within Chapter 2, the top eight teams from each NBA conference progress to the playoffs. Each position within the top eight awards a team a different end-of-season outcome, where the top four positions result in at least an initial home advantage during the
playoffs. Similar end-of-season outcomes also exist within first division Bundesliga football, where the top seven positions of the standings result in qualification to either the UEFA Champions League or Europa League. In second division Bundesliga, the top two positions in the standings result in automatic promotion to the first division, replacing the bottom two teams from division one who are relegated. Since the number of end-of-season outcomes differs between the two Bundesliga divisions, both will be included within this chapter.

To account for the different end-of-season outcomes within both sports, the probabilistic measure conceived by Bedford and Schembri (2006) will be modified. The importance of finishing within each position in the standings for both sports will be calculated, instead of solely focusing on finishing within the top eight. The methodology for applying the measure within NBA basketball is primarily detailed in the next section of this chapter, while adjustments to apply the method within Bundesliga football are also briefly described.
4.2. Methods

As mentioned, the importance of a match in NBA basketball is calculated by extending the probabilistic measure conceived by Bedford and Schembri (2006). The approach first determines a projected wins requirement for a team to finish within the top eight, which is based on a team’s total wins after game $g$. The probability that the team achieves the projected wins requirement, conditional on whether or not the team wins their next match, is calculated using a cumulative binomial distribution function. The difference between these probabilities then equate to the importance of the next match.

4.2.1. Projected wins

While the original work by Bedford and Schembri (2006) solely focused on finishing in eighth position, the measure is extended in this chapter to include other end-of-season outcomes in NBA basketball. In order to calculate the probability of team $i$ finishing in position $s$ after the completion of game $g$, a total wins requirement for each position needs to be determined. This is completed by multiplying the total wins ($TW$) of the team in position $s$ after game $g$ by the total number of matches played in the season (82), and dividing the result by the total number of games played by the team in position $s$ ($gs$). For $s = \{1, \ldots, 8\}$ and $g = \{1, \ldots, 81\}$, the projected wins, denoted by $ProjWins$, is presented in (4.1). Note that the $ProjWins$ are calculated after the completion of game $g$. This means that no calculations are completed after game 82 since this is the end of the season, and no calculations are completed prior to a team’s first match as no pre-match information is available.

$$ProjWins_s(g_s) = \left[ \frac{TW_s(g_s) \cdot 82}{gs} \right]$$

(4.1)

The wins requirement for team $i$ to finish in position $s$ is calculated by assessing the difference between the $ProjWins$ and the total wins of team $i$ after game $g$. However, initial
calculations observed that (4.1) is prone to over-estimating the projected wins for the team in position $s$. An example of this is presented in Figure 4.1, which compares the projected wins to the actual end-of-season wins for the first-place team from the Eastern conference of the 2005/2006 NBA season. In this example, the projected wins after each day of the season is greater than the actual total wins achieved by the first-place team (note that due to the NBA not having a round structure, the importance calculations are completed after the completion of each day of the season). Since the projected wins are critical when calculating the importance of a match using this approach, an alternative method is trialled here.
An alternative projected wins for the team in position $s$ is determined by forward projecting the maximum number of wins the team in position $s+1$ can achieve, which would ensure that the team in position $s$ maintains their spot. This is calculated through a summation of the total wins of the team in position $s+1$ and their number of games remaining in the season. This alternative projected wins will decrease as the season progresses and will at some point fall below the actual number of wins of the team in position $s$. Therefore, a $max$
function for the total wins of position $s$ and the maximum number of wins for position $s+1$ is applied. This is presented in (4.2).

\[
MaxWins_s(g_s) = \max(TW_s(g_s), TW_{s+1}(g_{s+1}) + (82 - g_{s+1}))
\]  

(4.2)

Figure 4.2: Comparison between $MaxWins$ function and its components from the 2005/2006 NBA Eastern conference
Figure 4.2 provides a graphical comparison of the components of (4.2) for positions one, three, five and seven of the NBA Eastern conference from the 2005/2006 season. Note that, again, the values are calculated after each day of the season since the NBA does not have a set round structure (see Chapter 2). As observed in Figure 4.2, the \( \text{MaxWins} \) function closely follows the position \( s+1 \) component for the majority of the season before switching to follow the position \( s \) component. The result is most noticeable for position one, while the remaining positions do not appear to be affected.

Figure 4.3 provides a comparison of (4.1) and (4.2) for the same positions from the NBA Eastern conference for the 2005/2006 season. For most positions, the \( \text{MaxWins} \) function provides a higher projected wins than the original \( \text{ProjWins} \) function. However, this does not occur within position one, where the \( \text{MaxWins} \) function becomes lower than the \( \text{ProjWins} \) function halfway through the season. Since position one is arguably the most desired position in the standings, as this provides the maximum amount of home court advantage, a minimum of the \( \text{MaxWins} \) and the \( \text{ProjWins} \) will be applied across all positions. The adjusted projected wins, denoted by \( \text{AdjProjWins} \), is presented in (4.3).

\[
\text{AdjProjWins}_s(g_s) = \min(\text{MaxWins}_s(g_s), \text{ProjWins}_s(g_s))
\]  

(4.3)

To calculate the required wins for team \( i \) to finish in position \( s \) after game \( g \), the difference between the \( \text{AdjProjWins} \) and the total wins of team \( i \) is assessed. If the result is a negative value, then the team has already surpassed position \( s \) in the standings, where a value of zero is applied instead. The final result is called the \( \text{ParWins} \) and is presented in (4.4).

\[
\text{ParWins}_{i,s}(g_i) = \max(\lfloor \text{AdjProjWins}_s(g_s) - TW_i(g_i) \rfloor, 0)
\]  

(4.4)
4.2.2. Conditional probabilities

To calculate the conditional probabilities required for the match importance definition, the probability of a team finishing in position $s$ is first required. This is calculated by applying a cumulative binomial distribution, where the number of successes is equal to the ParWins; the number of trials is equal to the number of matches remaining in the season for ParWins; the number of matches remaining in the season for
team $i$; and the probability of success is equal to 0.5 for all teams. Although the use of a constant success probability across all teams ignores critical factors such as team strength or home court advantage, it allows for a baseline model to be established. The probability of team $i$ finishing in position $s$ after the completion of game $g$ is calculated by using the formula presented in (4.5).

$$P_i(s|g_i) = 1_{(a)} + 1_{(b)}[1 - B(ParWins_i,s(g_i) - 1; 82 - g_i, 0.5)]$$ \hspace{1cm} (4.5)

Where,

$$a = (ParWins_{i,s}(g_i) = 0) \vee (Rank_i(g_i) < s)$$

$$b = (ParWins_{i,s}(g_i) > 0) \land (ParWins_{i,s}(g_i) \leq (82 - g_i))$$

(4.5) includes two indicator functions, which take the value one if condition $a$ or $b$ is true, and zero if false. Condition $a$ states that the probability is equal to one if either the $ParWins$ is equal to zero, or if team $i$ is ranked above position $s$. If both of these scenarios are false then condition $b$ is considered. This states that the cumulative binomial distribution is applied if the $ParWins$ is non-zero and less than or equal to the number of games remaining for team $i$.

To calculate the conditional probabilities required for the match importance, definition (4.5) is applied with some minor adjustments. The probability of team $i$ finishing in position $s$, given they win or lose their next match, is presented in (4.6) and (4.7), respectively.

$$P_i(s|W g_i + 1) = 1_{(c)} + 1_{(d)}[1 - B(ParWins_{i,s}(g_i) - 2; 82 - (g_i + 1), 0.5)]$$ \hspace{1cm} (4.6)

$$P_i(s|L g_i + 1) = 1_{(e)} + 1_{(d)}[1 - B(ParWins_{i,s}(g_i) - 1; 82 - (g_i + 1), 0.5)]$$ \hspace{1cm} (4.7)

Where,
\[ c = (\text{ParWins}_{i,s}(g_i) \leq 1) \lor (\text{Rank}_i(g_i) < s) \]

\[ d = (\text{ParWins}_{i,s}(g_i) > 1) \land (P_i(s|g_i) > 0) \land (\text{ParWins}_{i,s}(g_i) \leq (82 - (g_i + 1))) \]

\[ e = (\text{ParWins}_{i,s}(g_i) = 0) \lor (\text{Rank}_i(g_i) < s) \]

Condition \( c \) states that (4.6) will equal one if either the \( \text{ParWins} \) is less than or equal to one, or if team \( i \) is ranked above position \( s \) in the standings. If both of these scenarios are false then condition \( d \) is considered. This states that the cumulative binomial distribution is applied if the \( \text{ParWins} \) is greater than one, but less than the number of games remaining in the season for team \( i \); and if the probability that team \( i \) finishes in position \( s \) is greater than zero. Condition \( e \) is similar to condition \( c \), with the only change being that the \( \text{ParWins} \) must be equal to zero.

4.2.3. Match importance

The definition originally conceived by Schilling (1994) is applied to calculate the importance of a match for team \( i \) to finish in position \( s \). Since there are eight key positions within the NBA, eight match importance values can be calculated. These values will be called the positional importance, and are calculated by taking the difference of (4.6) and (4.7). This is presented in (4.8). The importance value for finishing in position \( s \) is measured on a 0-1 scale, with values closer to 1 signifying higher match importance.

\[ \text{IMP}_{i,s}(g_i + 1) = P_i(\text{Pos } s|W (g_i + 1)) - P_i(\text{Pos } s|L (g_i + 1)) \] (4.8)

While this approach allows for the match importance to be calculated with respect to achieving each individual position, it may be advantageous to create a single value which reflects the overall importance of the match to a team. One option is to apply a weighting to each of the positional importance values, where the weighting values reflect either the
hierarchy of the positions (i.e. position one being the highest and thus the best), or their proximity to the current position of team \( i \). A second option is to simply take a summation of all positional importance values without any weighting attached. Since these options are new to the literature, all three will be included within this dissertation.

The first approach, called the Outcome Sum (\( OS \)) determines a weight for each positional importance value according to the hierarchy of the positions within the standings, where position one receives a weighting of one, position two has a weight of \( \frac{1}{2} \), position three has a weight of \( \frac{1}{3} \), etc. This approach assumes that a team would desire to finish in the highest position possible, regardless of their current position in the standings. This is presented in (4.9).

\[
OS_i(g_i + 1) = \sum_{s=1}^{8} \frac{1}{s} IMP_{i,s}(g_i + 1)
\]  

(4.9)

The second approach, called the Positional Sum (\( PS \)), determines a weight for each positional importance value according to their proximity to team \( i \) in the standings. A team’s current position receives the full weighting, while the position directly above them receives a weighting of \( \frac{1}{2} \). The position two spots above team \( i \) receives a weighting of \( \frac{1}{3} \), while the weighting for the remaining positions follow a similar pattern. This approach assumes that a team would desire to move up in the standings while maintaining their current position. This is presented in (4.10).

\[
PS_i(g_i + 1) = \sum_{s=1}^{8} \frac{1}{|\text{Rank}_i(g_i) - s| + 1} IMP_{i,s}(g_i + 1)
\]  

(4.10)

The final approach, called the Total Sum (\( TS \)), takes a summation of the positional importance values without any weighting considered. This is presented in (4.11).

\[
TS_i(g_i + 1) = \sum_{s=1}^{8} IMP_{i,s}(g_i + 1)
\]  

(4.11)
While these three approaches allow for a single importance value to be calculated, it is also possible to sum together all the positional importance values that result in the same end-of-season outcome, resulting in more than one value to summarise the positional importance. For example, positions five through eight can be summed together as they all result in qualification for the NBA playoffs. The same can be completed for the top four positions, as they result in at least one round of home court advantage in the playoffs. These additional approaches are explored in the latter chapters of this dissertation, along with the three weighted-sum approaches.

To provide an understanding of the positional importance calculations, an example from the 2009/2010 NBA season will be assessed. The Eastern conference standings after the completion of play on the 20th of March, 2010 are presented in Table 4.1.

<table>
<thead>
<tr>
<th>Position</th>
<th>Team</th>
<th>Played</th>
<th>Wins</th>
<th>AdjProjWins</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CLE</td>
<td>70</td>
<td>55</td>
<td>61</td>
</tr>
<tr>
<td>2</td>
<td>ORL</td>
<td>70</td>
<td>49</td>
<td>57</td>
</tr>
<tr>
<td>3</td>
<td>BOS</td>
<td>69</td>
<td>45</td>
<td>53</td>
</tr>
<tr>
<td>4</td>
<td>ATL</td>
<td>68</td>
<td>44</td>
<td>52</td>
</tr>
<tr>
<td>5</td>
<td>MIL</td>
<td>68</td>
<td>38</td>
<td>46</td>
</tr>
<tr>
<td>6</td>
<td>MIA</td>
<td>70</td>
<td>36</td>
<td>42</td>
</tr>
<tr>
<td>7</td>
<td>CHA</td>
<td>69</td>
<td>35</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>TOR</td>
<td>68</td>
<td>34</td>
<td>41</td>
</tr>
<tr>
<td>9</td>
<td>CHI</td>
<td>69</td>
<td>32</td>
<td>38</td>
</tr>
<tr>
<td>10</td>
<td>NYK</td>
<td>69</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>11</td>
<td>PHI</td>
<td>70</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>12</td>
<td>DET</td>
<td>69</td>
<td>23</td>
<td>27</td>
</tr>
<tr>
<td>13</td>
<td>IND</td>
<td>69</td>
<td>23</td>
<td>27</td>
</tr>
<tr>
<td>14</td>
<td>WAS</td>
<td>67</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>15</td>
<td>NJN</td>
<td>69</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 4.1: NBA Eastern conference standings after 20th March, 2010
The team in focus is the fourth-place team, the Atlanta Hawks, who have won 44 of their 68 matches played. The adjusted projected end-of-season wins total (\(\text{AdjProjWins}\)) for each position is calculated using (4.3) and is presented in Table 4.1. The \(\text{ParWins}\) for Atlanta to finish in each position is calculated using (4.4), while the probability that they finish in each position, conditional on the outcome of their next match, is calculated using (4.5) – (4.7). The results are presented in Table 4.2.

<table>
<thead>
<tr>
<th>Position</th>
<th>(\text{ParWins})</th>
<th>(P(s))</th>
<th>(P(s \mid W_{g+1}))</th>
<th>(P(s \mid L_{g+1}))</th>
<th>(\text{IMP})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>0.0009</td>
<td>0.0017</td>
<td>0.0001</td>
<td>0.0016</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>0.2120</td>
<td>0.2905</td>
<td>0.1334</td>
<td>0.1571</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>0.3953</td>
<td>0.5000</td>
<td>0.2905</td>
<td>0.2095</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 4.2: Positional importance calculation for Atlanta after game 68

From Table 4.2, Atlanta requires a further 17 wins to finish in position one. However, this is not possible as they have only 14 games remaining in the season. Therefore, conditions \(a\) and \(b\) of (4.5) are false, and the probability that they finish in first position is equal to zero. This results in the positional importance of first position equal to zero. The remaining positions all result in a positional importance greater than zero, with the team’s current position, fourth, recording the highest value. Table 4.3 provides an example of the three weighted-sum approaches, where the \(OS\), \(PS\) and \(TS\) are equal to 0.1055, 0.2886 and 0.3682, respectively.
Table 4.3: Example of weighted-sum match importance approaches

<table>
<thead>
<tr>
<th>Position</th>
<th>OS</th>
<th>PS</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.0008</td>
<td>0.0005</td>
<td>0.0016</td>
</tr>
<tr>
<td>3</td>
<td>0.0524</td>
<td>0.0786</td>
<td>0.1571</td>
</tr>
<tr>
<td>4</td>
<td>0.0524</td>
<td>0.2095</td>
<td>0.2095</td>
</tr>
<tr>
<td>5</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>7</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>0.1055</td>
<td>0.2886</td>
<td>0.3682</td>
</tr>
</tbody>
</table>

4.2.4. Adaption to three-result sport

The adaption of the model within Bundesliga football requires some adjustment to the formula presented in the previous section. Firstly, the projected wins are adjusted to reflect the 34-round \( r \) Bundesliga season. Secondly, since the Bundesliga football standings are determined by the total match points, the projected wins are altered to reflect the projected match points of a team. The ProjPoints equation presented in (4.12) is applied in place of (4.1), while the MaxWins and ParWins functions become MaxPoints and ParPoints, respectively. Note that the model is applied only to league play, where non-league matches, such as Champions League and DFB-Pokal matches, are not considered at this time.

\[
ProjPoints_s(r) = \left\lfloor \frac{TP_s(r) \cdot 34}{r} \right\rfloor \quad (4.12)
\]

The importance for the Bundesliga is also calculated for all positions except for position 18, which has the same end-of-season outcome as position 17 and therefore is seen as redundant. The importance is also calculated using the win and loss conditional
probabilities presented in (4.6) and (4.7), with the draw outcome not considered at this time due to the binomial distribution only accounting for two outcomes. The OS and PS weighted-sum approaches are calculated using only the positions that result in a specific end-of-season outcome. For division one, this includes the top seven positions as well as the relegation positions 15-17; while for division two, this includes the top three positions as well as the same relegation positions. Note that, although position seven in first division may not result in Europa League qualification (as detailed in Chapter 3), it is included here as the result of the DFB-Pokal is not known until later in the season. The adjusted OS and PS equations for division one is presented in (4.13) and (4.14), respectively.

\[
OS_i(r + 1) = \sum_{s=1}^{7} \frac{1}{s} IMPI_{i,s}(r + 1) + \sum_{s=15}^{17} \frac{1}{(s-7)} IMPI_{i,s}(r + 1)
\]

(4.13)

\[
PS_i(r + 1) = \sum_{s=1}^{7} \frac{1}{|Rank_i(r) - s| + 1} IMPI_{i,s}(r + 1) + \sum_{s=15}^{17} \frac{1}{|Rank_i(r) - s| + 1} IMPI_{i,s}(r + 1)
\]

(4.14)
4.3. Results

This section presents the results of the ParWins method that is an extension of the work completed by Bedford and Schembri (2006). The results focus on the positional importance for both NBA basketball and Bundesliga football, while also presenting team-specific case studies that demonstrate the change in match importance over the duration of the season. While it was detailed in the previous section that the positional importance can be summarised into a single value, the results presented in this section will focus on the individual positional importance values to assess the model’s ability to quantify match importance with respect to achieving each position within the standings. All calculations were completed using Microsoft Excel.

4.3.1. NBA

The positional importance was calculated for the top eight positions within both the Eastern and Western conferences of the NBA. The calculations were completed after each day of the season due to the uneven schedule format. The nine NBA seasons from the primary data outlined in Chapter 3 were used, where the length of every season was approximately the same ($\mu=163.11$ days, $\sigma=0.78$). Despite the calculations being completed after each day of the season, the results are presented after the completion of each game to provide an understanding of how the importance changes over the course of the 82-match season. The results commence with an examination of the average importance by position across all teams and seasons, which is presented in Figure 4.4.
In Figure 4.4, the average importance for each of the lower positions (5-8) within each conference appears greater than the higher positions (1-4). This is expected as the seven teams ranked outside the top eight are more likely to finish within the lower positions compared to the higher positions. An interesting result from Figure 4.4 is that the average importance for each position increases towards the conclusion of the season. This is unexpected considering that, logically, the average importance across all teams should decrease towards the end of the season as there are fewer teams in contention for each
position compared to the start of the season. This observation is discussed further in Section 4.4 of this chapter.

While the distributions presented in Figure 4.4 provide a general summary of how the positional importance changes over the duration of a season, they do not provide information about the teams ranked in position $s$. To evaluate this, the average importance for each team ranked in position $s$ after the completion of game $g$ was calculated across the seasons. The position-specific importance distributions are presented for both conferences in Figure 4.5. In Figure 4.5, the average importance for the teams in each position slowly increases as the season progresses. The result differs from Figure 4.4, which is expected as the position-specific importance focuses on the team ranked in position $s$ after the completion of each game of the season. The increasing trend indicates that a team’s most important match when they currently are ranked in position $s$ would, on average, occur towards the conclusion of the season.
To further evaluate the positional importance, a case study of the two top teams from each conference from the 2013/2014 NBA season is assessed. For the Eastern conference, the Indiana Pacers (IND) finished two games ahead of the Miami Heat (MIA); while in the Western conference, the San Antonio Spurs (SAS) also finished two games ahead of the
Oklahoma City Thunder (OKC). Figure 4.6 provides a graphical comparison of the importance for finishing in first position for each of the four teams, split by conference.

![Graph showing importance for finishing in first position for top two ranked teams from Eastern and Western conferences during the 2013/2014 NBA season.](image)

Figure 4.6: First position importance comparison for top two ranked teams from Eastern and Western conferences during the 2013/2014 NBA season

Focusing on the Eastern conference, the importance for both IND and MIA increase towards the conclusion of the season. This is due to the teams having a similar number of
wins during the final ten games of the season. Both teams played their final match of the season on the final day, where, prior to play, IND led in the standings by one win. Despite this, the most important match according to the ParWins method for IND was their second-to-last, while the most important for MIA was their last. Logically, the final match for IND should have been their most important due to the close proximity of the teams within the standings, where MIA still has the opportunity of moving into first position. This may be due to the nature of the ParWins method, which focuses on achieving a required wins instead of the proximity of teams within the standings.

In the Western conference, OKC have a low importance for the entire season except for a small peak around game 60. This is due to a 19-game win streak by SAS between games 57 and 75, who eventually won 17 of their final 20 matches. During the same stretch of games, OKC won only 12 contests. The large amount of wins accumulated by SAS caused the ParWins of OKC to increase to seven wins after game 75, resulting in a probability equal to zero for finishing in first position. While the ParWins was equal to seven, the teams were only separated by four wins in the standings – meaning that the importance for OKC to finish in first position should be non-zero. For SAS, the importance throughout the first half of the season is equal to zero before an increase during their 19-game win streak, where they are close to achieving the projected wins requirement for finishing in first position. The increase in importance for SAS and the non-zero importance for OKC during this time period does not make logical sense as the teams are only separated by a small number of wins in the standings. Like the Eastern conference results, this may be due to the ParWins method focusing on a required wins total instead of the proximity of teams within the standings.
4.3.2. Bundesliga

The positional importance for Bundesliga football was calculated for each position within the standings, excluding position 18. The calculations were completed after each round of the season for both the first and second divisions using the primary data outlined in Chapter 3. Like the NBA, the average importance by position across all teams and seasons was calculated, with the results presented in Figure 4.7.

Figure 4.7: Average positional importance across all Bundesliga seasons split by division
The season average positional importance distributions for both divisions provide a similar result to those observed within the NBA, with the highest average importance occurring at the conclusion of the season. This is, again, an unexpected result, as the average importance for each position should be highest at the start of the season when there are a greater number of teams in contention for each position. Also like the NBA, the average importance for the lower positions is greater than the higher positions, which is expected given that teams ranked low in the standings are more likely to finish within these positions.

To evaluate the match importance across the season for individual teams, a case study of the top two teams from the 2006/2007 season for both divisions is assessed. For the first division, VfB Stuttgart finished two points ahead of Schalke 04; while for the second division, Karlsruher SC finished eight points ahead of Hansa Rostock. The importance for finishing in first position for all four teams is presented in Figure 4.8.
For division one, VfB Stuttgart won their final eight games of the season to claim the league championship from Schalke 04, who had been in first position for a majority of the season. With one round remaining, VfB Stuttgart led Schalke 04 by two points, which resulted in both teams recording a season-high importance for finishing in first position. This is an expected result, considering the close proximity of teams within the standings. However, despite the end-of-year result being correct, the low importance throughout the season for both teams is questionable.
For division two, Karlsruher SC led Hansa Rostock by three points after the completion of round 28. Despite the close proximity of the teams, the importance for Hansa Rostock is equal to zero for a majority of the season. After round 28, Hansa Rostock had a $ParPoints$ equal to 15, which was still possible to achieve given that there was 18 points still available over the final six matches. This result suggests that the $ParWins$ method is not correctly quantifying the importance of a match in Bundesliga football, as Hansa Rostock should have an importance greater than zero for finishing in first position when trailing by only three points. The results from both divisions are discussed in the next section of this chapter.
4.4. **Discussion**

The probabilistic measure of match importance presented in this chapter was an extension of the approach conceived by Bedford and Schembri (2006). While the original measure was conceived for application within Australian Rules football, minor adjustments were made so that the measure could be applied to both a two-result (NBA basketball) and a three-result sport (Bundesliga football). The measure was also applied to calculate the importance of a match with respect to finishing in different positions within the standings, instead of simply finishing within the top eight as originally designed. The different positions were selected based on their end-of-season outcomes, such as position four within first division Bundesliga that results in qualification for the UEFA Champions League.

When applying the measure to different positions within both sports, unexpected season average distribution results were observed. The season average results for each position steadily increased as the season drew to a close, which was unexpected considering that there are fewer teams in contention for each position late in the season compared to the start. Since there are fewer teams in contention late in the season, the average positional importance should be on the decline, as a majority of teams would be contributing importance values equal to zero. This questionable result casts doubt on the use of the ParWins method to model the importance of a match within both a NBA basketball and Bundesliga football.

A second questionable result was observed within the case studies of the importance for finishing in first position for the two top teams in each sport. For the second-place teams, the match importance was equal to zero despite the teams being in close proximity of the first-place team in the standings. For example, Hansa Rostock recorded an importance equal to zero for finishing in first with six rounds remaining, despite trailing Karlsruher SC by three points. In the NBA Western conference, Oklahoma City (OKC) recorded non-zero
importance values despite being within proximity of the first-place San Antonio (SAS) with a number of rounds remaining. These results are most likely due to the nature of the ParWins method, which focuses on a team achieving a wins requirement, instead of the proximity of teams within the standings. While there are no guarantees that the second-place teams would move ahead of the first-place teams, an importance value equal to zero when teams are close in the standings is illogical.

Like the original measure by Bedford and Schembri (2006), a constant win probability of 50% was applied across all teams within the cumulative binomial distribution function. This was applied in order for a baseline result to be obtained within both NBA basketball and Bundesliga football. However, the application of a constant win probability across all teams does not take into account factors such as home ground advantage, team strength, and the current form of teams. It is possible that allocating a larger win probability to stronger teams may improve the overall results of the ParWins model. While the equal win probability allowed for a baseline measurement to be obtained, future research on the ParWins method can focus on varying the match outcome probabilities, which could be completed using the bookmaker betting odds (Štrumbelj & Šikonja, 2010), or through ordinal logistic regression (Scarf & Shi, 2008).

A further drawback of applying a constant win probability within Bundesliga football is that the draw outcome is not taken into account. A win probability equal to 50% is not realistic due to the large amount of draw outcomes that can occur during a football season. This issue was not relevant to Bedford and Schembri (2006) when applying the ParWins method to Australian Rules football (AFL). While AFL is technically a three-result sport, the frequency of the draw outcome is small, with only 20 of the 2,034 (0.98%) of the matches played between 2005 and 2015 resulting in a draw (afltables.com, 2016). In Bundesliga football, approximately 26% of the matches played across the same period resulted in a draw.
outcome. Finally, applying a cumulative binomial distribution within a three-result sport is questionable, as the probability distribution only accounts for the possibility of two outcomes, success and failure.

In order for a probabilistic measure of match importance to account for the draw outcome in football, changes in both the definition and calculation of match importance are required. This includes applying an alternative statistical model to the cumulative binomial distribution function that was assessed within this chapter. Furthermore, a focus on alternative model inputs, instead of a projected wins total, may lead to improved results within both NBA basketball and Bundesliga football. Since the results contained within this chapter casts doubt on the use of the ParWins method to calculate the importance of a match within both sports, a new approach will be explored in the next chapter of this dissertation.
4.5. Summary

In this chapter, the importance of a match was quantified in both NBA basketball and Bundesliga football by extending an existing probabilistic approach. The approach calculated a wins requirement for a team to finish in different positions within the standings, while a cumulative binomial distribution was applied to assess the probability of achieving this wins requirement conditional on a win or a loss in a team’s next match. While the application of the measure showed that the importance of match can be calculated for individual positions in the standings, flawed results proved that the approach could not be applied to accurately measure the match importance in both a two-result and a three-result sport.
Chapter 5

Overtake importance

In this chapter, a new probabilistic approach for quantifying the importance of a match in both NBA basketball and Bundesliga football is presented. The new approach focuses on alternative model inputs to the ParWins method and introduces a Markov Chain model for quantifying match importance in both a two-result and a three-result sport. Like the previous chapter, the approach will be applied to calculate the importance of a match with respect to finishing in different positions within the standings, which will be completed for both sports.

The chapter is broken down into the following sections: Section 5.1 provides a brief introduction on the new approach and the improvements over the ParWins method. Section 5.2 details the new methodology for application within Bundesliga football, while also briefly describing its application to NBA basketball. Section 5.3 provides an exploration of how the importance by position varies throughout a season, while Section 5.4 provides a critical discussion of the new measure. Finally, Section 5.5 concludes and summarises the chapter.
5.1. Introduction

In Chapter 4, an extension of the probabilistic match importance measure conceived by Bedford and Schembri (2006) was applied to both NBA basketball and Bundesliga football. By first determining a projected wins requirement for a team to finish in a particular position, a cumulative binomial distribution function was applied to calculate the conditional probabilities of the Schilling (1994) match importance definition. However, the model’s application within both sports produced flawed results, including average importance distributions peaking at unexpected points during the season, and individual teams recording positional importance values equal to zero despite being in proximity of the position within the standings.

The ParWins method presented in Chapter 4 focused on two primary model inputs when completing importance calculations: the wins/points of the team of interest (team $i$) and the wins/points requirement to finish in a particular position ($ParWins/ParPoints$). However, as discussed in Chapter 4, the application of the second model input produced positional importance values equal to zero for some teams, despite the teams being within proximity of the position in the standings. Since the application of this model input produced poor results within both sports, the measure presented in this chapter will instead focus on the proximity of teams in the standings. It is believed that the focus on the total wins/points of the team in position $s$, instead of the projected final wins/points total of the team, will produce improved match importance results within both NBA basketball and Bundesliga football.

A further drawback of the ParWins method was its application within a three-result sport, where the probability of a draw outcome in football was not considered during the importance calculations. This was due to the ParWins method applying a cumulative binomial distribution to complete the match importance calculations, which only considers a
success (win) and a failure (loss). Furthermore, an equal success probability of 50% was applied within Bundesliga football, which is not realistic if the number of drawn matches throughout a season is considered. The new match importance measure presented in this chapter will attempt to incorporate the draw outcome within calculations.

In current literature, quantification of the importance of a draw outcome in football has been overlooked. This is particularly relevant within the Schilling (1994) definition of match importance, where a lack of extension to include the draw outcome has been noted in both Goossens et al. (2012) and Geenens (2014). Therefore, this chapter will aim to extend the Schilling definition to include the draw outcome in football, where the significance of this will be quantified using a new Markov Chain model. It is believed that the application of this model will allow for the measure to be adaptable to both a two-result and a three-result sport.

The new probabilistic approach presented within this chapter will seek to address the drawbacks of the ParWins methods. By extending the Schilling definition of match importance, the new approach will calculate the conditional probabilities using a new Markov Chain model for both NBA basketball and Bundesliga football. The new approach, called the Overtake approach, will also focus on alternative model inputs, which will allow for the proximity of the teams within the standings to be taken into account. Since the model inputs focus on the proximity of teams in the standings, the new approach will calculate the importance with respect to a lower-ranked team overtaking a higher-ranked season. The model’s application within Bundesliga football is detailed within the next section of this chapter, with the required adjustments for use within NBA basketball also briefly described.
5.2. Methods

To include the draw outcome in football, the definition by Schilling (1994) is adjusted. The draw outcome is included alongside the win outcome as a non-negative result with respect to a team achieving position \( s \), with the loss outcome being included as a negative result. The adjusted definition now states that the importance of a team’s next match \((r+1)\) is calculated by taking the difference between two conditional probabilities: the probability that team \( i \) finishes in position \( s \) given they achieve a non-negative result (win/draw) in their next match, and the probability that team \( i \) finishes in position \( s \) given they achieve a negative result (loss) in their next match. This adjusted definition means that an overall ‘result importance’ \((RIMP)\) is calculated instead of a win importance that ignores the draw outcome.

\[
RIMP_{i,s}(r+1) = P(Pos_s|WD(r+1)) - P(Pos_s|L(r+1)) \tag{5.1}
\]

In the previous chapter, the ParWins method applied a fixed (although estimated) final points target to complete the conditional probabilities calculations for the Schilling (1994) definition of match importance. However, as discussed in Chapter 4, the use of this points target produced flawed results; including teams recording a positional importance value equal to zero despite being within proximity of position \( s \) in the standings. The new method presented in this chapter will instead incorporate the current total points of the higher-ranked team who holds position \( s \) after the completion of round \( r \) within calculations to generate approximations of the conditional probabilities contained within (5.1).

In this approximation, the knowledge of the initial match points deficit between team \( i \) and the team who currently holds the position of interest (team \( j \)) can be used to simulate the final season points total; completed by applying a Markov Chain model. This deficit, taken as the difference between the current points totals, indicates how many season points team \( i \) must gain, in addition to the season points that team \( j \) gain, to be able to achieve the target
position. As a result, it is possible to calculate the probability that team \( i \) achieves a greater final points total than team \( j \), meaning that they have overtaken and captured the desired position. The application of this non-fixed deficit throughout the season allows for the new model to account for more variability in the end-of-season results, which cannot be accomplished when using a fixed wins requirement.

Before the methodology for the Markov Chain model is provided, win/draw/loss match outcome probabilities for football must be detailed. For simplicity, constant win/draw/loss match outcome probabilities will be applied across all teams. These probabilities were calculated empirically from the nine observed seasons contained within the primary data that was outlined in Chapter 3. Draws were found to have occurred in approximately 25% and 28% of the first and second division Bundesliga matches, respectively, meaning that the remaining percentage of matches comprise of non-drawn outcomes. As a result, the constant win/draw/loss probabilities \((p_w/p_d/p_l)\) are set to 0.375/0.25/0.375 for division one and 0.36/0.28/0.36 for division two. While the constant probabilities do not account for factors such as home ground advantage or team strength, it allows for a baseline model to be trialled.

### 5.2.1. Markov Chain model

As mentioned, the points deficit between team \( i \) and team \( j \) can be used to simulate the final seasons points total. The deficit between the two teams as the season is played out has a state space, \( S \), that can be modelled as a Markov Process, given the deficit itself varies predictably (however probabilistically) as each round is completed. \( S \) comprises of a set of integers from the range \([-102, 102]\), totalling 205 elements. This range comes about because the deficit can only move a total of 3 units (a win is worth 3 points) in the positive or negative directions; meaning over the course of the 34-round season, the maximum and minimum values of the deficit can actually take 102 and -102, respectively.
\[ S = \{s_{-102}, s_{-101}, \ldots, s_0, \ldots, s_{101}, s_{102}\} \] (5.2)

When progressing from round \( r \) to round \( r+1 \), there are only seven possible values that the deficit can become. These values depend on the match outcome of round \( r+1 \) for both team \( i \) and team \( j \). The possible changes in the deficit are presented in Table 5.1.

<table>
<thead>
<tr>
<th>Team ( i / ) Team ( j )</th>
<th>Wins</th>
<th>Draws</th>
<th>Loses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wins</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Draws</td>
<td>-2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Loses</td>
<td>-3</td>
<td>-1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.1: The change in deficit by match outcome dependent on team \( i \) and team \( j \)

The evolution of the deficit as the season progresses can be evaluated by generating a probability distribution after the completion of each round of the season. This probability distribution \( (P) \), presented in (5.2), approximates the probability that the end-of-season deficit will equal some value contained within \( S \). Since a positive end-of-season deficit would indicate that team \( i \) has overtaken team \( j \) in the standings, the probability distribution can be used to approximate the probability that team \( i \) will finish in position \( s \) after round \( r \) given the outcome of their next match.

\[
P = \begin{pmatrix}
p_{1,-102} & \cdots & p_{1,102} \\
\vdots & \ddots & \vdots \\
p_{34,-102} & \cdots & p_{34,102}
\end{pmatrix}
\] (5.2)

Each element of \( P \) is calculated through a product of two probability matrices, which is presented in (5.3).

\[ p_{34-r,s_d} = TP \cdot A^T \] (5.3)
Where,

\[ A = \begin{bmatrix}
p_{34-r-1,s_d-3} & p_{34-r-1,s_d-2} & p_{34-r-1,s_d-1} & p_{34-r-1,s_d} & p_{34-r-1,s_d+1} & p_{34-r-1,s_d+2} & p_{34-r-1,s_d+3}
\end{bmatrix} \]

The matrix \( TP \) consists of seven transition probabilities that are a function of the constant match outcome probabilities that were previously outlined. These transition probabilities are presented in Table 5.2 and allow for the deficit to evolve as the season progresses. In Table 5.2, a change in the deficit of -2 would indicate that team \( i \) has recorded a draw in their round \( r+1 \) match (accumulated one match point) while team \( j \) has recorded a win in their round \( r+1 \) match (accumulated three match points). The matrix \( A \) consists of seven probabilities that correspond to the seven possible values that the deficit can transition into when moving from one round to the next. This comes about because, as mentioned, there are only seven values that the deficit can become when transitioning from round \( r \) (34-\( r \) matches remaining) to round \( r+1 \) (34-\( r-1 \) matches remaining).

<table>
<thead>
<tr>
<th>Change in deficit</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>( p_l p_w )</td>
<td>( p_d p_w )</td>
<td>( p_l p_d )</td>
<td>( p_w^2 + p_d^2 + p_l^2 )</td>
<td>( p_d p_l )</td>
<td>( p_w p_d )</td>
<td>( p_w p_l )</td>
</tr>
<tr>
<td></td>
<td>0.14</td>
<td>0.09</td>
<td>0.09</td>
<td>0.34</td>
<td>0.09</td>
<td>0.09</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Table 5.2: Transition probabilities calculated from the first division Bundesliga constant match outcome probabilities

For all \( p_{1,s_d} \) in \( P \), an alternative probability matrix to \( A \) is applied to complete calculations. The probabilities contained within this alternative matrix \( (B) \) is presented in Table 5.3 and reflect the probability that the deficit will change given the possible result of team \( i \)'s next match. These are denoted as the conditional transition probabilities as they are
calculated with respect to the possible result of the next match for team \( i \). As there are three possible results for team \( i \)’s next match, the probability distribution \( P \) is generated three separate times after the completion of each round. Each probability distribution can then be evaluated to approximate the probability that team \( i \) finishes in position \( s \) given they win, draw, or lose their next match.

<table>
<thead>
<tr>
<th>Change in deficit</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability given team ( i ) win their ( r+1 ) match</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.14</td>
<td>0</td>
<td>0.09</td>
<td>0.14</td>
</tr>
<tr>
<td>Probability given team ( i ) draw their ( r+1 ) match</td>
<td>0</td>
<td>0.09</td>
<td>0</td>
<td>0.06</td>
<td>0.09</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Probability given team ( i ) lose their ( r+1 ) match</td>
<td>0.14</td>
<td>0</td>
<td>0.09</td>
<td>0.14</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.3: Conditional transitional probabilities calculated from the first division Bundesliga match outcome probabilities

Note that only three elements of the conditional transition probability matrix are non-zero as the outcome of team \( i \)’s next match has been assumed – meaning the deficit can only change by three values now, which depend on the outcome of team \( j \)’s next match. To approximate the probability that team \( i \) finishes in position \( s \) given the result of their next match, a summation of the probabilities within \( P \) for when the deficit is positive is taken. Again, a positive end-of-season deficit would indicate that team \( i \) has overtaken team \( j \) in the standings and secured the desired position.
A key assumption for this simplified approach is that the outcomes between teams are independent. Without a pre-defined season schedule, it is impossible to guarantee individual match outcomes remain mutually exclusive. As a result, circumstances where this is a necessary consideration for the current season scenario (e.g. prior to the final round when two teams are playing each other and competing for the same final position) are unrealistic. Such a drawback is unavoidable given the simplified approach being applied, but given 32 of a team’s 34 matches are played against other teams the effects of such an assumption can be seen as negligible. This assumption is discussed further in Section 5.4.

5.2.2. Result importance

The importance of a match is calculated in a similar manner to previous literature, but with the draw outcome being accounted for and an overall result importance being generated instead of solely a win importance, as presented in (5.1). Again, the outcome of a match is considered to be either negative (a loss and no increase in the current season points, meaning the deficit cannot be reduced) or non-negative (either a win or a draw, resulting in an increase in the current season points of three points or one point respectively, and the possibility of reducing the deficit). The components of (5.1) are calculated as follows:

\[
P(\text{Pos s}|WD(r + 1)) = P(\text{Pos s}|W(r + 1)) + P(\text{Pos s}|D(r + 1))
\]  

\[
P(\text{Pos s}|L(r + 1)) = \sum_{i>0}P(\Delta_{34} = i|L) + I_{P(\Delta_{34}=0|L)}
\]  

\[
P(\text{Pos s}|W(r + 1)) = \sum_{i>0}P(\Delta_{34} = i|W) + I_{P(\Delta_{34}=0|W)}
\]  

\[
P(\text{Pos s}|D(r + 1)) = \sum_{i>0}P(\Delta_{34} = i|D) + I_{P(\Delta_{34}=0|D)}
\]
Note that (5.5)-(5.7) include an indicator function \((I)\) that equals one when team \(i\) already holds position \(s\), and zero when it is currently ranked in a lower position. This is included to account for the situation where the final deficit \((\Delta_{34})\) is zero, meaning the two teams have equal total season points. In reality, two teams who share equal points are ranked by their goal difference, but given this is not considered in the modelling it is assumed that the team who is currently ranked higher remains higher (essentially, it is assumed the higher team will maintain a better goal difference throughout the season, such that if both teams finish with equal points, they would in reality be awarded position \(s\)). A final deficit of zero therefore is considered a success for the team who is currently ranked higher, meaning the probability of its occurrence must be included in their importance calculations.

Since there are two non-negative outcomes that can occur (win and draw), it is possible to break down the result importance into win and draw components. As the result importance is measured on a 0-1 scale, where higher values correspond to higher importance, the win and draw components also take values within this interval. The win and draw components, partitioned to show their contribution to the overall result importance, is presented in (5.8) and (5.9), respectively.

\[
WIMP_{i,s}(r + 1) = RIMP_{i,s}(r + 1) \cdot \frac{P(Pos_s|W(r+1))}{P(Pos_s|W(r+1))+P(Pos_s|D(r+1))}
\]  (5.8)

\[
DIMP_{i,s}(r + 1) = RIMP_{i,s}(r + 1) \cdot \frac{P(Pos_s|D(r+1))}{P(Pos_s|W(r+1))+P(Pos_s|D(r+1))}
\]  (5.9)

The win and draw components represent the importance of achieving a win and a draw, respectively, in team \(i\)'s next match with respect to finishing in position \(s\). Since a win (three points) rewards a team greater than a draw (one point), it is reasonable to assume that the win importance will always contribute more to the overall result importance than the draw importance. As the probabilities that form the result importance are seen as an approximation,
the result importance and its components should also be viewed as an approximation of a match’s criticality with respect to finishing in position \( s \). The results in the next section will determine the suitability of these approximations for quantifying the importance of a match with respect to finishing in position \( s \).

A custom VBA program was developed to complete the Markov Chain model calculations for each of the nine observed seasons for both first and second division Bundesliga. The program identifies the current position of team \( i \) in the standings after the completion of round \( r \) and compares their total points to team \( j \) who currently holds position \( s \). The program also assesses the proximity of team \( i \) to position \( s \) in the standings to determine if the team is ranked higher or lower. If a team is found to be ranked above position \( s \) in the standings, then their result importance is set to zero as they have already overtaken the position after round \( r \). If team \( i \) is currently holding position \( s \), then the total points of the team in position \( s+1 \) is substituted in as team \( j \) and the probabilities are calculated with respect to team \( i \) remaining as the higher team in position \( s \) (i.e. the previously mentioned indicator functions are equal to 1). The code for the VBA program is presented in Appendix A.

As mentioned in Chapter 4, it may be advantageous to summarise the positional importance into a single value. This can, again, be completed by assigning a weight to each positional importance value, where the weighting can reflect either the hierarchy of positions or the proximity of each position to the current rank of team \( i \). The single value can also reflect the total sum of the positional importance values that result in a desired end-of-season outcome. The three new weighted-sum approaches (Outcome Sum (\( OS \)), Positional Sum (\( PS \)) and Total Sum (\( TS \)) for first division Bundesliga football are presented in (5.10)-(5.12).

\[
OS_i(r + 1) = \sum_{s=1}^{7} \frac{1}{s} RIMP_{i,s}(r + 1) + \sum_{s=15}^{17} \frac{1}{(s-7)} RIMP_{i,s}(r + 1) \tag{5.10}
\]
It is also possible to summarise the positional importance into several different values, where each value approximates the overall importance of achieving the specific end-of-season outcome. For example, the summation of the importance for finishing in positions five, six and seven could approximate the overall importance of qualifying for the UEFA Europa League. The effectiveness of these summarised approaches is explored in the later chapters of this dissertation.

Applying these additive approaches to summarise the positional importance does introduce the issue of double counting of outcomes. However, no assumptions have been made with respect to how the importance values are to be used. Furthermore, the importance values are unitless approximations of a match’s criticality with respect to a team achieving position $s$. Since there are no restrictions on the use of these approximations, it is possible to summarise the values as detailed in the additive approaches.

5.2.3. Example

To demonstrate the new model, called the Overtake approach, consider an example from the 2008/2009 Bundesliga season. After the completion of round 30 ($r=30$), FC Bayern Munich held position two with 57 points while VfL Wolfsburg held position one with 60 points; meaning that the current deficit is equal to -3. The focus of this example will be the importance of finishing in first position. The first step is to generate the probability distribution of the deficit where the result of FC Bayern Munich’s next match ($r=31$) is known (i.e. the conditional probabilities in (5.5)-(5.7)). A sample of the probability
distribution, $P$, for when FC Bayern Munich has won/drawn/lost their next match, and VfL Wolfsburg has either won, drawn, or lost their next match, is presented in Table 5.4.

Each probability contained within Table 5.4 is calculated by applying (5.3). To demonstrate this, consider the element for when the deficit (-3) remains the same with one round remaining for when it is assumed that FC Bayern Munich win their next match. This value is equal to 0.08. Note that, since there is one round remaining for this element, the alternative probability matrix ($B$) from Table 5.3 is applied instead of the original matrix ($A$).

The calculations for this element are as follows:

$$p_{1,-3} = TP \cdot B^T$$

$$p_{1,-3} = \begin{bmatrix} 0.14 & 0.09 & 0.34 & 0.09 & 0.09 & 0.14 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 & 0.14 & 0 & 0.09 & 0.14 \end{bmatrix}^T$$

$$p_{1,-3} \approx 0.08$$

The remaining elements throughout the probability distribution are then calculated in a similar manner.
| FC Bayern Munich | VfL Wolfsburg | Deficit | -12 | -11 | -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|------------------|--------------|---------|-----|-----|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Round 31 result  | Round 31 result | Rounds remaining | | | | | | | | | | | | | | | | | | | |
| Win | Win/Draw/Lose | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.09 | 0.14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 | 0.03 | 0.08 | 0.04 | 0.06 | 0.08 | 0.02 | 0.03 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 2 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.02 | 0.03 | 0.05 | 0.04 | 0.05 | 0.05 | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 3 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.02 | 0.03 | 0.04 | 0.04 | 0.04 | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Draw | Win/Draw/Lose | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.09 | 0.09 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.05 | 0.02 | 0.04 | 0.05 | 0.01 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 2 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.04 | 0.03 | 0.03 | 0.04 | 0.02 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 3 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | 0.03 | 0.03 | 0.03 | 0.02 | 0.02 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lose | Win/Draw/Lose | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.09 | 0.14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 | 0.03 | 0.08 | 0.04 | 0.06 | 0.08 | 0.02 | 0.03 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 2 | 0.00 | 0.00 | 0.01 | 0.02 | 0.02 | 0.03 | 0.05 | 0.04 | 0.05 | 0.05 | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 3 | 0.00 | 0.01 | 0.01 | 0.02 | 0.03 | 0.04 | 0.04 | 0.04 | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Table 5.4: Sample of the transition probability distribution for FC Bayern Munich’s round 31 match from 2008/2009 Bundesliga season
The next step is to calculate the conditional probabilities from (5.5)-(5.7). This is completed by summing together the probabilities in Table 5.4 where the deficit for FC Bayern Munich is positive (i.e. they finish the season with a higher points total than VfL Wolfsburg). Since the results for round 31 have been assumed in Table 5.4, the probabilities for when there are three matches remaining ($34-r-1=3$) is summed together. For (5.6), where FC Bayern Munich has won their round 31 match, the summation of the probabilities from Table 5.4 is as follows:

$$P(1^{st}|W(31)) = 0.03 + 0.03 + 0.02 + 0.01 + 0.01 + 0.01 \approx 0.11$$

The same calculations are completed for when FC Bayern Munich draw their 31st match (5.7) and for when they lose their 31st match (5.5). The same procedure is completed separately to generate the probabilities for VfL Wolfsburg to remain as the higher team in first position. However, the indicator functions in (5.5)-(5.7) are applied; meaning the probability for when the deficit equals zero is included within their summation. The conditional probabilities for both teams are presented in Table 5.5.

<table>
<thead>
<tr>
<th></th>
<th>FC Bayern Munich</th>
<th>VfL Wolfsburg</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(1^{st}</td>
<td>W \text{ round } 31)$</td>
<td>0.1105</td>
</tr>
<tr>
<td>$P(1^{st}</td>
<td>D \text{ round } 31)$</td>
<td>0.0346</td>
</tr>
<tr>
<td>$P(1^{st}</td>
<td>L \text{ round } 31)$</td>
<td>0.0297</td>
</tr>
</tbody>
</table>

Table 5.5: Conditional probabilities for round 31 for FC Bayern Munich and VfL Wolfsburg from 2008/2009 Bundesliga season
The conditional probabilities contained within Table 5.5 are then applied to create the result importance. Using (5.1) and (5.4), the result importance for FC Bayern Munich to finish in first position for their round 31 match is equal to 0.1154, while the result importance for VfL Wolfsburg is equal to 0.2783. The final step is to apply the result importance values to calculate both the win and draw importance components, as detailed in (5.8) and (5.9). For FC Bayern Munich, the win and draw importance is calculated as:

\[
WIMP_{1st}(31) = 0.1154 \cdot \frac{0.1105}{(0.1105 + 0.0346)} = 0.0879
\]

\[
DIMP_{1st}(31) = 0.1154 \cdot \frac{0.0346}{(0.1105 + 0.0346)} = 0.0275
\]

The win and draw importance, along with the result importance, for both FC Bayern Munich and VfL Wolfsburg is presented in Table 5.6. Note that the importance values are not probabilities and do not sum to one. The values simply approximate how critical the next match is for each team to achieve first position and win the league championship, where higher values correspond to a greater match importance.

<table>
<thead>
<tr>
<th></th>
<th>FC Bayern Munich</th>
<th>VfL Wolfsburg</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIMP(31)</td>
<td>0.0879</td>
<td>0.1760</td>
</tr>
<tr>
<td>DIMP(31)</td>
<td>0.0275</td>
<td>0.1023</td>
</tr>
<tr>
<td>RIMP_{1st}(31)</td>
<td>0.1154</td>
<td>0.2783</td>
</tr>
</tbody>
</table>

Table 5.6: Win and draw importance for round 31 FC Bayern Munich and VfL Wolfsburg from the 2008/2009 Bundesliga season
5.2.4. Adaption to two-result sport

Since NBA basketball is a two-result sport, the Markov Chain model is applied without a draw outcome being considered; meaning that the win and loss match outcome probabilities are both set to 50%. The state space, $S$, also reduces to 165 elements as the maximum change in the deficit between two teams has a magnitude 1 and the total number of matches in a season is 82. As match points are not allocated in NBA basketball, the total number of wins is applied within the model. Following the ParWins model, the Overtake approach updates the importance after each day of the season to account for the uneven NBA schedule. Although this may mean that teams $i$ and $j$ have played a different amount of matches after the completion of each day, the importance values are viewed as an approximation where validation of the results are to be completed in Chapter 6. The VBA program for the NBA calculations is presented in Appendix B.

Like the Bundesliga importance, the NBA positional importance can be summarised into a single value by using the three weighted sum approaches: Outcome sum (OS), Positional Sum (PS), and Total Sum (TS). These approaches are presented in (5.13)-(5.15).

\[
OS_i(g_t + 1) = \sum_{s=1}^{8} \frac{1}{s} RIMP_{i,s}(g_t + 1) \quad (5.13)
\]

\[
PS_i(g_t + 1) = \sum_{s=1}^{8} \frac{1}{|\text{Rank}_{i}(g_t) - s|+1} RIMP_{i,s}(g_t + 1) \quad (5.14)
\]

\[
TS_i(g_t + 1) = \sum_{s=1}^{8} RIMP_{i,s}(g_t + 1) \quad (5.15)
\]
5.3. Results

This section aims to explore the application of the new method, called the Overtake approach, to calculate match importance within both Bundesliga football and NBA basketball. The model is applied to the primary data for both sports that was outlined in Chapter 3. The average importance distributions by position are assessed, while the case studies from Chapter 4 are re-evaluated to determine whether an improvement has been made from the ParWins method. Although it was mentioned in the previous section that the positional importance can be summarised into a single value, the focus of this section will be the importance for finishing in each position as this allows for the results from the Overtake approach to be directly compared to those of the ParWins method.

5.3.1. Bundesliga

Like the ParWins method, the Overtake approach was applied to both division one and division two Bundesliga football. Also, non-league matches, such as DFB-Pokal matches, are again not considered. A result that arose from the ParWins method was that the average positional importance for both divisions peaked at the conclusion of the season. This was an illogical result as the number of teams in contention for position $s$ at the end of the season would be less than the number of teams in contention at the commencement of the season. To compare the new model to the ParWins method, the average positional importance across all teams and seasons were calculated for the Overtake approach. The results are presented in Figure 5.1.
Figure 5.1: *Overtake* average positional importance for first and second division Bundesliga

An initial observation from Figure 5.1 is that, according to the *Overtake* approach, the average importance for each position within each division is at a maximum at the start of the season before steadily declining towards the end. This is expected, as there are a greater number of teams in contention for position \( s \) at the start compared to the end; so the average importance across teams should be greater at the commencement of the season. To confirm this, the average number of teams in contention for the key positions across all first division Bundesliga seasons, and their average importance, is presented in Figure 5.2.
Figure 5.2: Average teams in contention and average positional importance for first division Bundesliga football

Two approaches were considered when determining if a team was in contention for position $s$. The first approach considered a team in contention for position $s$ if their result importance for the position was greater than the average across all teams within the season.
The second approach assessed the difference between a team’s total points and the total points of the team in position \( s \). If the difference was less than the maximum number of points a team could achieve for the remainder of the season, then the team was considered in contention for position \( s \). However, the second approach meant that some teams were considered to be in contention despite there being a considerable distance between them and position \( s \) in the standings – resulting in an unrealistic reflection of the number of teams in direct competition for each position. Therefore, the first approach was applied.

In Figure 5.2, it can be observed that the average number of teams in contention for the key division one positions slowly decreases as the season progresses. As observed within the first panel, the average positional importance across all teams also decreases. Focusing on the third panel, the average importance for only the teams in contention actually increases towards the conclusion of the season. The results presented in Figure 5.2 confirm the hypothesis that the average importance across all teams is at a maximum at the start of the season before steadily declining, which corresponds with a decline in the average number of teams in contention within each position.

A similar procedure was completed for the second division Bundesliga seasons to confirm that the result occurs across both divisions. The observed results for the key second division positions are presented in Figure 5.3. Like the first division, the average number of teams in contention for each position decreases towards the conclusion of the season, which corresponds with a decrease in the average positional importance across all teams. A notable result across both divisions is the low number of teams in contention for positions 15 and 16. This is due to the low number of teams ranked below these positions, meaning the number of teams in contention for each position would be consistently low throughout the season.
Figure 5.3: Average teams in contention and average positional importance for second division Bundesliga football

One of the key features of the Overtake approach is that the result importance for finishing in position $s$ can be broken down into win and draw components. To demonstrate this, the case studies from Chapter 4 are re-evaluated. For division one, the result importance
for VfB Stuttgart and Schalke 04, along with the win and draw importance components, are presented in Figure 5.4. Note that the draw importance is always smaller than the win component, since a win (three points) awards a team greater than a draw (one point). In other words, a win will always be more beneficial to a team than a draw.

Figure 5.4: Win, draw and result importance for position one for VfB Stuttgart and Schalke 04 during 2006/2007 Bundesliga season
As observed in Figure 5.4, the importance for both teams is consistent across the season before peaking with one round remaining. Before this final round, Schalke 04 trailed VfB Stuttgart by two points, meaning they required nothing less than a win from their final match to finish in first position. This is observed within Figure 5.4, where the draw importance for Schalke 04 is equal to zero before their final match. This result indicates that, despite including the draw outcome as a non-negative result, the Overtake approach is correctly identifying when the draw outcome can be beneficial. For VfB Stuttgart, both the win and draw components are at a season high before their final match, meaning that they merely had to achieve a non-negative result in their final match to secure the league championship. Both teams went on to win their final matches, resulting in VfB Stuttgart winning the league title.

Figure 5.5 presents the first position importance results for the 2006/2007 second division Bundesliga case study, where Karlsruher SC won the league championship by two points ahead of Hansa Rostock. Unlike the results from Chapter 4, the position one importance for Hansa Rostock is non-zero for a majority of the season. The importance then equals zero with three rounds remaining, when it is mathematically impossible for them to finish in first position. For Karlsruher SC, their importance is consistent across the season before reverting to zero with three rounds remaining.

Within both case studies, the new Overtake approach is providing improved results over the ParWins method with regards to calculating the importance of a match for finishing in first position. The new model is also correctly identifying when a draw outcome can be beneficial for a team despite always including the draw as a non-negative result, such as Schalke 04 within the first division case study. The promising case study results, along with the improved average positional importance distributions, lend support to the use of the Overtake approach for measuring the importance of a match in a three-result sport.
Figure 5.5: Win, draw and result importance for position one for Karlsruher SC and Hansa Rostock during 2006/2007 2. Bundesliga season
5.3.2. NBA

Like the Bundesliga, the average importance was calculated for each position across the primary data using the new *Overtake* approach. In Chapter 4, it was observed using the *ParWins* method that the average importance by position was at a maximum at the conclusion of the season, which was illogical due if the number of teams in contention for position $s$. For the *Overtake* approach, the results for both the Eastern and Western conferences are presented in Figure 5.6.

Similar to the Bundesliga results, the average positional importance for each NBA conference is at a maximum at the start of the season before steadily declining towards the conclusion. This is an improvement over the results observed in Chapter 4 when applying the *ParWins* method, which saw the average positional importance peak at the conclusion of the season. Similarly to the Bundesliga, the steady decline across the season is expected when the number of teams in contention is considered. To explore this, the average number of teams in contention for position $s$ across the seasons was calculated. As described in the previous section, a team was considered in contention for position $s$ if their match importance was greater than the average for a season.
Figure 5.6: Overtake average positional importance for NBA Eastern and Western conferences

Figure 5.7 and Figure 5.8 present the results for the teams in contention within the Eastern and Western conferences, respectively. As observed within both conferences, the average number of teams in contention for position $s$ declines as the season progresses, which corresponds to a decrease in the average positional importance. Like the Bundesliga results, the average importance for only the teams in contention increases as the season progresses,
which was the expected result. These promising results suggest that the *Overtake* approach, adjusted for application within a two-result sport, is correctly quantifying the importance of a match in NBA basketball.

Figure 5.7: Average teams in contention and average positional importance for NBA Eastern Conference
To further explore the _Overtake_ approach within NBA basketball, the case studies first presented in Chapter 4 are re-evaluated. The importance of finishing in first position for the first and second-placed teams from the 2013/2014 season for both conferences is presented in
Figure 5.9. For the Eastern conference, Indiana (IND) finished two games ahead of Miami (MIA); while in the Western conference, San Antonio (SAS) also finished two games ahead of Oklahoma City (OKC).

In the previous chapter, it was observed that the calculated importance for finishing in position one was flawed as it was equal to zero for the second-place teams despite the teams being within proximity of position one with a number of matches remaining. This was most noticeable within the Western conference, where OKC recorded an importance equal to zero for a majority of the season, despite being closely ranked in the standings to SAS. The result is improved in Figure 5.9, where the importance for OKC is a consistent non-zero value throughout the season before declining at the conclusion. The decline at the end of the season is due to a 19-match win streak for SAS, who secured the conference championship with a small number of matches remaining in the season.

In the Eastern conference, the importance for both IND and MIA are consistent throughout the season. However, the importance for both teams equals zero before each team’s final match, despite IND only leading by one game in the standings. This result differs slightly from those observed when applying the ParWins method, which saw only the importance for IND equalling zero before the final match. The difference in results can be explained by the key inputs of each model, where the importance for IND under the ParWins method equals zero because the team has achieved the projected wins requirement. Under the Overtake approach, MIA can only equal them not overtake them; resulting in an importance value for both teams equal to zero. Despite the similar results within the Eastern conference, the improved results within the Western conference supports the application of the Overtake approach for quantifying the importance of a match in NBA basketball.
Figure 5.9: Position one importance for top two teams in NBA Eastern and Western conferences from 2013/2014 NBA season
5.4. Discussion

The new probabilistic measure of match importance presented in this chapter was designed to account for the drawbacks of the ParWins method. By focusing on the probability of a lower-ranked team overtaking a higher-ranked team, the new Overtake approach calculated the match importance for multiple positions using a Markov Chain model for both Bundesliga football and NBA basketball. The application of a Markov Chain model allowed the new approach to be adaptable to both a two-result (basketball) and a three-result (football) sport; which was not successfully achieved when applying the ParWins method.

To demonstrate the new Overtake approach, past season results for both Bundesliga football and NBA basketball were assessed. The results from both sports showed that the average positional importance was at a maximum at the commencement of the season before steadily declining towards the conclusion; which was an improvement over the results obtained from the ParWins method. The decreasing average positional importance results also coincided with the average number of teams in contention for position $s$ throughout the season, which steadily declined as the season progressed. Despite these sensible results, a confirmation of the average importance distributions is required to confirm the model’s ability to calculate the match importance. This will be explored in the next chapter of this dissertation.

While the importance distributions for a majority of positions were at a maximum at the start of the season, the result was slightly different for the relegation positions within Bundesliga football. As observed in Figure 5.1, the average importance for the lower-positions was consistent throughout the season, with a distinct decreasing trend not observable. This result is due to the number of teams in contention for these low positions
throughout the season, which is detailed in Figure 5.2 (first division) and Figure 5.3 (second division). For these positions, the number of teams in contention remains consistent throughout the season, which, like the other positions, corresponds with the shape of the positional importance distribution. The constant result suggests that there is always competition to avoid relegation in Bundesliga football. This should be expected given it is doubtful that any team would aspire to be relegated down a division.

One of the key differences between the ParWins method and the Overtake approach was the focal model inputs. The Overtake approach replaced the projected wins/points with the actual total wins/points of the team in position $s$, which led to an overall improvement in the positional importance results. This was noticeable within the case studies from both sports, which now saw the second-place teams recording a non-zero importance of finishing in first position while they remained within proximity of the first-place team in the standings. The improved results within both sports demonstrate that the new Overtake approach can be successfully applied to both a two-result and a three-result sport.

The application of a Markov Chain model to calculate the importance of a match in Bundesliga football allowed for the significance of the draw outcome to be quantified for the first time in the literature. By defining the draw outcome as a non-negative result, the draw importance was calculated by partitioning the overall result importance for position $s$. However, the draw outcome may not always be a non-negative result to a team, especially if they require nothing less than a win to finish in position $s$. Despite this, the inclusion of the draw outcome as a non-negative result still produced sensible results within the case studies, illustrated by the draw importance equalling zero for Schalke 04’s final match when they trailed VfB Stuttgart by two points. Nevertheless, future research can focus on varying the draw outcome between a negative and a non-negative result for individual teams.
One of the key assumptions with the calculation of the conditional probabilities was the independence of match results of the two teams. Although this is not the case when two teams play each other, the effect on the overall calculations is minimal since two teams only play each other twice in Bundesliga football, and at most four times in NBA basketball. Furthermore, an effect would only be observed when the two teams are playing each other late in the season. To determine if this assumption has an effect on the overall results, a comparison with a simulation model that accounts for the dependency of match results will be completed in the next chapter of this dissertation.

To create a baseline model, the new Overtake approach applied constant match outcome probabilities across all teams to complete the match importance calculations within both sports. However, application of constant match outcome probabilities does not take into account factors such as team strength, the current form of teams, or home ground advantage; where the latter has been found to be prominent within both basketball and football (Schwartz & Barsky, 1977). For example, across the nine Bundesliga football seasons, FC Bayern Munich recorded an overall win percentage of 67%, which increased to 76% when playing at home. As mentioned in the Chapter 4 discussion, there are different statistical approaches that can be applied to generate team-specific match outcome probabilities, which may lead to an enhancement of the results contained within this chapter. However, the use of the constant match outcome probabilities allowed for a baseline model to be established, which can be further developed in future research. Nevertheless, the next chapter of this dissertation will explore applying a Bayesian approach to account for the current form of teams within the importance calculations.

A potential weakness of the new model is that it fails to account for the teams who are ranked between team $i$ and team $j$ in the standings. The results of these teams’ matches could have an effect on the probability that team $i$ overtakes team $j$, given these teams are currently
ranked above team \( i \) and therefore closer to team \( j \) in the standings. However, as mentioned when detailing the Markov Chain model, the generated importance values are approximations of how critical the matches are to team \( i \) with respect to achieving position \( s \). A comparison of the results with a Monte Carlo simulation model, which will be completed in the next chapter of this dissertation, will help determine if this failure to account for the in-between teams is a major weakness of the Overtake approach.

A drawback of the Overtake approach is that the model is designed to only calculate the importance for the next match in the season. Consequently, the importance of future matches cannot be calculated. In order to achieve this, a computer simulation model, such as Monte Carlo simulation (Lahvička, 2015; Scarf & Shi, 2008), would be a suitable alternative. However, since simulation models can be computationally more complex and require a great deal of runtime to complete calculations, the Overtake approach can be applied to swiftly and reasonably calculate the importance of an immediate match. Moreover, it can be applied to calculate the match importance for past season results, which would be advantageous for those seeking to include match importance as a variable within a statistical model.

While the measure presented within this chapter was primarily detailed for application in Bundesliga football and NBA basketball, the structure of the methodology allows for the approach to naturally be applied to other professional sports. This includes the top European football leagues, such as the English Premier League (EPL) and Spanish La Liga, which both follow a similar format to the Bundesliga. In terms of NBA basketball, as well as being applicable to international basketball leagues, the new measure could be applied to other major North American sports, such as American football (NFL), Ice Hockey (NHL) and Major League Baseball (MLB). A comparison between the results obtained from other professional sports with those observed within this chapter would provide an interesting
discussion of when the most important matches occur throughout a season; which can be completed within future research on the Overtake approach.

While past results were assessed to demonstrate the functionality of the new Overtake approach within both sports, it should be emphasised that any measure of match importance is an estimation of the true value. Since there is no ‘gold standard’ in quantifying the importance of a match in professional sports, all measures should be seen as an estimation of the true importance, where the validity of each measure should be confirmed by assessing results like those presented in this chapter. While each measure assesses different team information, the most complete approach is Monte Carlo simulation as this takes into account the match results of all teams. As mentioned, a comparison between this model and the Overtake approach will be completed in the next chapter to further validate the observed results for both sports.

The new measure presented in this chapter quantified the importance of a match in both Bundesliga football and NBA basketball, with the significance of the draw outcome evaluated for the first time within football. While the application to past seasons provided sensible results within both sports, confirmation of the average positional importance distributions is required. Furthermore, confirmation of the match independence assumption and a variation of the match outcome probabilities are required to further validate the use of the Overtake approach for calculating match importance.
5.5. Summary

In this chapter, a new probabilistic measure of match importance, called the *Overtake* approach, was presented. The new measure focused on alternative model inputs to the previous chapter’s model, and instead approximated the conditional probabilities with respect to a lower-ranked team overtaking a higher-ranked team. While being applied to multiple positions, the *Overtake* approach was found to be adaptable to both a two-result (NBA basketball) and a three-result (Bundesliga football) sport. Furthermore, the application of a Markov Chain model within football enabled the draw outcome to be quantified for the first time in the match importance literature. While the observed results were logical, confirmation on the average positional importance distributions and a variation of the match outcome probabilities is required.
Chapter 6

Simulation

In this chapter, a comparison between the *Overtake* approach and a Monte Carlo simulation model will be completed. The comparison will focus on comparing the average positional importance distributions for both NBA basketball and Bundesliga football to determine if the new model, with its reduced computational complexity, produces similar results to the complete simulation procedure. A secondary objective of this chapter is to explore whether a variation of the match outcome probabilities has a pronounced effect on the season average importance distributions.

The chapter is broken down into the following sections: Section 6.1 provides a brief introduction to Monte Carlo simulation and its application within the match importance literature. Section 6.2 details the methodology for applying Monte Carlo simulation to calculate the importance of a match within both sports, as well as varying the match outcome probabilities to account for the current form of teams. Section 6.3 provides a detailed comparison of the simulation model with the *Overtake* approach, while Section 6.4 critically discusses the observed results within both sports. Finally, Section 6.5 will conclude and summarise the chapter.
6.1. Introduction

In Chapter 5, the new *Overtake* approach was applied to both Bundesliga football and NBA basketball. By focusing on alternative model inputs, the new approach provided improved average positional importance results to those observed from the *ParWins* method in Chapter 4. However, confirmation of the positional importance distributions, and validation of the match independence assumption, is required. To complete this, a Monte Carlo simulation model will be applied within both sports, with the results providing a basis for comparison with those presented in the previous chapter.

Monte Carlo simulation has been applied in past literature to calculate the importance of a match in professional sports. As detailed in Chapter 1, Scarf and Shi (2008) applied this procedure to English Premier League football, where the probabilities from the Schilling (1994) definition of match importance were estimated dependent on whether team *i* achieved an outcome of interest, conditional on a favourable outcome in the current match of interest and the results of all matches played up to time *t*. A favourable outcome was then defined as a win while an unfavourable outcome was defined as a loss, with the draw outcome not considered. By defining match importance as the strength between match results and a season outcome, Lahvička (2015) also applied the model to EPL football. The author argued that the model is more suitable for quantifying the importance of a match than the Schilling method, despite failing to compare results between the two approaches.

Like any of the match importance measures, there are both benefits and drawbacks associated with Monte Carlo simulation. While the method takes into account the results of all matches, a great deal of runtime is required to complete the calculations; which can be disadvantageous to those seeking a straightforward method of calculating the importance of a match. A possible alternative is to apply the *Overtake* approach, which relaxes match
independence assumptions and requires fewer model inputs. The comparison between the two models within this chapter will help determine if the reduction in computational complexity of the Overtake approach provides similar results to a complete Monte Carlo simulation procedure.

While the primary objective of this chapter is to compare the Overtake approach to a Monte Carlo simulation model, a secondary objective is to explore a variation of the match outcome probabilities within both NBA basketball and Bundesliga football. To account for the current form of a team during the season, Bayes’ Rule will be applied to both sports to alter the constant match outcome probabilities outlined in the previous chapter. While this approach does not account for other factors such as team strength and home ground advantage, it will allow for the Overtake approach to be tested with different match outcome probabilities for each team.

The Monte Carlo simulation model presented in this chapter will follow a similar structure to other simulation models from past literature. Like the Overtake approach, the simulation model will calculate the importance with respect to finishing in individual positions for both Bundesliga football and NBA basketball. The average importance across the simulated seasons will then serve as the basis for comparison with the results from the Overtake approach. The methodology for applying the Monte Carlo simulation model within both sports is detailed in the next section of this chapter.
6.2. Methods

This section details the application of the Monte Carlo simulation model to calculate the importance of a match in both Bundesliga football and NBA basketball. This section also outlines the application of Bayes’ Rule to vary the match outcome probabilities with respect to the current form of the two competing teams.

6.2.1. Monte Carlo simulation

A Monte Carlo simulation model provides the basis for comparison with the Overtake approach. Developed using a custom VBA program in Microsoft Excel, the procedure first generates a full season schedule following the format of either Bundesliga football or NBA basketball. Each generated season schedule is then simulated a number of times to completion to produce success probabilities for finishing in position \( s \), conditional on individual match outcomes throughout the season. Match importance is then evaluated using these probabilities conditional on the current round/game and a team’s current position in the standings, completed using the Schilling (1994) definition of match importance. The VBA programs for Bundesliga football and NBA basketball are presented in Appendices C and D, respectively.

While the Overtake approach split the result importance for football into win and draw components, the Monte Carlo simulation model will only focus on calculating the importance according to the original Schilling (1994) definition. Quantification of the draw outcome was trialled during initial modelling testing, with the end result producing an unrealistic scale for the draw importance compared to the win importance. Furthermore, the primary objective of this chapter is to validate the average positional importance distributions, which is calculated using the overall result importance in football. Therefore, the simulation model for football will focus on only the win and loss outcomes.
For the Monte Carlo simulation model, 1,000 season-schedule iterations will be applied, with a further 100 season-result iterations required to generate the success outcome probabilities for the match importance calculations within each season schedule. While the 100 season-result iterations is a low number, it was determined through trial and error that a larger number dilated the outcome probabilities; which led to unrealistic match importance results, such as importance values being equal to zero early in the season. As for the season-schedule iterations, 1,000 were selected due to the large amount of runtime required to complete the calculations. The end result is a total of 100,000 iterations (1000 x 100), which will be applied to both Bundesliga football and NBA basketball.

In terms of computational runtime for Bundesliga football, the Monte Carlo simulation for the ten key positions within the first division took approximately 2.2 days (52.7 hours) to complete the forthcoming importance calculations. This time was dramatically reduced with the *Overtake* approach, where a single season took just 33.5 minutes to complete the importance calculations for seventeen positions within the standings. For NBA basketball, the Monte Carlo simulation took 6.3 days (151.1 hours) to complete the forthcoming importance calculations for the top eight positions, while the *Overtake* approach took just 29.7 minutes per season to complete the importance calculations for the same top eight positions within both conferences. This highlights the reduced runtime required for the *Overtake* approach compared to the Monte Carlo simulation model.

6.2.2. Match outcome probabilities

For the Monte Carlo simulation model, two sets of match outcome probabilities will be applied. The first set is the static (constant) match outcome probabilities outlined in Chapter 5, which were applied across all teams in Bundesliga football and NBA basketball. The second set is determined by applying Bayes’ Rule, where the static match outcome
probabilities are adjusted according to the outcome/result ($X$) of the previous match for both team $i$ and their opponent ($k$). This is presented in (6.1).

$$P(X_{i,r}|X_{i,r-1},X_{k,r-1}) = \frac{P(X_{i,r-1},X_{k,r-1}|X_{i,r})P(X_{i,r})}{P(X_{i,r-1},X_{k,r-1})}$$  \hspace{1cm} (6.1)

The conditional probabilities required to adjust the static outcome probabilities using Bayes’ Rule in Bundesliga football, split by division and season period, are presented in Table 6.1. The conditional probabilities presented in Table 6.1 were determined by assessing the influence of each team’s previous match outcome on team $i$’s current match outcome across the nine observed seasons for each Bundesliga season; which was completed by determining the proportion of times that team $i$ would win/draw/lose their current match given the result of their and team $k$’s previous match.
Table 6.1: Bayesian conditional probabilities for Bundesliga football split by division and season period

The conditional probabilities for NBA basketball, split by season period, are presented in Table 6.2. Like the Bundesliga, the conditional probabilities in Table 6.2 were calculated by assessing the influence of each team’s previous match on team \( i \)'s current match outcome across all nine observed NBA seasons.
<table>
<thead>
<tr>
<th>Games</th>
<th>Opp. Prev. Outcome</th>
<th>Team i previous outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>P(W</td>
</tr>
<tr>
<td>1-41</td>
<td>W</td>
<td>0.5441</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>0.4559</td>
</tr>
<tr>
<td>42-82</td>
<td>W</td>
<td>0.5442</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>0.4558</td>
</tr>
</tbody>
</table>

Table 6.2: Bayesian conditional probabilities for NBA basketball split by season period
6.3. Results

This section provides a graphical comparison between the Overtake approach and the Monte Carlo simulation model; completed by comparing the average positional importance distributions for the key end-of-season outcomes. This includes applying both the static (constant) and varying match outcome probabilities, which will be completed for both the Monte Carlo simulation procedure and the Overtake approach. This section is split into two sub-sections: Bundesliga football and NBA basketball.

6.3.1. Bundesliga football comparison

Figure 6.1 presents the average importance distributions for the top four positions within first division Bundesliga football. Four models are included: Overtake with static probabilities, Overtake with varying probabilities, Monte Carlo with static probabilities, and Monte Carlo with varying probabilities. An initial observation from Figure 6.1 is that the two distributions are not identical to each other. However, both models produce the same conclusion regarding the average importance across the season: the average importance for each position is at a maximum at the start of the season before steadily declining towards the conclusion. The confirmation of this decreasing importance trend using the Monte Carlo simulation model suggests that the Overtake approach is reasonably approximating the importance of a match by position in Bundesliga football as it is providing a similar result to the computer simulation procedure.

In terms of applying Bayes’ Rule to vary the match outcome probabilities, the overall effect on the average positional importance distributions appears to be minimal. For the Overtake approach, a small amount of variation within the average importance distributions can be observed. However, the results have neither improved nor worsened; meaning the application of Bayes’ Rule to vary the match outcome probabilities has not drastically
changed the model’s results. As for the Monte Carlo model, a change is unrecognisable between the model types, with the importance distributions presenting as nearly identical across the season for each position.

Figure 6.1: Comparison between Overtake approach and Monte Carlo simulation with two sets of match outcome probabilities for top-four first division Bundesliga positions
Figure 6.2: Comparison between *Overtake* approach and Monte Carlo simulation with two sets of match outcome probabilities for bottom-four first division Bundesliga positions

Figure 6.2 provides the same comparison between the models for the lower positions within first division Bundesliga football. For the relegation positions (15-17), a large amount of movement can be observed within the two Monte Carlo simulation models. While there appears to be a greater amount of movement within these models, the overall trend is similar to the results from the *Overtake* approach. The comparatively low scale of each of these
positions should also be noted, which is due to the low number of teams in contention for these positions throughout the season.

Comparing the use of two sets of match outcome probabilities, the results are similar to those observed with the top four positions. Both the static and varying match outcome probabilities produce similar importance distributions for both the Overtake approach and Monte Carlo simulation. While a small amount of deviation can be observed, a distinct change in the distributions is not identifiable. These results across the positions cast doubt on the use of Bayes’ Rule to vary the match outcome probabilities within both model types. Similar results are also observed within the second division Bundesliga, where select key positions (1, 2, 15 and 16) are presented in Figure 6.3.
6.3.2. NBA basketball comparison

Figure 6.4 presents the average importance distributions for the top four positions from each conference in NBA basketball. There are six models included: East Overtake with static probabilities, East Overtake with varying probabilities, West Overtake with static probabilities, West Overtake with varying probabilities, Monte Carlo with static probabilities,
and Monte Carlo with varying probabilities. Note that the Monte Carlo simulation model is not split by conference as generic teams are used within the calculations. This means that the same result would be produced if the simulation process was completed for the East and West. Therefore, a single Monte Carlo simulation model, with both static and varying probabilities, is provided.

Figure 6.4: Comparison between Overtake approach and Monte Carlo simulation with two sets of match outcome probabilities for top-four NBA basketball positions
Like the Bundesliga results, the average positional importance from the *Overtake* approach does not follow an identical distribution to the Monte Carlo simulation model; where the average importance for the former tends to be lower than the latter. However, both models draw the same conclusion: the average importance across positions is at a maximum at the start of the season before steadily declining towards the conclusion. If Monte Carlo simulation is viewed as the most complete method for evaluating the importance of a match, then the *Overtake* approach can be seen as a good alternative that provides a reasonable approximation of the importance across the season. This is also completed while reducing the information required during calculations, since the *Overtake* approach focuses only on the current team of interest and the current team in position $s$.

In terms of varying the match outcome probabilities by applying Bayes’ Rule, there is marginal change in the importance distribution for the Monte Carlo simulation model. For the *Overtake* approach, a slight variation in the importance distributions can be observed across all four positions. However, the overall change within both the *Overtake* and Monte Carlo models is minimal, with the two sets of match outcome probabilities still drawing the same conclusion regarding when the most important match for each position, on average, occurs throughout a season of NBA basketball. Since this result is similar to those observed in Bundesliga football, it can be concluded that applying Bayes’ Rule to account for the current form of teams is not an effective method for defining team-specific match outcome probabilities. This is discussed further in the next section of this chapter.
6.4. Discussion

In this chapter, a Monte Carlo simulation model was applied to both Bundesliga football and NBA basketball. The objective was to validate the application of the Overtake approach from Chapter 5 by comparing the results to those observed through a complete Monte Carlo simulation procedure. The results from both sports indicated that the reduction in model inputs and computational complexity of the Overtake approach provided similar positional importance distributions to the Monte Carlo simulation model. While the results were not identical, they still demonstrated that the new Overtake provides a reasonable approximation of the average match importance generated from the simulation model.

As observed in Section 6.3, the average importance by position for each model within both sports followed a similar distribution, with the average importance across teams reaching a maximum at the commencement of the season before declining towards the conclusion. However, a key observation from Bundesliga football was that the average importance for the relegation positions was consistent across the season; which was different to the results from the higher positions. This result was due to the consistently low number of teams in contention for the relegation positions across the season, which would cause the average importance for each position to be low (yet consistent) across the Bundesliga season.

One observation from the NBA results is that the average positional importance distributions generated from the Overtake approach are lower than those from the Monte Carlo simulation model. This result may be due to the generic nature of teams within the simulation model, where there are no distinct dominant teams like those in the observed NBA seasons. Therefore, the number of teams in contention for position $s$ would be greater than those calculated from the observed seasons in Chapter 5. However, the similarity in results between the two models indicates that the Overtake approach is still reasonably
approximating the importance of a match, albeit on a lower scale to the Monte Carlo simulation model.

A secondary objective of this chapter was to determine if a variation in the match outcome probabilities via Bayes’ Rule would have an effect on the results of both models within the two sports. Bayes’ Rule was applied to adjust the constant (static) match outcome probabilities to account for the current form of the competing teams. However, the results from both sports found that there was little change in the average importance distributions. It follows that applying Bayes’ Rule to account for the current form of teams is an ineffective approach for creating team-specific match outcome probabilities. Therefore, future research on the Overtake approach should focus on varying the match outcome probabilities to account for other critical factors, such as home ground advantage and team strength. This could be completed using different statistical modelling techniques, such as a bivariate Poisson distribution (Dixon & Coles, 1997); or applying the pre-match bookmaker odds (Forrest & Simmons, 2008).

One of the key assumptions when calculating the importance of a match using the Overtake approach is the independence of match results of two teams. This assumption was not present within the Monte Carlo simulation model, as the results between all teams is considered when completing the match importance calculations. Since the results between the two models are similar for both sports, it can be concluded that a relaxation of the dependence of match results within the Overtake approach does not have a pronounced effect on the overall match importance calculations. Therefore, the Overtake approach is suitable for application within both Bundesliga football and NBA basketball as it provides similar results to the Monte Carlo simulation model while reducing the computational complexity and relaxing key calculation assumptions.
As discussed in Chapter 5, a limitation of the *Overtake* approach is that it can only calculate the importance of the next match in the season and not future contests; which is not the case within the Monte Carlo simulation model. However, as detailed in this chapter, the two models produce similar results and hence, the *Overtake* approach can be applied in place of the Monte Carlo simulation model without a dramatic loss of key information. This would be advantageous to those seeking a simple method of calculating the importance for a team’s next match without having to implement a computer simulation model. Nevertheless, if the desire is to calculate the importance of future matches, the Monte Carlo simulation is a favourable alternative to the *Overtake* approach.

While the Monte Carlo simulation model was applied as a basis for comparison with the *Overtake* approach, it would be interesting to compare the latter to other measures of match importance, such as the ad hoc approach by Jennett (1984). A comparison with other established measures could further validate the use of the *Overtake* approach to calculate the importance of a match in both Bundesliga football and NBA basketball. Furthermore, comparing the models within other sports, such as English Premier League football or Australian Rules football, would help create a definitive conclusion regarding both the Monte Carlo simulation and *Overtake* models. However, the results presented in this chapter still validate the use of the *Overtake* approach for quantifying the importance of a match within Bundesliga football and NBA basketball. This validation means that the match importance from the *Overtake* approach can now be applied within sports modelling, which is explored in the next two chapters of this dissertation.
6.5. Summary

In this chapter, a comparison was completed between the *Overtake* approach and a Monte Carlo simulation model to determine if the former provides similar results to the latter while reducing the number of required model inputs and computational complexity. The average importance distributions from both NBA basketball and Bundesliga football demonstrated that the *Overtake* approach produces similar results to those observed from the Monte Carlo simulation. Furthermore, there was minimal deviation in the importance distributions when the match outcome probabilities were varied by applying Bayes’ Rule. Despite this, the similarity in results validates the application of the *Overtake* approach for calculating the importance of a match in both sports. This validation allows for the match importance from the *Overtake* approach to now be assessed within sports modelling, which is completed in the next two chapters.
Chapter 7

Elo ratings

In this chapter, the match importance generated from the Overtake approach is applied to several versions of the Elo ratings system to determine if it has an effect on the overall predictive accuracy of the model. Different importance categories based on the end-of-season outcomes in both Bundesliga football and NBA basketball will be applied to three variations of the Elo ratings system; with the results being split according to season period and the level of match importance to the competing teams. This chapter aims to identify the circumstances in which the match outcome predictions are affected by the match importance.

This chapter is split into the following sections: Section 7.1 introduces the Elo ratings system and provides a brief discussion of the application of Elo ratings within the current literature. Section 7.2 details the three variations of the Elo ratings system that will be applied to both Bundesliga football and NBA basketball. Section 7.3 provides an analysis on the change in two performance metrics when matches are split by their level of importance and season period. Section 7.4 will provide a critical discussion of the observed results and finally, Section 7.5 summarises and concludes the chapter.
7.1. Introduction

In Chapter 6, the suitability of the Overtake approach for quantifying the importance of a match in both Bundesliga football and NBA basketball was validated through a comparison with a Monte Carlo simulation model. Since the validation of this new measure has been completed, the application of the Overtake approach within statistical modelling can now be assessed. In this chapter, the match importance generated from the Overtake approach is applied to the Elo ratings system to determine if it has an effect on the overall predictive accuracy of the model.

Originally developed to rate the overall ability of chess players (Elo, 1978), the Elo ratings system has been frequently applied to professional sports to model the strength of teams, such as Australian Rules football (Ryall & Bedford, 2010) and international football (soccer) (Leitner et al., 2010). Since the Elo ratings system determines the strength of a team, they are often applied to predict the outcome of sporting contests, where the team with a higher pre-match rating is considered the favourite to win. The model has also been further developed; with Glickman (1999) presenting a non-iterative updating algorithm that improves large paired comparison experiments, and Herbrich, Minka, and Graepel (2006) demonstrating that a Bayesian skill rating algorithm provides an increase in accuracy and convergence speed for rating online players compared to a standard Elo model. A common exclusion from existing Elo ratings systems is the importance of a match, with only the World Football Elo rating system including a simplistic approach to distinguish between World Cup matches and international friendly’s (Lasek et al., 2015).

As detailed in Chapter 1, match importance has frequently been applied in past research as an explanatory variable within a predictive model, including Bivariate Poisson models (Goddard, 2005) and probit models (Forrest et al., 2005; Goddard & Asimakopoulos, 2004);
where it has been commonly been defined as a binary value. While significant results have been found in past research (Goddard, 2005), this application of match importance may not be the best approach for determining the potential effect that it may have on a predictive model. Furthermore, the use of a binary value to signify the match importance does not allow for the level of importance to the two competing teams to be considered; where matches of differing levels of importance to the two competing teams may affect the predictive ability of a statistical model.

The *Overtake* approach, however, can be applied to create different levels of match importance as it provides a scale for the importance to be measured on. This can be completed by binning the importance values into three groups, where each group represents a different level of importance with respect to a team achieving a desirable end-of-season outcome, such as qualifying for Champions League in Bundesliga football. For example, one group can be described as “high” importance, which means that the match is of critical importance to the team of interest with respect to achieving an end-of-season outcome. While past research has concentrated on models such as a Bivariate Poisson model, this chapter will focus on the Elo ratings system as little research has been conducted on the effect of match importance on this type of statistical model.

In this chapter, the potential effect that the match importance has on the predictive accuracy of the Elo ratings system will be investigated in both Bundesliga football and NBA basketball. The match importance, generated from the *Overtake* approach, will not be included as a variable within the Elo model. Instead, matches will be binned according to their level of importance to both the pre-match “favourite” and “underdog”. Three variations of the Elo ratings system will be assessed within both sports, as well as different summations of the positional importance. The predictive accuracy of the Elo ratings system will be
assessed using two different performance metrics, which are detailed in the next section of this chapter.
7.2. Methods

This section details the application of the Elo ratings system within both Bundesliga football and NBA basketball. The additional two variations of the system are explained, since the baseline model for both sports was detailed in Chapter 3. This section also presents details for the different summations of the positional importance, and includes information about the process for binning matches according to their level of importance to the competing teams.

7.2.1. Bundesliga models

As detailed in Chapter 3, the Elo ratings system takes the following form:

\[
R_t^H = R_{t-1}^H + k(O^H - E^H) \quad (7.1)
\]

The methodology for the base Elo ratings system for Bundesliga football is presented in Chapter 3 of this dissertation. The second Elo model will incorporate a Pythagorean projection (James, 1987) that alters the observed result of the match \((O^H)\) to reflect the actual scores of both the home and away teams. This is presented in (7.2). Note that, although research has suggested that the optimal value for the exponent in (7.2) for football is approximately equal to 1.82 (Miller, 2007), the original Pythagorean projection formula as outlined in James (1987) is applied for simplicity.

\[
O^H = \frac{Score^H + 1}{Score^H + Score^A + 2} \quad (7.2)
\]

The third Elo ratings system for Bundesliga football is an existing approach conceived by Hvattum and Arntzen (2010). In this approach, the \(k\) value is adjusted depending on the absolute observed end-of-match goal difference between teams \((\delta)\). This approach allows teams to be rewarded for larger margins of victory. The adjustment is presented in (7.3), where \(\lambda\) is set to 1 as completed in Hvattum and Arntzen (2010).
\[ k_{GD} = k(1 + \delta)^2 \] (7.3)

As detailed in Chapter 3, the \( c \) and \( d \) values of the Elo ratings system are set to 10 and 400, respectively. The initial ratings for teams in division one and division two are set to 1,500 and 1,200, respectively. While the first value is commonly applied as the initial Elo rating (as described in Chapter 3), the second value was chosen arbitrarily as a way of distinguishing the two divisions by the strength of their teams (where it is reasonable to assume that division two comprises of weaker teams than division one). Finally, the \( k \) and \( h \) values are determined by maximising the log-likelihood (see equation (3.7) in Chapter 3) across the secondary data set outlined in Section 3.1.1 of this dissertation.

There are a number of methods available for evaluating the predictive performance of the models. Hvattum and Arntzen (2010) applied a number of loss functions to assess their Elo ratings model within association football. This included the Brier score (Brier, 1950), which measures the mean square difference between the predicted probability of the possible outcomes and the actual outcome; and the pseudo-likelihood statistic (Rue & Salvesen, 2000), which determines the average informational loss over a set of matches. Bailey and Clarke (2004) suggested three methods, including the absolute difference between expected and actual margins, the return on investment (ROI), and the percentage of correctly classified matches. Since a new application of match importance is being assessed within this chapter, a simple performance metric would be beneficial when evaluating the overall results. Therefore, the percentage of correctly classified matches and the Brier score will be the primary performance metrics assessed.

The Brier score is applied to statistical models to determine the accuracy of probabilistic predictions (Brier, 1950). The score is calculated using the difference between the observed outcome of an event \((o)\) and the predicted probability of that event occurring \((f)\).
Typically the former is restricted to binary values, where a match that has been correctly predicted takes the value of 1, and an incorrect prediction takes the value of 0. The Brier score is measured on a 0-1 scale, with values close to zero indicating that the model has good predictive accuracy. The formula for calculating the Brier score is presented in (7.4), where \( N \) is the total number of observations being assessed.

\[
Brier = \frac{1}{N} \sum_{i=1}^{N} (f_i - o_i)^2
\]  

(7.4)

Table 7.1 presents the optimised model values, the overall correct prediction percentages, and the Brier scores for the three Elo ratings systems. Note that the three Elo ratings systems are optimised separately for both the first and second division, as it was found to produce the highest overall correct prediction percentage within both divisions. As observed, the overall correct prediction percentage within each model is relatively poor, with only around half the number of matches being correctly classified. This is due to the difficulty of predicting a drawn match, which occurs across approximately 26% of the included Bundesliga matches. Further, the Elo ratings system can only be used to predict binary outcomes, meaning it will only predict the winner and loser of a match, never a drawn outcome. These values will serve as the basis for comparison when the matches are binned by their level of importance to the competing teams.
<table>
<thead>
<tr>
<th></th>
<th>Log-likelihood</th>
<th>$k$</th>
<th>$h$</th>
<th>%</th>
<th>Brier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Div. 1</td>
<td>Base</td>
<td>-170.95</td>
<td>25.89</td>
<td>67.20</td>
<td>49.79</td>
</tr>
<tr>
<td></td>
<td>Pythag.</td>
<td>-179.03</td>
<td>36.71</td>
<td>41.93</td>
<td>49.87</td>
</tr>
<tr>
<td></td>
<td>Goal Diff.</td>
<td>-169.95</td>
<td>7.96</td>
<td>68.97</td>
<td>49.35</td>
</tr>
<tr>
<td>Div. 2</td>
<td>Base</td>
<td>-176.53</td>
<td>14.68</td>
<td>81.75</td>
<td>46.65</td>
</tr>
<tr>
<td></td>
<td>Pythag.</td>
<td>-181.76</td>
<td>27.99</td>
<td>47.86</td>
<td>46.32</td>
</tr>
<tr>
<td></td>
<td>Goal Diff.</td>
<td>-176.32</td>
<td>4.68</td>
<td>82.17</td>
<td>46.24</td>
</tr>
</tbody>
</table>

Table 7.1: Bundesliga Elo ratings model parameters. overall correct prediction percentage and Brier scores

To determine the effect of match importance on the Elo ratings system, different importance categories will be created. The importance categories relate to the different end-of-season outcomes associated with each position in the Bundesliga standings (see Chapter 2); which will be calculated through a summation of the positional importance from the Overtake approach. For example, since positions 2-4 in division one Bundesliga football result in qualification for Champions League, the summation of the importance for these positions will equate to the ‘Champions League’ category. Note that position one is not included within the ‘Champions League’ category as this position also rewards a team with the Bundesliga championship. This position will be assessed separately in the ‘League Champion’ category. Position two is included within the second division ‘League Champion’ category as both these positions result in automatic promotion to division one; where it can be argued that this is more desirable to a team than winning the second division championship. The positions that will be contained within each importance category are...
presented in Table 7.2. The three weighted-sum approaches, Positional Sum, Outcome Sum and Total Sum, which were detailed in Chapter 5, will also be assessed.

<table>
<thead>
<tr>
<th>League</th>
<th>League Champion</th>
<th>Champions League</th>
<th>Europa League</th>
<th>Promotion</th>
<th>Relegation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bundesliga</td>
<td>1</td>
<td>2-4</td>
<td>5-7</td>
<td>15-16</td>
<td></td>
</tr>
<tr>
<td>2. Bundesliga</td>
<td>1-2</td>
<td></td>
<td></td>
<td>3</td>
<td>15-16</td>
</tr>
</tbody>
</table>

Table 7.2: Bundesliga importance categories with required positions split by division

After this has been completed, teams within each match are determined to be either the pre-match favourite to win (FAV), or the pre-match underdog (UND); which is accomplished by assessing the expected result of the match \( E^H \) contained within the Elo ratings calculations. The importance values within each category across the seasons are then binned into three levels of match importance: high (H), medium (M) and low (L). This process is completed separately for both the pre-match favourite and underdog. The binning process is completed using the visual binning technique in IBM SPSS Statistics 23, which evaluates the data to create three groups, each consisting of approximately 33.33% of the importance values. The binning cut-off values for the Bundesliga Elo ratings models are presented in Appendix E.

Note that the ‘Relegation’ category has only four classifications due to the compressed importance values contained within the category. After the importance categories for each team are binned, each match is allocated a classification based on the level of match importance for both the pre-match favourite and underdog. For example, if the pre-match favourite has a high importance (H) and the underdog has a low importance (L), then the
match is classified as HL. If the level of importance for the teams is reversed, then the match is classified as LH. This process is completed for all importance categories, where a total of nine match classifications per category are created. To determine the effect of the match importance on the Elo ratings system, the change in the correct prediction percentage and the Brier score within these match classifications will be assessed.

### 7.2.2. NBA models

The methodology for the NBA base Elo ratings system is presented in Chapter 3 of this dissertation, while the second model for the NBA will incorporate the Pythagorean projection presented in (7.2). The third and final Elo ratings system will again be an existing model. This model was conceived by fivethirtyeight.com (FiveThirtyEight.com, 2015). The FiveThirtyEight model introduces a margin of victory \((MOV)\) variable that is multiplied onto the \(k\) value. This is presented in (7.5).

\[
MOV^H = \frac{(\delta + 3)^{0.8}}{7.5 + 0.006(R_{t-1}^H - R_{t-1}^A)} \tag{7.5}
\]

Note that \(\delta\) is equal to the absolute end-of-match score difference. The FiveThirtyEight model also includes a season-to-season carry over, where only a proportion of a team’s end-of-season rating is carried over to the next season. From the model, 75% of a team’s end-of-season rating is carried over to the next season, where it is summed to 25% of the initial rating. For example, if a team’s end-of-season rating is equal to 1750 and the initial rating for each team was set to 1,500, then the team’s starting value within the new season is equal to 1,687.5.

Like the Bundesliga football models, the \(c\) and \(d\) values for each NBA Elo model are set to 10 and 400, respectively. The initial rating for all teams is set to 1,500, while the \(k\) and \(h\) values are determined by maximising the log-likelihood (see equation (3.7) in Chapter 3).
across the secondary data set outlined in Section 3.1.2 of this dissertation. The optimised values for each model, along with the overall correct prediction percentages and the Brier scores, are presented in Table 7.3.

<table>
<thead>
<tr>
<th></th>
<th>Log-likelihood</th>
<th>k</th>
<th>h</th>
<th>%</th>
<th>Brier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>-645.65</td>
<td>28.64</td>
<td>85.89</td>
<td>67.42</td>
<td>0.21</td>
</tr>
<tr>
<td>Pythag.</td>
<td>-726.30</td>
<td>30.09</td>
<td>12.25</td>
<td>67.93</td>
<td>0.24</td>
</tr>
<tr>
<td>FiveThirtyEight</td>
<td>-640.41</td>
<td>24.74</td>
<td>88.54</td>
<td>67.45</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Table 7.3: NBA Elo model parameters, overall correct prediction percentage and Brier Scores

Like the Bundesliga models, different importance categories are created for the NBA based on a summation of the positional importance. Along with the three weighted-sum approaches (Positional Sum, Outcome Sum, and Total Sum) detailed in Chapter 5, the NBA importance categories focus on winning the conference championship (‘Conference Champion’), finishing within the top four positions (‘Top 4’), and finishing within the bottom half of the top eight (‘Positions 5-8’). Each category is then binned into high/medium/low levels of importance for both the pre-match favourite and underdog; which is completed separately for each conference. The cut-off values for each level of importance for each Elo model split by conference and importance category is presented in Appendix E.

After this process is completed, each match receives a classification similar to the Bundesliga models. Since the binning process is completed separately for each conference, match classifications consist of East-only teams, West-only teams, and a mixture of the two conferences (inter-conference matches). Therefore, the results on the percentage of correctly
classified matches are split depending on whether the match is played between East teams, West teams, or a combination of the two.
7.3. Results

This section presents the results for the Elo ratings systems when matches are classified by their level of importance to both the pre-match favourite and underdog. The results for both Bundesliga football and NBA basketball are split by season period to explore different situations where match importance may affect the overall correct prediction percentage and the Brier score. In terms of naming conventions, if match classification ‘HM’ is being assessed in period two, then the scenario is referred to as ‘HM2’.

The primary focus of this section will be to assess the correct prediction percentage, with the Brier score results also briefly detailed. For both sports, it is hypothesised that a shift in the prediction percentage will correspond with a change in the level of importance for either the favourite or underdog. For example, it is anticipated that the correct prediction percentage will increase when the match is of higher importance to the favourite, such as in classifications HM and ML. When the match is of lower importance to the favourite, such as classifications MH or LM, it is anticipated that the correct prediction percentage will decrease. This hypothesised outcome is referenced throughout this section for both sports.

7.3.1. Bundesliga

The analysis of the Bundesliga football results commence with an evaluation of the correct prediction percentages for each importance category. Excluding the first round of the season due to no pre-match information being available for importance calculations, the 33 rounds are split evenly into three periods; with period one consisting of rounds two through twelve, period two consisting of rounds 13 through 23, and period three consisting of rounds 24 through 34. The correct prediction percentage for each importance category, split by match classification and season period, for all three division one Elo models are presented in
Table 7.4, along with the overall correct prediction percentage for each season period. The results for division two are presented in Table 7.5.
<table>
<thead>
<tr>
<th>League</th>
<th>Champion-League</th>
<th>Champion-League</th>
<th>Europa-League</th>
<th>Relegation</th>
<th>Residential-Ban</th>
<th>Outcome-Ban</th>
<th>Final-Ban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ill</td>
<td>30.26 (24)</td>
<td>58.08 (38)</td>
<td>55.56 (14)</td>
<td>57.14 (11)</td>
<td>45.75 (33)</td>
<td>44.35 (23)</td>
<td>51.46 (15)</td>
</tr>
<tr>
<td>IHM</td>
<td>51.38 (17)</td>
<td>64.12 (13)</td>
<td>53.52 (31)</td>
<td>62.20 (29)</td>
<td>50.88 (122)</td>
<td>42.11 (38)</td>
<td>51.73 (14)</td>
</tr>
<tr>
<td>Ill</td>
<td>50.30 (51)</td>
<td>63.04 (51)</td>
<td>57.14 (31)</td>
<td>68.42 (21)</td>
<td>45.09 (100)</td>
<td>52.61 (57)</td>
<td>38.59 (36)</td>
</tr>
<tr>
<td>ML</td>
<td>57.00 (17)</td>
<td>56.28 (20)</td>
<td>55.19 (14)</td>
<td>46.07 (15)</td>
<td>40.09 (34)</td>
<td>41.67 (182)</td>
<td>52.50 (40)</td>
</tr>
<tr>
<td>MH</td>
<td>47.20 (12)</td>
<td>46.09 (326)</td>
<td>47.06 (17)</td>
<td>49.45 (17)</td>
<td>45.45 (526)</td>
<td>50.46 (48)</td>
<td>49.66 (177)</td>
</tr>
<tr>
<td>MH</td>
<td>43.52 (216)</td>
<td>36.51 (61)</td>
<td>46.09 (12)</td>
<td>41.38 (51)</td>
<td>45.77 (52)</td>
<td>52.30 (42)</td>
<td>48.16 (175)</td>
</tr>
<tr>
<td>LL</td>
<td>47.74 (107)</td>
<td>47.56 (629)</td>
<td>46.43 (87)</td>
<td>47.14 (44)</td>
<td>55.09 (79)</td>
<td>58.28 (259)</td>
<td>50.89 (557)</td>
</tr>
<tr>
<td>LM</td>
<td>33.98 (108)</td>
<td>32.98 (28)</td>
<td>51.23 (28)</td>
<td>40.27 (47)</td>
<td>43.28 (175)</td>
<td>64.95 (59)</td>
<td>57.71 (30)</td>
</tr>
<tr>
<td>L1</td>
<td>25.00 (4)</td>
<td>57.14 (14)</td>
<td>55.19 (14)</td>
<td>46.07 (15)</td>
<td>40.09 (34)</td>
<td>41.67 (182)</td>
<td>52.50 (40)</td>
</tr>
<tr>
<td>Overall</td>
<td>48.15 50.17 51.07</td>
<td>48.15 50.17 51.07</td>
<td>48.15 50.17 51.07</td>
<td>48.15 50.17 51.07</td>
<td>48.15 50.17 51.07</td>
<td>48.15 50.17 51.07</td>
<td></td>
</tr>
<tr>
<td>Peru</td>
<td>50.81 (24)</td>
<td>58.56 (90)</td>
<td>55.56 (14)</td>
<td>57.14 (11)</td>
<td>45.75 (33)</td>
<td>44.35 (23)</td>
<td>51.46 (15)</td>
</tr>
<tr>
<td>IHM</td>
<td>50.30 (51)</td>
<td>63.04 (51)</td>
<td>57.14 (31)</td>
<td>68.42 (21)</td>
<td>45.09 (100)</td>
<td>52.61 (57)</td>
<td>38.59 (36)</td>
</tr>
<tr>
<td>Ill</td>
<td>49.31 (907)</td>
<td>60.06 (64)</td>
<td>53.63 (61)</td>
<td>68.42 (11)</td>
<td>47.90 (918)</td>
<td>50.98 (22)</td>
<td>30.56 (58)</td>
</tr>
<tr>
<td>ML</td>
<td>57.00 (17)</td>
<td>56.28 (20)</td>
<td>55.19 (14)</td>
<td>46.07 (15)</td>
<td>40.09 (34)</td>
<td>41.67 (182)</td>
<td>52.50 (40)</td>
</tr>
<tr>
<td>MH</td>
<td>46.08 (123)</td>
<td>52.01 (320)</td>
<td>47.06 (17)</td>
<td>49.45 (17)</td>
<td>45.45 (526)</td>
<td>50.46 (48)</td>
<td>49.66 (177)</td>
</tr>
<tr>
<td>MH</td>
<td>47.08 (232)</td>
<td>38.36 (88)</td>
<td>46.09 (12)</td>
<td>41.38 (51)</td>
<td>45.77 (52)</td>
<td>52.30 (42)</td>
<td>48.16 (175)</td>
</tr>
<tr>
<td>LL</td>
<td>47.12 (106)</td>
<td>47.56 (633)</td>
<td>46.31 (88)</td>
<td>45.21 (44)</td>
<td>50.60 (79)</td>
<td>78.24 (258)</td>
<td>40.53 (535)</td>
</tr>
<tr>
<td>LM</td>
<td>34.11 (123)</td>
<td>36.36 (22)</td>
<td>51.23 (28)</td>
<td>40.27 (47)</td>
<td>43.28 (175)</td>
<td>64.95 (59)</td>
<td>57.71 (30)</td>
</tr>
<tr>
<td>L1</td>
<td>25.00 (4)</td>
<td>57.14 (14)</td>
<td>55.19 (14)</td>
<td>46.07 (15)</td>
<td>40.09 (34)</td>
<td>41.67 (182)</td>
<td>52.50 (40)</td>
</tr>
<tr>
<td>Overall</td>
<td>48.26 50.39 50.95</td>
<td>48.26 50.39 50.95</td>
<td>48.26 50.39 50.95</td>
<td>48.26 50.39 50.95</td>
<td>48.26 50.39 50.95</td>
<td>48.26 50.39 50.95</td>
<td></td>
</tr>
<tr>
<td>Goal-Def</td>
<td>50.30 (51)</td>
<td>58.08 (38)</td>
<td>55.56 (14)</td>
<td>57.14 (11)</td>
<td>45.75 (33)</td>
<td>44.35 (23)</td>
<td>51.46 (15)</td>
</tr>
<tr>
<td>IHM</td>
<td>51.22 (41)</td>
<td>63.36 (29)</td>
<td>53.52 (31)</td>
<td>62.20 (29)</td>
<td>49.25 (122)</td>
<td>51.35 (15)</td>
<td>51.35 (35)</td>
</tr>
<tr>
<td>Ill</td>
<td>49.41 (52)</td>
<td>56.91 (62)</td>
<td>54.72 (33)</td>
<td>61.11 (10)</td>
<td>46.06 (1030)</td>
<td>47.75 (57)</td>
<td>50.88 (122)</td>
</tr>
<tr>
<td>ML</td>
<td>56.00 (58)</td>
<td>54.78 (48)</td>
<td>54.78 (48)</td>
<td>20.00 (3)</td>
<td>49.09 (60)</td>
<td>56.49 (14)</td>
<td>51.49 (149)</td>
</tr>
<tr>
<td>MH</td>
<td>49.54 (121)</td>
<td>48.42 (106)</td>
<td>53.81 (34)</td>
<td>40.14 (15)</td>
<td>42.11 (76)</td>
<td>52.17 (49)</td>
<td>48.41 (177)</td>
</tr>
<tr>
<td>LL</td>
<td>45.03 (103)</td>
<td>47.57 (648)</td>
<td>50.00 (88)</td>
<td>42.23 (42)</td>
<td>54.79 (79)</td>
<td>78.24 (255)</td>
<td>50.89 (535)</td>
</tr>
<tr>
<td>LM</td>
<td>32.08 (125)</td>
<td>46.09 (32)</td>
<td>50.00 (32)</td>
<td>46.09 (32)</td>
<td>52.12 (520)</td>
<td>52.45 (205)</td>
<td>46.01 (228)</td>
</tr>
<tr>
<td>L1</td>
<td>60.00 (5)</td>
<td>44.69 (9)</td>
<td>56.21 (57)</td>
<td>46.47 (97)</td>
<td>55.56 (36)</td>
<td>64.06 (64)</td>
<td>49.61 (228)</td>
</tr>
<tr>
<td>Overall</td>
<td>48.26 49.53 50.95</td>
<td>48.26 49.53 50.95</td>
<td>48.26 49.53 50.95</td>
<td>48.26 49.53 50.95</td>
<td>48.26 49.53 50.95</td>
<td>48.26 49.53 50.95</td>
<td></td>
</tr>
</tbody>
</table>

Sample size corrected to participants

Table 7.4: Bundesliga division one correct prediction percentage by Elo type, match classification, season period and importance category
| League Champions | Promotion | Relegation | Professional | Overall | Total | League | LM | LL | MH | HL | LL | MH | HL | LL | MH | HL | LL | MH | HL | LL | MH | HL | LL | MH | HL |
|------------------|----------|-----------|--------------|---------|-------|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Bundesliga | 55.13 (70) | 2.00 (9) | 43.27 (20) | 82.10 (37) | 39.33 (14) | 52.99 (32) | 50.00 (2) | 58.76 (7) | 61.45 (23) | 48.40 (107) | 54.84 (62) | 39.39 (15) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) |
| English Premier  | 45.93 (137) | 0.00 (0) | 57.24 (13) | 61.91 (30) | 56.80 (46) | 55.86 (35) | 50.00 (2) | 46.43 (133) | 54.01 (124) | 58.76 (7) | 61.45 (23) | 48.40 (107) | 54.84 (62) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) |
| La Liga | 50.12 (236) | 37.95 (80) | 20.41 (27) | 51.39 (75) | 56.92 (26) | 53.16 (135) | 57.97 (217) | 46.99 (257) | 44.65 (118) | 45.59 (248) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) |
| Premier League | 51.89 (106) | 52.00 (135) | 39.20 (143) | 55.29 (120) | 49.42 (80) | 45.38 (15) | 46.30 (26) | 48.40 (107) | 54.84 (62) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) |
| 2. Bundesliga | 50.98 (2) | 38.51 (161) | 50.43 (351) | 62.86 (35) | 50.00 (16) | 53.03 (66) | 53.33 (15) | 52.81 (89) | 46.75 (77) | 47.90 (157) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) | 50.00 (16) | 41.81 (17) | 40.00 (25) | 56.74 (76) | 42.41 (197) | 48.40 (107) | 54.84 (62) |

Table 7.5: Bundesliga division two correct prediction percentage by Elo type, match classification, season period and importance category
While the results for both divisions are presented, the primary focus will be on the first division. In Table 7.4, an initial observation is that the three Elo models produce similar results across the different match classifications and importance categories. For example, all three models produce a correct prediction percentage of 70.83% with a sample size of 24 for HL2 of the ‘League Champion’ category. While there are some classifications that differ, such as HH3 of the ‘Relegation’ category, the similarity of results for a majority of classifications suggest that the adjustments to the Bundesliga Elo ratings systems have not provided an enhanced result. Due to this similarity, the focus of the analysis will shift to the first division base Elo model.

In the ‘League Champion’ category, a change in the correct prediction percentage was observed when the match was of high importance to the favourite, with the greatest increase occurring within HL2 (+20.66%, \( n=24 \)). When the importance of the match for the favourite shifts from high to medium, the prediction percentages converge toward the overall, with a decrease of 13.66% \( (n=63) \) observed within MH2. Finally, the prediction percentages drop below the overall when the match is of low importance to the favourite, with LH2 recording a correct prediction percentage of 25%, albeit with a small sample size \( (n=4) \). A similar decreasing trend was also observed within period three of the category, where an increase in the correct prediction percentage corresponded with a high level of match importance for the favourite.

To compare the results between the categories that have a specific end-of-season outcome, the difference between the observed and the overall correct prediction percentage for the first four categories are presented in Figure 7.1. Note that period one is excluded due to some classifications observing a sample size of zero. In Figure 7.1, the aforementioned decreasing trend across the match classifications for the ‘League Champion’ category is evident within period two. A similar decreasing trend is less obvious within the remaining
three categories, where the prediction percentage differences across the classifications are close to zero. However, the hypothesised result (where the importance is greater for one team) is still observed within certain match classifications, such as HL3 (+4.49%, \( n=90 \)) and MH3 (15.6%, \( n=12 \)) within the ‘Champions League’ category. Nonetheless, the obvious trend within the ‘League Champion’ category indicates that a distinct change in the prediction percentage is more pronounced when focusing solely on the importance of finishing in first position.

![Figure 7.1: Base Elo change in prediction percentage for periods two and three of outcome-specific importance categories for first division Bundesliga](image)

<table>
<thead>
<tr>
<th>Match classification</th>
<th>Percentage difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period 2</td>
<td></td>
</tr>
<tr>
<td>HL</td>
<td>-30</td>
</tr>
<tr>
<td>HM</td>
<td>-15</td>
</tr>
<tr>
<td>HH</td>
<td>0</td>
</tr>
<tr>
<td>ML</td>
<td>15</td>
</tr>
<tr>
<td>MM</td>
<td>30</td>
</tr>
<tr>
<td>MH</td>
<td>-30</td>
</tr>
<tr>
<td>LL</td>
<td>-15</td>
</tr>
<tr>
<td>LM</td>
<td>0</td>
</tr>
<tr>
<td>LH</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Match classification</th>
<th>Percentage difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period 3</td>
<td></td>
</tr>
<tr>
<td>HL</td>
<td>-30</td>
</tr>
<tr>
<td>HM</td>
<td>-15</td>
</tr>
<tr>
<td>HH</td>
<td>0</td>
</tr>
<tr>
<td>ML</td>
<td>15</td>
</tr>
<tr>
<td>MM</td>
<td>30</td>
</tr>
<tr>
<td>MH</td>
<td>-30</td>
</tr>
<tr>
<td>LL</td>
<td>-15</td>
</tr>
<tr>
<td>LM</td>
<td>0</td>
</tr>
<tr>
<td>LH</td>
<td>15</td>
</tr>
</tbody>
</table>

Figure 7.1: Base Elo change in prediction percentage for periods two and three of outcome-specific importance categories for first division Bundesliga
Figure 7.2 provides a graphical comparison of the percentage difference for the three weighted-sum categories, ‘Positional Sum’, ‘Outcome Sum’ and ‘Total Sum’, for periods two and three. While a distinct trend across the match classifications was not obvious, the hypothesised result still occurs under certain circumstances. For example, the correct prediction percentage decreases below the overall for ‘Positional Sum’ when the match is of low importance to the favourite during both periods two and three. However, for LH3 of ‘Total Sum’, the prediction percentage actually increases, meaning that the opposite of the hypothesised result has occurred.
The most obvious change in the predictive accuracy occurs within the ‘League Champion’ category so this is will be the focus of the Bundesliga results from here on. To apply the Brier score to the category, the expected win probability for the favourite (\(E^H\)) generated from the Elo ratings model is first binned into three equal size groups for each season period. Matches are binned according to whether the favourite has a low advantage (bin 1), a medium advantage (bin 2), or a high advantage (bin 3) over the underdog. A Brier
score is then calculated for each bin within each period, where the average expected win probability for each bin is applied within the calculations (variable $f$ in (7.3)). Once a base score for each period and bin has been established, a Brier score is calculated for each match classification within each bin and period. The base Brier score, along with the correct prediction percentage, for each season period, bin and match classification is presented in Table 7.6 and Table 7.7 for first and second division Bundesliga, respectively.

A lower Brier score paired with an improvement in predictive accuracy provides evidence that the improvement is reliable. While not a statistically significant result, the use of this predictive accuracy metric, combined with reasonable sample sizes, allows the focus to shift to the more reliable improvements in predictive accuracy of the Elo ratings system. There are circumstances where the predictive accuracy improves but it not met with a low Brier score and thus, as a result, it is difficult to accept the improvement on face value. Therefore, the focus is on those that are reliable.

While not evident for every match classification, there are scenarios where an improvement in the Brier score is observed within the first division Bundesliga. For example, HM of bin two within period two produces a Brier score of 0.17 ($n=44$), which is lower than the base Brier score for the period and bin (0.28). This result suggests that when the favourite has a medium advantage in terms of Elo expected win probability, the predictive accuracy of the model is improved when the match importance for the favourite is greater than the underdog. This is also reflected within Table 7.4, which observes an improvement in the prediction percentage of approximately 14% for HM2 of the ‘League Champion’ category. Other noteworthy results that are supported by a large sample size ($n>30$) in Table 7.8 include MM of bin three within period one (0.23 compared to base Brier score of 0.27), and HM of bin three within period two (0.19 compared to base Brier score of 0.22). A similar comparison can be completed with the second division Bundesliga results in Table 7.7.
<table>
<thead>
<tr>
<th>Favourite Bin</th>
<th>Period one</th>
<th>Period two</th>
<th>Period three</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Avg. Elo predictive home W%</td>
<td>0.54</td>
<td>0.64</td>
<td>0.76</td>
</tr>
<tr>
<td>Base Brier</td>
<td>0.26</td>
<td>0.28</td>
<td>0.27</td>
</tr>
<tr>
<td>Correct prediction %</td>
<td>40.47</td>
<td>44.22</td>
<td>59.73</td>
</tr>
<tr>
<td>HL Brier</td>
<td>0.29</td>
<td>0.21</td>
<td>0.09</td>
</tr>
<tr>
<td>%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM Brier</td>
<td>0.28</td>
<td>0.33</td>
<td>0.19</td>
</tr>
<tr>
<td>%</td>
<td>38.46 (13)</td>
<td>41.66 (12)</td>
<td>75.00 (12)</td>
</tr>
<tr>
<td>HH Brier</td>
<td>0.25</td>
<td>0.27</td>
<td>0.26</td>
</tr>
<tr>
<td>%</td>
<td>42.69 (171)</td>
<td>47.05 (170)</td>
<td>60.46 (172)</td>
</tr>
<tr>
<td>ML Brier</td>
<td>0.26</td>
<td>0.25</td>
<td>0.22</td>
</tr>
<tr>
<td>%</td>
<td>47.36 (19)</td>
<td>57.89 (19)</td>
<td>68.42 (19)</td>
</tr>
<tr>
<td>MM Brier</td>
<td>0.26</td>
<td>0.29</td>
<td>0.23</td>
</tr>
<tr>
<td>%</td>
<td>38.09 (42)</td>
<td>39.02 (41)</td>
<td>64.28 (42)</td>
</tr>
<tr>
<td>MH Brier</td>
<td>0.26</td>
<td>0.26</td>
<td>0.30</td>
</tr>
<tr>
<td>%</td>
<td>31.94 (72)</td>
<td>50.70 (71)</td>
<td>47.94 (73)</td>
</tr>
<tr>
<td>LL Brier</td>
<td>0.26</td>
<td>0.31</td>
<td>0.24</td>
</tr>
<tr>
<td>%</td>
<td>44.44 (36)</td>
<td>34.28 (35)</td>
<td>61.11 (36)</td>
</tr>
<tr>
<td>LM Brier</td>
<td>0.26</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>%</td>
<td>30.00 (40)</td>
<td>28.20 (39)</td>
<td>43.58 (39)</td>
</tr>
<tr>
<td>LH Brier</td>
<td>0.25</td>
<td>0.38</td>
<td>0.24</td>
</tr>
<tr>
<td>%</td>
<td>50.00 (2)</td>
<td>0.00 (2)</td>
<td>66.66 (3)</td>
</tr>
</tbody>
</table>

Sample size presented in parenthesis

Table 7.6: Brier score comparison for Bundesliga Division One ‘League Champion’ importance classifications split by season period and Elo favourite expected win probability bin
<table>
<thead>
<tr>
<th></th>
<th>Period one</th>
<th></th>
<th>Period two</th>
<th></th>
<th>Period three</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Favourite Bin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Elo predictive home W%</td>
<td>0.54</td>
<td>0.62</td>
<td>0.71</td>
<td>0.54</td>
<td>0.62</td>
<td>0.72</td>
</tr>
<tr>
<td>Base Brier</td>
<td>0.26</td>
<td>0.28</td>
<td>0.23</td>
<td>0.26</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>Correct prediction %</td>
<td>35.35</td>
<td>43.58</td>
<td>64.43</td>
<td>38.38</td>
<td>41.02</td>
<td>51.51</td>
</tr>
<tr>
<td><strong>Avg. Elo predictive home W%</strong></td>
<td>0.54</td>
<td>0.62</td>
<td>0.72</td>
<td>0.54</td>
<td>0.62</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>Base Brier</strong></td>
<td>0.26</td>
<td>0.28</td>
<td>0.23</td>
<td>0.26</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td><strong>Correct prediction %</strong></td>
<td>35.35</td>
<td>43.58</td>
<td>64.43</td>
<td>38.38</td>
<td>41.02</td>
<td>51.51</td>
</tr>
</tbody>
</table>

Table 7.7: Brier score comparison for Bundesliga Division Two ‘League Champion’ importance classifications split by season period and Elo favourite expected win probability bin

Sample size presented in parenthesis
7.3.2. NBA

The results for NBA basketball are split into eight season periods instead of three, which was completed due to the length of a NBA standard season. Each period consists of ten matches excluding the final period, which includes eleven matches. Like the Bundesliga football models, the first match of the season was excluded as there is no pre-match information available to complete the importance calculations. Since the NBA schedule does not follow a specific round structure, matches are excluded if they fall into a different season period for the two competing teams. Therefore, a total of 1,348 matches are excluded from the analysis.

As mentioned in Section 7.2, the results for the NBA Elo models are split into three sub-groups: East, West and inter-conference. Due to the large number of season periods, the results for each model are split into three tables consisting of two importance categories. This section will primarily focus on the base NBA Elo results, with the results for the remaining models presented in Appendix F. Table 7.8 present the prediction percentages for categories ‘Conference Champion’ and ‘Top 4’; Table 7.9 present the results for ‘Positions 5-8’ and ‘Positional Sum’; and Table 7.10 present the results the results for ‘Outcome Sum’ and ‘Total Sum’.
<table>
<thead>
<tr>
<th>East</th>
<th>Overall</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>80.80 (2)</td>
<td>71.19 (37)</td>
<td>65.96 (36)</td>
<td>70.00 (11)</td>
<td>100.00 (2)</td>
<td>100.00 (1)</td>
<td>71.80 (11)</td>
<td></td>
</tr>
<tr>
<td>HM</td>
<td>65.47 (34)</td>
<td>75.36 (139)</td>
<td>70.19 (61)</td>
<td>58.26 (38)</td>
<td>72.08 (27)</td>
<td>58.33 (11)</td>
<td>70.00 (12)</td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>40.00 (1)</td>
<td>70.00 (50)</td>
<td>75.64 (76)</td>
<td>77.72 (11)</td>
<td>73.33 (98)</td>
<td>80.77 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH</td>
<td>54.24 (8)</td>
<td>63.87 (135)</td>
<td>64.71 (85)</td>
<td>79.17 (26)</td>
<td>61.54 (17)</td>
<td>75.00 (6)</td>
<td>40.00 (1)</td>
<td></td>
</tr>
<tr>
<td>MH</td>
<td>57.45 (94)</td>
<td>50.31 (185)</td>
<td>58.16 (18)</td>
<td>42.06 (14)</td>
<td>78.59 (17)</td>
<td>50.00 (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td>50.00 (2)</td>
<td>56.02 (29)</td>
<td>70.29 (311)</td>
<td>82.32 (178)</td>
<td>62.46 (255)</td>
<td>66.39 (159)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>50.00 (6)</td>
<td>51.72 (29)</td>
<td>70.00 (16)</td>
<td>55.56 (9)</td>
<td>100.00 (3)</td>
<td>66.67 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LH</td>
<td>100.00 (1)</td>
<td>100.00 (3)</td>
<td>57.14 (7)</td>
<td>50.00 (6)</td>
<td>57.14 (7)</td>
<td>80.00 (5)</td>
<td>70.00 (12)</td>
<td>76.02 (13)</td>
</tr>
<tr>
<td>Overall</td>
<td>62.15</td>
<td>67.22</td>
<td>66.96</td>
<td>61.96</td>
<td>75.07</td>
<td>71.68</td>
<td>78.19</td>
<td>64.81</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>West</th>
<th>Overall</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>36.80 (18)</td>
<td>38.33 (20)</td>
<td>34.00 (2)</td>
<td>55.31 (18)</td>
<td>35.59 (18)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM</td>
<td>66.00 (35)</td>
<td>66.00 (35)</td>
<td>63.60 (6)</td>
<td>50.00 (10)</td>
<td>64.71 (17)</td>
<td>68.33 (15)</td>
<td>66.67 (9)</td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>80.00 (5)</td>
<td>75.17 (41)</td>
<td>78.00 (75)</td>
<td>80.77 (76)</td>
<td>75.34 (98)</td>
<td>75.76 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH</td>
<td>58.99 (55)</td>
<td>72.67 (126)</td>
<td>67.06 (45)</td>
<td>65.00 (45)</td>
<td>66.40 (19)</td>
<td>100.00 (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH</td>
<td>66.30 (71)</td>
<td>62.10 (139)</td>
<td>58.33 (65)</td>
<td>59.52 (42)</td>
<td>58.08 (26)</td>
<td>64.29 (14)</td>
<td>100.00 (1)</td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td>60.00 (1)</td>
<td>57.06 (28)</td>
<td>56.98 (36)</td>
<td>67.66 (14)</td>
<td>70.91 (206)</td>
<td>60.00 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>70.00 (10)</td>
<td>55.17 (29)</td>
<td>46.15 (39)</td>
<td>64.61 (17)</td>
<td>38.46 (13)</td>
<td>66.67 (12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LH</td>
<td>50.00 (2)</td>
<td>100.00 (1)</td>
<td>75.00 (6)</td>
<td>100.00 (3)</td>
<td>100.00 (6)</td>
<td>67.74 (1)</td>
<td>59.26 (27)</td>
<td>61.96 (21)</td>
</tr>
<tr>
<td>Overall</td>
<td>66.04</td>
<td>66.12</td>
<td>78.54</td>
<td>67.07</td>
<td>65.48</td>
<td>71.96</td>
<td>69.32</td>
<td>78.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inter</th>
<th>Overall</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>83.72 (4)</td>
<td>76.47 (85)</td>
<td>73.03 (78)</td>
<td>54.55 (33)</td>
<td>54.55 (33)</td>
<td>100.00 (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM</td>
<td>67.90 (44)</td>
<td>69.95 (205)</td>
<td>65.79 (29)</td>
<td>62.16 (17)</td>
<td>66.67 (21)</td>
<td>50.00 (6)</td>
<td>25.00 (4)</td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>58.33 (12)</td>
<td>60.68 (80)</td>
<td>73.00 (14)</td>
<td>71.20 (111)</td>
<td>70.20 (47)</td>
<td>87.58 (8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH</td>
<td>74.50 (55)</td>
<td>63.91 (106)</td>
<td>67.44 (128)</td>
<td>74.47 (47)</td>
<td>79.17 (24)</td>
<td>100.00 (7)</td>
<td>100.00 (1)</td>
<td></td>
</tr>
<tr>
<td>MH</td>
<td>66.96 (122)</td>
<td>68.83 (162)</td>
<td>67.06 (85)</td>
<td>64.08 (26)</td>
<td>48.48 (17)</td>
<td>69.23 (11)</td>
<td>75.66 (12)</td>
<td>50.00 (2)</td>
</tr>
<tr>
<td>LL</td>
<td>100.00 (1)</td>
<td>66.67 (42)</td>
<td>61.97 (42)</td>
<td>65.79 (28)</td>
<td>67.85 (31)</td>
<td>65.79 (316)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>47.62 (21)</td>
<td>71.18 (52)</td>
<td>65.29 (29)</td>
<td>70.00 (50)</td>
<td>73.33 (98)</td>
<td>100.00 (6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LH</td>
<td>50.00 (2)</td>
<td>50.00 (18)</td>
<td>62.50 (53)</td>
<td>33.33 (13)</td>
<td>70.00 (10)</td>
<td>33.33 (3)</td>
<td>70.31 (29)</td>
<td>82.43 (26)</td>
</tr>
<tr>
<td>Overall</td>
<td>67.36</td>
<td>69.98</td>
<td>66.45</td>
<td>66.75</td>
<td>69.26</td>
<td>67.73</td>
<td>67.77</td>
<td>67.77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conference Champions</th>
<th>Top 4</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Inter</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

<p>| Table 7.8: NBA base Elo prediction percentage split by season period and match classification for ‘Conference Champion’ and ‘Top 3’ categories |</p>
<table>
<thead>
<tr>
<th>Positions</th>
<th>5%</th>
<th>10%</th>
<th>15%</th>
<th>20%</th>
<th>25%</th>
<th>30%</th>
<th>35%</th>
<th>40%</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IH</td>
<td>60.68 (67)</td>
<td>57.69 (64)</td>
<td>67.31 (62)</td>
<td>70.67 (61)</td>
<td>70.67 (61)</td>
<td>70.67 (61)</td>
<td>70.67 (61)</td>
<td>70.67 (61)</td>
</tr>
<tr>
<td>IH</td>
<td>54.41 (134)</td>
<td>51.34 (131)</td>
<td>61.29 (128)</td>
<td>64.62 (125)</td>
<td>64.62 (125)</td>
<td>64.62 (125)</td>
<td>64.62 (125)</td>
<td>64.62 (125)</td>
</tr>
<tr>
<td>IH</td>
<td>71.72 (14)</td>
<td>68.67 (11)</td>
<td>88.67 (8)</td>
<td>98.89 (5)</td>
<td>98.89 (5)</td>
<td>98.89 (5)</td>
<td>98.89 (5)</td>
<td>98.89 (5)</td>
</tr>
<tr>
<td>LM</td>
<td>78.57 (14)</td>
<td>75.00 (10)</td>
<td>66.67 (7)</td>
<td>76.67 (4)</td>
<td>76.67 (4)</td>
<td>76.67 (4)</td>
<td>76.67 (4)</td>
<td>76.67 (4)</td>
</tr>
<tr>
<td>MH</td>
<td>65.31 (9)</td>
<td>62.00 (6)</td>
<td>71.57 (5)</td>
<td>72.73 (4)</td>
<td>72.73 (4)</td>
<td>72.73 (4)</td>
<td>72.73 (4)</td>
<td>72.73 (4)</td>
</tr>
<tr>
<td>MH</td>
<td>70.83 (9)</td>
<td>68.00 (6)</td>
<td>76.00 (5)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
</tr>
<tr>
<td>LH</td>
<td>61.63 (114)</td>
<td>58.00 (86)</td>
<td>70.00 (58)</td>
<td>82.00 (41)</td>
<td>82.00 (41)</td>
<td>82.00 (41)</td>
<td>82.00 (41)</td>
<td>82.00 (41)</td>
</tr>
<tr>
<td>LH</td>
<td>63.16 (11)</td>
<td>60.00 (8)</td>
<td>72.00 (5)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
</tr>
<tr>
<td>LH</td>
<td>66.04</td>
<td>63.00 (6)</td>
<td>75.00 (4)</td>
<td>87.00 (3)</td>
<td>87.00 (3)</td>
<td>87.00 (3)</td>
<td>87.00 (3)</td>
<td>87.00 (3)</td>
</tr>
<tr>
<td>LH</td>
<td>59.38 (18)</td>
<td>56.00 (12)</td>
<td>68.00 (8)</td>
<td>78.00 (4)</td>
<td>78.00 (4)</td>
<td>78.00 (4)</td>
<td>78.00 (4)</td>
<td>78.00 (4)</td>
</tr>
<tr>
<td>LH</td>
<td>62.86 (26)</td>
<td>60.00 (16)</td>
<td>72.00 (8)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
<td>84.00 (4)</td>
</tr>
<tr>
<td>LH</td>
<td>58.82 (17)</td>
<td>55.00 (10)</td>
<td>67.00 (7)</td>
<td>79.00 (4)</td>
<td>79.00 (4)</td>
<td>79.00 (4)</td>
<td>79.00 (4)</td>
<td>79.00 (4)</td>
</tr>
<tr>
<td>LH</td>
<td>66.67 (9)</td>
<td>63.33 (6)</td>
<td>75.00 (4)</td>
<td>87.00 (3)</td>
<td>87.00 (3)</td>
<td>87.00 (3)</td>
<td>87.00 (3)</td>
<td>87.00 (3)</td>
</tr>
<tr>
<td>LH</td>
<td>69.77 (86)</td>
<td>66.67 (58)</td>
<td>78.67 (41)</td>
<td>90.67 (28)</td>
<td>90.67 (28)</td>
<td>90.67 (28)</td>
<td>90.67 (28)</td>
<td>90.67 (28)</td>
</tr>
<tr>
<td>LH</td>
<td>74.14 (116)</td>
<td>71.43 (89)</td>
<td>83.43 (53)</td>
<td>95.43 (30)</td>
<td>95.43 (30)</td>
<td>95.43 (30)</td>
<td>95.43 (30)</td>
<td>95.43 (30)</td>
</tr>
<tr>
<td>LH</td>
<td>64.81 (108)</td>
<td>62.00 (70)</td>
<td>74.00 (43)</td>
<td>86.00 (26)</td>
<td>86.00 (26)</td>
<td>86.00 (26)</td>
<td>86.00 (26)</td>
<td>86.00 (26)</td>
</tr>
<tr>
<td>Overall</td>
<td>62.15</td>
<td>60.98 (50)</td>
<td>72.67 (40)</td>
<td>84.00 (25)</td>
<td>84.00 (25)</td>
<td>84.00 (25)</td>
<td>84.00 (25)</td>
<td>84.00 (25)</td>
</tr>
<tr>
<td>West</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL</td>
<td>65.85 (19)</td>
<td>62.96 (16)</td>
<td>75.00 (12)</td>
<td>87.00 (8)</td>
<td>87.00 (8)</td>
<td>87.00 (8)</td>
<td>87.00 (8)</td>
<td>87.00 (8)</td>
</tr>
<tr>
<td>IL</td>
<td>69.63 (21)</td>
<td>66.67 (17)</td>
<td>78.67 (13)</td>
<td>90.67 (7)</td>
<td>90.67 (7)</td>
<td>90.67 (7)</td>
<td>90.67 (7)</td>
<td>90.67 (7)</td>
</tr>
<tr>
<td>IL</td>
<td>66.41 (177)</td>
<td>63.33 (133)</td>
<td>75.33 (93)</td>
<td>87.33 (49)</td>
<td>87.33 (49)</td>
<td>87.33 (49)</td>
<td>87.33 (49)</td>
<td>87.33 (49)</td>
</tr>
<tr>
<td>IL</td>
<td>77.78 (9)</td>
<td>74.00 (6)</td>
<td>86.00 (3)</td>
<td>98.00 (1)</td>
<td>98.00 (1)</td>
<td>98.00 (1)</td>
<td>98.00 (1)</td>
<td>98.00 (1)</td>
</tr>
<tr>
<td>IL</td>
<td>69.00 (5)</td>
<td>65.33 (5)</td>
<td>77.33 (3)</td>
<td>89.33 (1)</td>
<td>89.33 (1)</td>
<td>89.33 (1)</td>
<td>89.33 (1)</td>
<td>89.33 (1)</td>
</tr>
<tr>
<td>IL</td>
<td>78.57 (27)</td>
<td>75.00 (18)</td>
<td>87.50 (11)</td>
<td>99.50 (5)</td>
<td>99.50 (5)</td>
<td>99.50 (5)</td>
<td>99.50 (5)</td>
<td>99.50 (5)</td>
</tr>
<tr>
<td>IL</td>
<td>75.31 (22)</td>
<td>72.00 (14)</td>
<td>84.00 (7)</td>
<td>96.00 (2)</td>
<td>96.00 (2)</td>
<td>96.00 (2)</td>
<td>96.00 (2)</td>
<td>96.00 (2)</td>
</tr>
<tr>
<td>IL</td>
<td>77.38 (13)</td>
<td>74.14 (7)</td>
<td>86.00 (3)</td>
<td>98.00 (1)</td>
<td>98.00 (1)</td>
<td>98.00 (1)</td>
<td>98.00 (1)</td>
<td>98.00 (1)</td>
</tr>
<tr>
<td>IL</td>
<td>64.81 (108)</td>
<td>62.00 (70)</td>
<td>74.00 (43)</td>
<td>86.00 (26)</td>
<td>86.00 (26)</td>
<td>86.00 (26)</td>
<td>86.00 (26)</td>
<td>86.00 (26)</td>
</tr>
<tr>
<td>Overall</td>
<td>62.15</td>
<td>60.98 (50)</td>
<td>72.67 (40)</td>
<td>84.00 (25)</td>
<td>84.00 (25)</td>
<td>84.00 (25)</td>
<td>84.00 (25)</td>
<td>84.00 (25)</td>
</tr>
</tbody>
</table>

**Table 7.9: NBA base Elo prediction percentage split by season period and match classification for 'Positions 5-8' and 'Positional Sum' categories**
### Table 7.10: NBA base Elo prediction percentage split by season period and match classification for ‘Outcome Sum’ and ‘Total Sum’ categories

<table>
<thead>
<tr>
<th>Season</th>
<th>Outcome Sum</th>
<th>Total Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>100.00 (%)</td>
<td>100.00 (%)</td>
</tr>
<tr>
<td>1-2</td>
<td>62.00 (14)</td>
<td>60.00 (14)</td>
</tr>
<tr>
<td>2-4</td>
<td>78.00 (15)</td>
<td>78.00 (15)</td>
</tr>
<tr>
<td>4-5</td>
<td>50.00 (14)</td>
<td>50.00 (14)</td>
</tr>
<tr>
<td>5-7</td>
<td>60.00 (14)</td>
<td>60.00 (14)</td>
</tr>
<tr>
<td>7-9</td>
<td>40.00 (14)</td>
<td>40.00 (14)</td>
</tr>
<tr>
<td>9-11</td>
<td>50.00 (14)</td>
<td>50.00 (14)</td>
</tr>
<tr>
<td>11-13</td>
<td>60.00 (14)</td>
<td>60.00 (14)</td>
</tr>
<tr>
<td>13-15</td>
<td>70.00 (14)</td>
<td>70.00 (14)</td>
</tr>
<tr>
<td>15-17</td>
<td>80.00 (14)</td>
<td>80.00 (14)</td>
</tr>
<tr>
<td>17-19</td>
<td>90.00 (14)</td>
<td>90.00 (14)</td>
</tr>
</tbody>
</table>

Sample size: predicted in brackets (in parenthesis)
Focusing on the ‘Conference Champion’ category within Table 7.8, a change in the prediction percentage is observed within periods four and five. In HL5, the prediction percentages for all three sub-groups increased above the overall for each period, which are all supported by large sample sizes. The greatest increase for HL5 was observed within the Western conference (+17.85%), which has a large sample size \((n=30)\). In MH5, a decrease in the prediction percentage was observed within the inter-conference group (-18.12%, \(n=33\)), which indicates that the underdog, independent of the conferences between the two teams, is winning more matches when the match is of slightly higher importance to them during the middle section of the season.

In the ‘Top 4’ category, similar results are observed within the middle periods of the season. In HL5 for the inter-conference group, the prediction percentage increases by 17.78% with a sample size of 32. A further increase within the HL classification is observed in periods six through eight for the Western conference, where the prediction percentage increases by as much as 16.04% within HL8 \((n=92)\). A decrease in the prediction percentage is observed for LH4 (-11.89%, \(n=31\)), which suggests that, when focusing on inter-conference matches, the underdog is performing when matches are of high importance to them with respect to finishing within the top four.

In Table 7.9, in the ‘Positions 5-8’ category, an increase in the correct prediction percentage is observed within the HL classification for both the West and inter-conference groups, where the greatest increase with a suitable sample size observed was within HL6 (+16.74%, \(n=41\)). For classification LH, the hypothesised result was not observed across a majority of the season periods for each sub-group, where the percentages increase instead of decrease. For example, LH7 for the inter-conference group produces a prediction percentage of 77.78% with a sample size of 36. While the hypothesised result does not occur within LH,
the results from HL indicate that a change in prediction percentage still occurs when focusing on the importance of finishing in positions five through eight.

For the three-weighted sum approaches, the expected outcome is observed within certain classifications. For example, for the ‘Positional Sum’ category, the correct prediction percentage increased by 15.94% \((n=63)\) within HL5 of the inter-conference group. Additionally, a decrease of 11.6% \((n=40)\) was observed within LH5 of the same sub-group. A decreasing trend across the classifications was also observed within some season periods, such as period four of the inter-conference group in the ‘Outcome Sum’ category, where the correct prediction percentage steadily decreased from HL \((77.78\%, n=27)\) to LH \((50\%, n=24)\).
To explore the change across the season periods, the difference in correct prediction percentage for each period within the ‘Conference Champion’ category, split by sub-group, is presented in Figure 7.3. Note that only the match classifications with a differing level of importance is included to allow the focus to turn to the hypothesized result. For the match classifications where the favourite has a greater importance (HL, HM, ML), the prediction
percentage across the season periods for each sub-group increased, which was expected given the favourite has more to play for than their underdog opponent. The opposite is also observed for classifications MH and LM, where the correct prediction percentage decreases across the season periods. For LH, the only decrease was observed within the inter-conference group, where the other two sub-groups observe either an increase or a sample size of zero across the periods.

Like the Bundesliga model, the Brier score is applied to the ‘Conference Champion’ category, where a base Brier score is calculated for each season period and bin. To simplify the presentation, the results for the Brier score have not been split by conference. The Brier score for periods one through four, along with the correct prediction percentage and same size, are presented in Table 7.11; and the results for periods five through eight are presented in Table 7.12. As completed with the Bundesliga model, a lower Brier score paired with an improvement in predictive accuracy provides evidence that the improvement is reliable. Again, while not a statistically significant result, the application allows for the focus to shift the more reliable results within the base NBA Elo ratings system.

Focusing on Table 7.11, there are scenarios where an improvement in the Brier score is observed. For example, classification ML of bin three within period four produces a Brier score of 0.14 ($n=51$), which is lower than the base brier score for the period/bin (0.14). This result also observes an increase in the prediction percentage (82.35% compared to the base of 79.69%), which demonstrates that when the favourite has a strong advantage in terms of Elo expected win probability, the prediction percentage of the model is improved when the match importance for the favourite is slightly greater than the underdog. In the scenarios where the match is of greater importance to the underdog, the Brier score increases. This includes classification MH of bin three within period three (0.21 compared to base of 0.18), which
demonstrates that the predictive accuracy of the Elo ratings system will decrease when the match is of higher importance to the underdog compared to the favourite.

In Table 7.12, similar results are observed within the match classifications for the periods five through eight. This includes classification HL, where a decrease in the Brier score is observed across a majority of the season periods and bins. This classification’s result means that the predictive accuracy of the Elo ratings system will improve during the second half of the season when the matches are of high importance to the favourite and low importance to the underdog with respect to finishing in first position. A similar conclusion can also be drawn from classification ML, where an improvement in the Brier score, along with a large sample size, is observed in all but one season period (period eight). In the final season period, the largest sample sizes are observed within classification LL, which follows the observed results from Chapter 5 where the average importance across teams in contention for position one decreased towards the conclusion of the season.
<table>
<thead>
<tr>
<th></th>
<th>Period one</th>
<th></th>
<th></th>
<th>Period two</th>
<th></th>
<th></th>
<th></th>
<th>Period three</th>
<th></th>
<th></th>
<th>Period four</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Favourite Bin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Elo predictive home W%</td>
<td>0.57</td>
<td>0.69</td>
<td>0.84</td>
<td>0.57</td>
<td>0.71</td>
<td>0.85</td>
<td>0.56</td>
<td>0.69</td>
<td>0.85</td>
<td>0.56</td>
<td>0.69</td>
<td>0.84</td>
</tr>
<tr>
<td>Base Brier</td>
<td>0.25</td>
<td>0.24</td>
<td>0.18</td>
<td>0.25</td>
<td>0.23</td>
<td>0.16</td>
<td>0.24</td>
<td>0.22</td>
<td>0.18</td>
<td>0.25</td>
<td>0.24</td>
<td>0.16</td>
</tr>
<tr>
<td>Correct prediction %</td>
<td>53.35</td>
<td>63.50</td>
<td>77.49</td>
<td>55.53</td>
<td>65.83</td>
<td>80.65</td>
<td>60.41</td>
<td>66.24</td>
<td>76.86</td>
<td>54.29</td>
<td>62.34</td>
<td>79.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Elo predictive away W%</td>
<td>0.57</td>
<td>0.69</td>
<td>0.84</td>
<td>0.57</td>
<td>0.71</td>
<td>0.85</td>
<td>0.56</td>
<td>0.69</td>
<td>0.85</td>
<td>0.56</td>
<td>0.69</td>
<td>0.84</td>
</tr>
<tr>
<td>Base Brier</td>
<td>0.25</td>
<td>0.24</td>
<td>0.18</td>
<td>0.25</td>
<td>0.23</td>
<td>0.16</td>
<td>0.24</td>
<td>0.22</td>
<td>0.18</td>
<td>0.25</td>
<td>0.24</td>
<td>0.16</td>
</tr>
<tr>
<td>Correct prediction %</td>
<td>53.35</td>
<td>63.50</td>
<td>77.49</td>
<td>55.53</td>
<td>65.83</td>
<td>80.65</td>
<td>60.41</td>
<td>66.24</td>
<td>76.86</td>
<td>54.29</td>
<td>62.34</td>
<td>79.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Elo predictive home Brier</td>
<td>0.25</td>
<td>0.24</td>
<td>0.18</td>
<td>0.25</td>
<td>0.23</td>
<td>0.16</td>
<td>0.24</td>
<td>0.22</td>
<td>0.18</td>
<td>0.25</td>
<td>0.24</td>
<td>0.16</td>
</tr>
<tr>
<td>Correct prediction %</td>
<td>53.35</td>
<td>63.50</td>
<td>77.49</td>
<td>55.53</td>
<td>65.83</td>
<td>80.65</td>
<td>60.41</td>
<td>66.24</td>
<td>76.86</td>
<td>54.29</td>
<td>62.34</td>
<td>79.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Elo predictive away Brier</td>
<td>0.25</td>
<td>0.24</td>
<td>0.18</td>
<td>0.25</td>
<td>0.23</td>
<td>0.16</td>
<td>0.24</td>
<td>0.22</td>
<td>0.18</td>
<td>0.25</td>
<td>0.24</td>
<td>0.16</td>
</tr>
<tr>
<td>Correct prediction %</td>
<td>53.35</td>
<td>63.50</td>
<td>77.49</td>
<td>55.53</td>
<td>65.83</td>
<td>80.65</td>
<td>60.41</td>
<td>66.24</td>
<td>76.86</td>
<td>54.29</td>
<td>62.34</td>
<td>79.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Elo predictive home %</td>
<td>0.25</td>
<td>0.24</td>
<td>0.18</td>
<td>0.25</td>
<td>0.23</td>
<td>0.16</td>
<td>0.24</td>
<td>0.22</td>
<td>0.18</td>
<td>0.25</td>
<td>0.24</td>
<td>0.16</td>
</tr>
<tr>
<td>Correct prediction %</td>
<td>53.35</td>
<td>63.50</td>
<td>77.49</td>
<td>55.53</td>
<td>65.83</td>
<td>80.65</td>
<td>60.41</td>
<td>66.24</td>
<td>76.86</td>
<td>54.29</td>
<td>62.34</td>
<td>79.69</td>
</tr>
</tbody>
</table>

Sample size presented in parenthesis

Table 7.11: Brier score comparison for NBA ‘League Champion’ for periods one to four split by Elo favourite expected win probability bin.
| Favourite Bin | Period five | | | Period six | | | | Period seven | | | | Period eight | | |
|-------|-------------|---|---|-------------|---|---|-------------|---|---|-------------|---|---|-------------|---|---|
| | Avg. Elo predictive home W% | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| | Base Brier | | | | | | | | | | | | | |
| | Correct prediction % | | | | | | | | | | | | | |
| LH | Brier | 0.25 | 0.11 | 0.06 | 0.18 | 0.16 | 0.08 | 0.25 | 0.14 | 0.09 | 0.29 | 0.19 | 0.10 |
| | % | 57.14 (42) | 87.80 (41) | 92.85 (42) | 78.04 (41) | 78.94 (38) | 90.47 (42) | 53.19 (47) | 82.97 (47) | 89.13 (46) | 41.66 (24) | 73.91 (23) | 87.5 (24) |
| HM | Brier | 0.23 | 0.21 | 0.24 | 0.25 | 0.26 | 0.16 | 0.18 | 0.09 | 0.18 | 0.10 | 0.07 | |
| | % | 65.38 (26) | 69.23 (26) | 68.00 (25) | 45.45 (11) | 54.54 (11) | 80.00 (10) | 100.00 (4) | 100.00 (2) | 75.00 (4) | 100.00 (2) | 100.00 (2) | |
| HH | Brier | 0.24 | 0.18 | 0.25 | 0.23 | 0.23 | 0.29 | 0.23 | 0.26 | 0.24 | 0.19 | 0.11 | 0.39 |
| | % | 57.89 (19) | 78.94 (19) | 61.11 (18) | 66.66 (12) | 63.63 (11) | 50.12 (12) | 71.42 (7) | 50.00 (6) | 62.50 (8) | 100.00 (4) | 100.00 (3) | 25.00 (4) |
| ML | Brier | 0.22 | 0.18 | 0.14 | 0.23 | 0.15 | 0.12 | 0.25 | 0.22 | 0.06 | 0.21 | 0.11 | 0.18 |
| | % | 65.16 (89) | 76.4 (89) | 83.14 (89) | 62.79 (86) | 80.95 (84) | 84.88 (86) | 53.44 (58) | 68.96 (58) | 92.98 (57) | 72.22 (18) | 88.23 (17) | 77.77 (18) |
| MM | Brier | 0.23 | 0.21 | 0.16 | 0.23 | 0.19 | 0.20 | 0.24 | 0.21 | 0.14 | 0.11 | 0.09 | 0.50 |
| | % | 67.56 (37) | 69.44 (36) | 78.94 (38) | 68.42 (19) | 73.68 (19) | 72.22 (18) | 57.14 (7) | 71.42 (7) | 83.33 (6) | 100.00 (2) | 100.00 (1) | 0.00 (1) |
| MH | Brier | 0.24 | 0.27 | 0.21 | 0.24 | 0.24 | 0.10 | 0.26 | 0.19 | 0.18 | 0.27 | 0.09 | 0.31 |
| | % | 53.84 (26) | 40.00 (25) | 68.00 (25) | 58.33 (12) | 58.33 (12) | 91.66 (12) | 33.33 (6) | 80.00 (5) | 80.00 (5) | 50.00 (2) | 100.00 (1) | 50.00 (2) |
| LL | Brier | 0.24 | 0.22 | 0.18 | 0.24 | 0.22 | 0.16 | 0.24 | 0.23 | 0.11 | 0.24 | 0.22 | 0.17 |
| | % | 56.66 (120) | 64.40 (118) | 76.47 (119) | 58.94 (190) | 64.55 (189) | 78.83 (189) | 55.03 (258) | 64.20 (257) | 86.48 (259) | 56.48 (370) | 65.85 (369) | 77.62 (371) |
| LM | Brier | 0.24 | 0.23 | 0.3 | 0.24 | 0.23 | 0.15 | 0.24 | 0.24 | 0.23 | 0.25 | 0.19 | 0.22 |
| | % | 67.85 (28) | 62.96 (27) | 44.82 (29) | 52.63 (19) | 61.11 (18) | 84.21 (19) | 63.63 (11) | 54.54 (11) | 63.63 (11) | 50.00 (6) | 100.00 (5) | 66.66 (6) |
| LH | Brier | 0.24 | 0.25 | 0.22 | 0.25 | 0.22 | 0.32 | 0.24 | 0.20 | 0.09 | 0.23 | 0.23 | 0.18 |
| | % | 66.66 (3) | 50.00 (2) | 66.66 (3) | 50.00 (4) | 66.66 (3) | 33.33 (3) | 60.00 (5) | 75.00 (4) | 100.00 (5) | 100.00 (4) | 66.66 (3) | 80 (5) |

Table 7.12: Brier score comparison for NBA ‘League Champion’ for periods five to eight split by Elo favourite expected win probability bin

Sample size presented in parenthesis
7.4. Discussion

The primary objective of this chapter was to determine, in general terms, if match importance has a measureable effect on the predictive accuracy of the Elo ratings system. By classifying matches based on their level of importance to both the pre-match favourite and underdog, a shift in the correct prediction percentage and the Brier score was observed within both Bundesliga football and NBA basketball. While a number of different importance categories were applied to summarise the positional importance from the Overtake approach, the most pronounced effects were observed when solely focusing on finishing in first position.

For first division Bundesliga football, an increase in the base Elo prediction percentage was observed when the importance of a match was classified as high for the favourite; while a decrease was observed when a match was classified as low importance to the favourite. This decreasing trend indicated that the correct prediction percentage of the Elo ratings system could be affected by classifying matches based on their level of importance. However, the trend was less obvious within the other importance categories, such as ‘Champions League’ and ‘Europa League’, where the prediction percentage was consistent across the match classifications and season periods. Since these categories were based on a summation of the positional importance, the pronounced effect within ‘League Champion’ indicates that a change in the prediction percentage can be primarily observed when classifying matches based on a single importance value rather than a summation.

Similar results could be observed within the second division Bundesliga base Elo model. Despite having a lower overall prediction percentage (46.65% compared to 49.79% for first division), a distinct change was observed within certain match classifications, such as HL (+8.26%, n=52) and LH (-8.66%, n=40) within period two of the ‘League Champion’
category. Like the first division, the results are less pronounced within the other importance categories, such as period two of the ‘Total Sum’, where no match classification differs greater than 2.8% from the period’s overall percentage (43.66%). The less obvious trend within the remaining importance categories supports the previous statement that classifying matches based on a single value, instead of a summation, leads to a more pronounced effect.

In the NBA basketball models, the results were split into three sub-groups: East, West and inter-conference. Within each sub-group, a change in the correct prediction percentage of the base Elo model was observed in certain match classifications within different importance categories. The most pronounced effects were observed during the middle section of the season within the ‘Conference Champion’ category, where the correct prediction percentage increased when the match was of greater importance to the favourite compared to the underdog. Furthermore, the promising results within the middle section of the season suggests that, when applying the match importance to an NBA Elo model, the greatest effect on the prediction percentage will occur during the middle period of the season.

In terms of season periods within both sports, the most pronounced effects occurred within the middle stages of the season. As detailed in Section 7.3, a decreasing trend across the match classifications was observed within period two of the ‘League Champion’ category for first division Bundesliga football. For NBA basketball, positive results were observed within certain match classifications for periods four through five of the ‘Conference Champion’ category. While positive results were also observed within other season periods for both sports, the consistent findings from the middle periods of the season suggests that this is when the greatest effect on the prediction percentage is observed.

A common observation from both the Bundesliga football and NBA basketball models is that the results for the weighted-sum approaches were less pronounced than those observed
Within the outcome-specific categories. Furthermore, for each model, the most obvious change in the prediction percentage occurred when solely focusing on the importance of finishing in first position. The lack of distinct change in the correct prediction percentage across the classifications within the weighted-sum categories indicates that summarising all positional importance values is less effective than summing the positions with a common end-of-season outcome. Nevertheless, future research can explore alternative methods for summarising the match importance into a single value, such as calculating the maximum level of importance across the outcome-specific importance categories to reflect the overall importance of a match to a team.

As mentioned previously, three variations of the Elo ratings system was applied to both Bundesliga football and NBA basketball. The aim of applying different variations of the Elo model was to determine if they further enhanced the results of the base model. Despite the variations improving the overall correct prediction percentage, the sample sizes and prediction percentage of each match classification for the different importance categories was almost identical to the base model results. While minimal improvement was observed with the additional Elo models, future research can evaluate other variations of the model, such as a seasonal decay model (Ryall & Bedford, 2010), or further adjusting the model to account for the time period in which a match occurs (Stefani & Pollard, 2007).

In this chapter, the percentage of correctly classified matches and the Brier score were applied to evaluate the performance of the Elo ratings system within both Bundesliga football and NBA basketball. While these methods provided a simple approach for evaluating the models, they do not provide a statistically significant result. Future research could apply alternative performance metrics, including the evaluating the return on investment (ROI), and the absolute difference between the predicted and actual margins (Bailey & Clarke, 2004). However, the focus of this chapter was not to determine a betting strategy but rather
investigate the effect of match importance on the Elo ratings system. Therefore, the percentage of correctly classified matches and the Brier score were suitable performance metrics. Nevertheless, future research on the effect of match importance on the Elo ratings system could evaluate the performance of the model using an alternative method.

One drawback of applying the Elo ratings system to professional football is that the model does not successfully identify when a draw outcome will occur. The result of this is a lower prediction percentage (49.79% for first division Bundesliga) compared to a two-result sport such as NBA basketball (67.42%). One approach of improving the model within professional football is to predict that a draw will occur if the pre-match ratings difference between teams falls within some specified interval. This was trialled within the Bundesliga football models but the end result was a lower prediction percentage than the original. This result further highlights the difficulty of predicting the draw outcome in football, which has been thoroughly discussed in past research (Deschamps & Gergaud, 2007). While it has been shown in this chapter that classifying matches according to their level of importance can affect the prediction percentage, future research should focus on improving the suitability of the Elo ratings system to predict draw outcomes in football.

A further drawback of the Elo ratings systems applied within this chapter is the small amount of matches used to calibrate the model parameters. For both Bundesliga football and NBA basketball, a total of two seasons were optimised to determine the model parameters for the next nine seasons. Furthermore, for Bundesliga football, teams move between divisions through promotion and relegation; meaning that a team could have entered the division without having featured within the test seasons. Future research should increase the amount of matches within the test data in an attempt to improve the overall performance of the Elo ratings system for match prediction. Moreover, for football, future research should introduce
a more specific initial Elo rating for teams that do not originally feature within the test data to allow for a more accurate reflection of the team’s strength.

In terms of future application, classification of matches by importance to affect the predictive accuracy of the Elo ratings system could formulate a betting strategy in both Bundesliga football and NBA basketball. As mentioned, a decrease in the prediction percentage is observed when the match is of higher importance to the underdog compared to the favourite. In terms of a betting strategy, it would be advantageous for a betting person to avoid matches that fall into these classifications. Alternatively, they could bet against what is being predicted by the Elo ratings system. While these are two possible options, determining an optimal betting strategy using the match importance classifications is beyond the scope of this chapter, but would still make for an interesting future research topic.

While it has been established that there are circumstances where the match importance has an effect on the predictive accuracy of the Elo ratings system, the next step is to determine if the importance generated from the Overtake approach is a statistically significant predictor of the outcome of a match. This will be explored in the next chapter of this dissertation, where the different match importance categories that were detailed previously will be included as explanatory variables within a predictive model; which, as described in Chapter 1, is the most common application of match importance within past research. The binning process that was completed for the Elo ratings system will also be applied in the next chapter to further explore the effect of match importance on a model’s predictive ability.
7.5. Summary

In this chapter, an investigation into the possible effect that match importance has on the predictive accuracy of an Elo ratings system was completed. By classifying matches based on their level of importance to both the pre-match favourite and underdog, it was concluded that there were circumstances where the correct prediction percentage of the Elo model was affected in both Bundesliga football and NBA basketball. While the results were promising, further analysis on the match importance from the Overtake approach can be completed by re-visiting its application as an explanatory variable within a linear model.
Chapter 8

Regression analysis

In this chapter, the match importance generated from the *Overtake* approach is applied within a linear regression model to determine if it is a significant predictor of the final score difference between teams. This is completed for both Bundesliga football and NBA basketball, where a number of additional independent variables are included to complete the models. This chapter will also explore the possible effect of match importance on the regression model’s performance, which will be completed by binning matches according to their level of importance to the competing teams.

The chapter is broken down into the following sections: Section 8.1 discusses applying the importance of a match as a predictor variable within the current literature. Section 8.2 details the additional independent variables that are included within the regression models for both Bundesliga football and NBA basketball. Section 8.3 presents the results for the regression models, while Section 8.4 provides a critical discussion of the findings. Finally, Section 8.5 summarises and concludes the chapter.
8.1. Introduction

In Chapter 7, the general effect of match importance on the predictive accuracy of the Elo ratings system was explored. It was determined that the correct prediction percentage of an Elo ratings system could be affected by the level of match importance to both the pre-match favourite and underdog. While the results were promising, the next step is to evaluate the match importance generated from the Overtake approach as an explanatory variable within a predictive model. Particularly, the focus will be on determining if the match importance is a significant predictor of the final score difference in both Bundesliga football and NBA basketball; which is the primary aim of this chapter.

A number of past research studies have explored modelling the outcome of a sporting contest using different statistical procedures. A common approach has been to apply a Poisson distribution to model the scores of the competing teams. Maher (1982) applied independent Poisson distributions to model the goals scored and conceded by English football teams, while Dixon and Coles (1997) and Dixon and Pope (2004) both elected to apply a bivariate Poisson distribution to model the same outcomes. Karlis and Ntzoufras (2003) also applied a bivariate Poisson model but instead elected to model the final score difference between teams instead of the total goals scored. However, when defining the model parameters within the Poisson distribution, match importance has often been overlooked in favour of team-specific metrics, such as offensive and defensive measures.

Another common statistical approach for modelling outcomes is to apply an ordered probit regression model (Audas et al., 2002; Forrest & Simmons, 2000). Goddard and Asimakopoulos (2004) both applied the model to forecast the results of English football matches, concluding that the significance of the match positively contributed to the model’s performance. This approach was compared to a Poisson distribution model by Goddard
(2005), where it was concluded that there was little difference in terms of forecasting performance between the methods, and a binary match importance variable was a significant determinant of the match outcome for both models. Other statistical approaches include logistic regression, where Morley and Thomas (2005) found that the importance of a match was a significant determinant of English one-day county cricket contests; and generalised estimation equations, where Lei and Humphreys (2013) found that match importance had a significant relationship with the outcome of Major League Baseball matches.

Past literature has also focused on modelling the attendance at sporting events while applying the importance of a match as an explanatory variable (Forrest & Simmons, 2006; Jennett, 1984). Borland and Lye (1992) evaluated match attendance in Australian Rules football, determining that the importance of a match had a positive effect on the crowd size. Determining match importance with respect to winning the championship or avoiding relegation, Dobson and Goddard (1992) found that the importance of a match for the home team had a positive effect on attendance within English football. A comprehensive review of modelling attendance figures can be found in Borland and MacDonald (2003) and Villar and Guerrero (2009).

While current literature has established that the importance of a match can be a significant determinant of match outcomes, binary variables have typically been applied to signify match importance within the models. Since the Overtake approach provides a scale for the importance to be measured on, it will be applied within a stepwise regression model to help increase the knowledge of scale measures of match importance. This type of statistical model will be assessed due to its originality compared to the established literature, where a majority of studies applied either a Poisson distribution or an ordered probit regression model (Goossens et al., 2012). Current literature has also elected to focus on the total goals scored and conceded by teams, with only a small number of studies modelling the end-of-match
score difference. Therefore, the dependent variable within this chapter will be the end-of-match score difference between teams.

The stepwise regression model will be applied to the primary data outlined in Chapter 3 for both sports. The primary objective is to determine whether the match importance generated from the Overtake approach is a significant predictor of the observed end-of-match score difference. To complete the models within each sport, additional independent variables that have commonly been applied within past literature will be included. A description of these additional variables, along with the different match importance variables, is presented in the next section of this chapter. A secondary objective of this chapter is to further explore the effect of match importance on the predictive ability of the regression model. This will be completed by classifying matches based on their level of importance to the competing teams; which was completed in the previous chapter when assessing the Elo ratings system.
8.2. Methods

For both Bundesliga football and NBA basketball, the importance of a match within the stepwise regression model will be denoted by the different summations of the positional importance detailed in Chapter 7. Additional independent variables will also be included to complete each regression model. These variables were selected based on their application within the existing literature, where a description of each variable is provided in Table 8.1. As mentioned in the previous section of this chapter, the dependent variable for both sports will be the end-of-match score difference.
The calculation of the scored-based variables within Table 8.1 is completed by assessing the match results across the primary data. Note that each variable is calculated for both the home (H) and away (A) teams where applicable. For both sports, the base Elo ratings from Chapter 7 will be applied for both teams, where the $ELO_{DIFF}$ will be equal to the home rating minus the away rating. Finally, the geographical distances between teams ($Distance$) for both sports were collected from www.sportmapworld.com.
The stepwise regression modelling technique will be applied to both Bundesliga football and NBA basketball, where the forward selection criteria will be implemented to reduce the amount of independent variables within the final model. A brief description of the forward selection criteria can be found in Chapter 3, while a detailed description is provided in Devore (2008). The coefficient of determination ($R^2$) will be assessed to determine the suitability of the final model to predict the end-of-match score difference ($\text{Score}_\text{Diff}$), where values close to one will indicate that there is a strong relationship between the selected independent variables and the response variable.

To further evaluate the results of the stepwise regression model, matches will be binned into importance classifications similar to the Elo ratings in Chapter 7. This approach will help determine if an improvement in the $R^2$ of a stepwise regression model can be observed when matches are binned by their importance. The binning process is completed for each importance category detailed in Chapter 7, where the classifications are created with respect to the home and away team instead of the pre-match favourite and underdog. Due to the similar binning results to Chapter 7, the cut-off values for the three importance levels are not presented within this chapter.

Since the response variable is the end-of-match score difference (measured in terms of the home team), the constant within the stepwise regression model can be viewed as a variable to distinguish the two competing teams. However, all information that distinguishes the two teams is captured within the additional variables presented in Table 8.1, such as the relative strength difference between teams being described by the Elo ratings difference ($\text{ELO}_\text{DIFF}$). Since this information is already contained within the independent variables, the constant would not provide any further information about the competing teams. Therefore, the constant is excluded from the stepwise regression models.
8.3. Results

This section details the forward stepwise regression results for both Bundesliga football and NBA basketball. For the former, the stepwise regression model is applied separately for both the first and second division. This section also provides the results for when matches are binned by their level of importance to the home and away teams. The descriptive statistics for all variables for both sports are presented in Table 8.2.
Table 8.2: Descriptive statistics of regression variables for both Bundesliga football and NBA basketball

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bundesliga</th>
<th>2. Bundesliga</th>
<th>NBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score_Diff</td>
<td>0.35 1.83</td>
<td>0.39 1.70</td>
<td>3.02 12.91</td>
</tr>
<tr>
<td>H_AvgGlsFor</td>
<td>1.44 0.53</td>
<td>1.34 0.49</td>
<td>H_AvgPtsFor</td>
</tr>
<tr>
<td>H_AvgGlsAg</td>
<td>1.46 0.50</td>
<td>1.37 0.47</td>
<td>H_AvgPtsAg</td>
</tr>
<tr>
<td>H_GlsForLast</td>
<td>1.28 1.20</td>
<td>1.17 1.13</td>
<td>H_PtsForLast</td>
</tr>
<tr>
<td>H_GlsAgForLast</td>
<td>1.58 1.31</td>
<td>1.51 1.28</td>
<td>H_PtsAgFor</td>
</tr>
<tr>
<td>H_Pts12</td>
<td>49.17 12.58</td>
<td>2.83 1.16</td>
<td>H_Pts12</td>
</tr>
<tr>
<td>H_Pts24</td>
<td>100.41 21.09</td>
<td>86.49 13.67</td>
<td>H_Pts24</td>
</tr>
<tr>
<td>H_Max5</td>
<td>2.93 1.22</td>
<td>2.76 1.16</td>
<td>A_Max5</td>
</tr>
<tr>
<td>H_Min5</td>
<td>0.26 0.49</td>
<td>0.20 0.43</td>
<td>A_Min5</td>
</tr>
<tr>
<td>A_AvgGlsFor</td>
<td>1.46 0.54</td>
<td>1.37 0.49</td>
<td>A_AvgGlsFor</td>
</tr>
<tr>
<td>A_AvgGlsAg</td>
<td>1.44 0.50</td>
<td>1.34 0.46</td>
<td>A_AvgGlsAg</td>
</tr>
<tr>
<td>A_GlsForLast</td>
<td>1.58 1.32</td>
<td>1.48 1.28</td>
<td>A_GlsForLast</td>
</tr>
<tr>
<td>A_GlsAgForLast</td>
<td>1.28 1.21</td>
<td>1.15 1.12</td>
<td>A_GlsAgForLast</td>
</tr>
<tr>
<td>A_Pts12</td>
<td>49.54 12.71</td>
<td>44.09 9.98</td>
<td>A_Pts12</td>
</tr>
<tr>
<td>A_Pts24</td>
<td>100.88 21.31</td>
<td>86.81 13.68</td>
<td>A_Pts24</td>
</tr>
<tr>
<td>A_Max5</td>
<td>2.96 1.25</td>
<td>2.83 1.15</td>
<td>ELO_DIFF</td>
</tr>
<tr>
<td>A_Min5</td>
<td>0.28 0.51</td>
<td>0.22 0.44</td>
<td>Distance</td>
</tr>
<tr>
<td>H_ELO</td>
<td>1515.65 87.32</td>
<td>1328.58 62.02</td>
<td>H_Champ</td>
</tr>
<tr>
<td>A_ELO</td>
<td>1515.56 88.10</td>
<td>1328.78 61.79</td>
<td>H_T4</td>
</tr>
<tr>
<td>ELO_DIFF</td>
<td>0.0969 125.12</td>
<td>-0.20 89.45</td>
<td>H_B4</td>
</tr>
<tr>
<td>Distance</td>
<td>292.31 147.41</td>
<td>313.72 147.11</td>
<td>H_PS</td>
</tr>
<tr>
<td>H_Champ</td>
<td>0.05 0.07</td>
<td>0.06 0.07</td>
<td>H_OS</td>
</tr>
<tr>
<td>H_EL</td>
<td>0.18 0.17</td>
<td>0.11 0.09</td>
<td>H_TS</td>
</tr>
<tr>
<td>H_PR</td>
<td>0.19 0.18</td>
<td>0.13 0.13</td>
<td>A_T4</td>
</tr>
<tr>
<td>H_Re</td>
<td>0.06 0.12</td>
<td>0.07 0.13</td>
<td>A_B4</td>
</tr>
<tr>
<td>H_PS</td>
<td>0.19 0.13</td>
<td>0.11 0.12</td>
<td>A_PS</td>
</tr>
<tr>
<td>H_OS</td>
<td>0.06 0.08</td>
<td>0.03 0.06</td>
<td>A_OS</td>
</tr>
<tr>
<td>H_TS</td>
<td>0.48 0.29</td>
<td>0.52 0.32</td>
<td>A_TS</td>
</tr>
<tr>
<td>A_Champ</td>
<td>0.05 0.07</td>
<td>0.06 0.07</td>
<td>A_Champ</td>
</tr>
<tr>
<td>A_CL</td>
<td>0.18 0.17</td>
<td>0.11 0.09</td>
<td>A_CL</td>
</tr>
<tr>
<td>A_EL</td>
<td>0.19 0.18</td>
<td>0.13 0.13</td>
<td>A_EL</td>
</tr>
<tr>
<td>A_Pr</td>
<td>0.06 0.12</td>
<td>0.06 0.13</td>
<td>A_Pr</td>
</tr>
<tr>
<td>A_Re</td>
<td>0.20 0.13</td>
<td>0.11 0.12</td>
<td>A_Re</td>
</tr>
<tr>
<td>A_PS</td>
<td>0.07 0.09</td>
<td>0.04 0.06</td>
<td>A_PS</td>
</tr>
<tr>
<td>A_OS</td>
<td>0.48 0.29</td>
<td>0.52 0.31</td>
<td>A_OS</td>
</tr>
</tbody>
</table>
8.3.1. Division one Bundesliga

For first division Bundesliga football, all first round matches were removed from the analysis due to no pre-match information being available to complete the importance calculations. A further six matches were identified as potential outliers during the analysis and were subsequently removed from the final model. Therefore, a total of 2,667 division one Bundesliga matches were included in the final analysis.

Table 8.3 provides the forward stepwise regression results for first division Bundesliga football. The assumption of collinearity was met within the final stepwise model (Tolerance > 0.1, VIF < 10 for all independent variables). The final model also met the assumption of independent errors (Durbin-Watson = 1.963), while a histogram of the standardised residuals indicated that the data contained approximately normally distributed errors. Finally, the assumptions of homogeneity of variance and linearity were confirmed through a scatterplot of the standardised residuals.

Using the forward stepwise selection criteria, a statistically significant regression equation was found (F(5, 2662) = 117.543, \( p > 0.000 \)), with an \( R^2 \) of 0.181. Two importance variables, \( A_{PS} \) and \( H_{EL} \), were both found to be significant within the final stepwise model (\( p = 0.006 \) and \( p = 0.028 \), respectively). Despite the final stepwise model containing significant independent variables, a low \( R^2 \) value indicates that only 18.1% of the variation in the final score difference can be explained by the independent variables. This \( R^2 \) value will serve as the benchmark for comparison with matches that are classified by their level of importance to both the home and away teams.
### Stepwise models

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ELO</strong></td>
<td>0.005***</td>
<td>0.005***</td>
<td>0.005***</td>
<td>0.005***</td>
<td>0.005***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>H_M</strong></td>
<td>0.120***</td>
<td>0.090***</td>
<td>0.103***</td>
<td>0.128***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.017)</td>
<td>(0.019)</td>
<td>(0.022)</td>
<td></td>
</tr>
<tr>
<td><strong>A_P</strong></td>
<td>0.504**</td>
<td>0.620***</td>
<td>0.658***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.232)</td>
<td>(0.239)</td>
<td>(0.239)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>H_E</strong></td>
<td>-0.363**</td>
<td>-0.383**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.174)</td>
<td>(0.174)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>H_G</strong></td>
<td>-0.061**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **R^2** | 0.136       | 0.177      | 0.178      | 0.180      | 0.181      |
|         | 0.135       | 0.176      | 0.177      | 0.178      | 0.179      |
| **No. observations** | 2667       |

Standard error presented in parenthesis

*, **, *** indicates significance at the 90%, 95% and 99% confidence level, respectively

Table 8.3: Forward stepwise coefficients for first division Bundesliga matches

Like the Elo ratings system from Chapter 7, matches were sorted into nine classifications based on their level of importance for both the home and away teams. Since there are seven importance categories within first division Bundesliga, a total of 58 partitioned data sets can be assessed. Note that ‘Relegation’ has only four classifications due to the low importance values across teams. The stepwise model was re-applied to all 58 partitioned data sets, with the change in $R^2$ presented in Figure 8.1.
During calculations, the issue of multicollinearity persisted within certain match classifications, including HL of ‘Relegation’, ML of ‘Positional Sum’, and HM and LM of ‘Outcome Sum’. For the remaining 54 classifications, 33 (61.11%) experienced an increase in the $R^2$. This means that, for a slight majority, the results of the stepwise regression model have been improved by classifying matches based on their level of importance to the competing teams.

Figure 8.1: $R^2$ difference for first division Bundesliga matches classified by importance
A bar chart representing the number of times each independent variable appeared within a partitioned model is presented in Appendix G. The most common variable included within each partitioned model was $ELO_{DIFF}$, which appeared in 52 of the 58 models (89.67%); while the second most common variable was $H_{Max}$, which was included 15 times (25.86%). For the importance variables, $A_{PS}$ was included within a model seven times, while $H_{PS}$ and $A_{Re}$ were both included five times. All scored-based variables appeared in at least one of the partitioned models, while $A_{ELO}$, $H_{EL}$, $H_{Re}$ and $A_{Champ}$ were never included.

In terms of the first variable entered into the model, $ELO_{DIFF}$ was included in 44 of the 58 partition models (75.86%), with the second largest being $H_{AvgGlsFor}$ (4 times). Of the importance variables, only $H_{Champ}$ was entered first into the model multiple times (twice), while $H_{CL}$, $A_{EL}$ and $A_{OS}$ were all entered first into the model once. The large number of times that the Elo ratings difference was included first indicates that the information contained within this variable is a critical factor when modelling the final score difference. Inversely, the low number of times that a match importance variable was entered first indicates that the information provided from these variables is not an essential factor.

### 8.3.2. Division two Bundesliga

Like the first division, all first round matches were removed from the second division model due to no pre-match information being available for the importance calculations. Eleven matches were also identified as potential outliers and were subsequently removed from the final analysis. Therefore, a total of 2,662 division two Bundesliga matches were included within the stepwise regression analysis.

Table 8.4 provides the forward stepwise regression results for second division Bundesliga football. The assumption of collinearity was met within the final stepwise model.
(Tolerance > 0.1, VIF < 10 for all independent variables), but not before \( A_{Pts12}, A_{Pts24}, H_{ELO} \) and \( A_{ELO} \) were removed due to multicollinearity concerns. The final stepwise model also met the assumption of independent errors (Durbin-Watson = 1.932), while a histogram of the standardised residuals indicated that the data contained approximately normally distributed errors. Finally, the assumptions of homogeneity of variance and linearity were confirmed through a scatterplot of the standardised residuals.

A statistically significant regression equation was found (\( F(4, 2658) = 87.056, p > 0.000 \)) with an \( R^2 \) of 0.116. The low \( R^2 \) indicates that only 11.6% of the variation in the final score difference can be explained by the four significant independent variables (\( H_{AvgGlsFor}, p=0.001; ELO\_DIFF, p > 0.000; A_{AvgGlsAg}, p > 0.000; A_{GlsAgLast}, p=0.002 \)). No match importance variables were found to be significant predictors of the final score difference during the forward selection procedure.
Stepwise models

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H_AvgGlsFor</strong></td>
<td>0.309***</td>
<td>0.274***</td>
<td>0.167***</td>
<td>0.164***</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.049)</td>
<td>(0.049)</td>
</tr>
<tr>
<td><strong>ELODIFF</strong></td>
<td>0.004***</td>
<td>0.004***</td>
<td>0.004***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td><strong>A_AvgGlsAg</strong></td>
<td></td>
<td></td>
<td>0.121**</td>
<td>0.204***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.049)</td>
<td>(0.056)</td>
</tr>
<tr>
<td><strong>A_GlsForLast</strong></td>
<td></td>
<td></td>
<td></td>
<td>-0.093***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.031)</td>
</tr>
<tr>
<td>R²</td>
<td>0.064</td>
<td>0.111</td>
<td>0.113</td>
<td>0.116</td>
</tr>
<tr>
<td>Adj. R²</td>
<td>0.063</td>
<td>0.110</td>
<td>0.112</td>
<td>0.115</td>
</tr>
<tr>
<td>No. observations</td>
<td>2662</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard error presented in parenthesis
* *, **, *** indicates significance at the 90%, 95% and 99% confidence level, respectively

Table 8.4: Forward stepwise coefficients for second division Bundesliga matches

The \( R^2 \) of 0.116 will serve as the benchmark for comparison with the results generated by classifying matches based on their level of importance to the competing teams. Figure 8.2 provides a graphical comparison of the \( R^2 \) change within each match classification for the six division two importance categories. The issue of multicollinearity occurred within one classification (ML of ‘Total Sum’), while one classification (LM of ‘League Champion’) could not complete calculations as no variables met the selection criteria.

Of the remaining 47 partitioned models, 28 observed an increase in the \( R^2 \) (59.57%), with the largest increase occurring within HL of the ‘League Champion’ category. Within this model, only \( ELO\_DIFF \) and \( A\_Max5 \) were found to be a significant determinant of the final score difference. The increase in the \( R^2 \) for a majority of the partitioned models indicates that the stepwise regression model can be improved when matches are classified by the level
of importance to the competing teams. It also casts doubt on the application of the match importance as an independent variable when modelling second division Bundesliga results, where no match importance variables were included in the final stepwise model in Table 8.4.

Figure 8.2: $R^2$ difference for second division Bundesliga matches classified by importance

The frequency of each variables inclusion within the second division Bundesliga partitioned models is presented in Appendix G. Like the first division, the most commonly
included variable was $ELO\_DIFF$, which featured in 34 of the 48 models (70.83%); including 19 times (39.58%) where it was the first variable entered into the model. For the match importance variables, the most frequently included was the $H\_Champ$ (five times), while the next highest were $H\_TS$, $A\_Champ$ and $A\_TS$ (four times). The frequent inclusion of $ELO\_DIFF$ combined with the low inclusion rate of the match importance variables indicate that the information contained within the former is more critical to modelling the second division Bundesliga score difference than the information contained within the latter.

Assessing the overall performance of both the first and second division models, it appears that the stepwise regression model is a poor choice for predicting the score difference in Bundesliga football. This is highlighted by the low $R^2$ value for the final stepwise models within both divisions, with only 18.1% and 11.6% of the variation in score difference within the first and second division models, respectively, being explained by the selected independent variables. While this statistical model had not been explored in past research on Bundesliga football, the results contained within this section suggests that future research should focus on alternative models, especially when assessing the potential significance of match importance on the final score difference.

8.3.3. NBA

For the NBA stepwise regression model, a total of 714 of the original 11,069 matches from the primary data were removed due to incompleteness. This included all matches that featured a team playing their first match, where no pre-match information is available to complete their importance calculations. An additional 22 matches were identified as potential outliers and were subsequently removed from the analysis. Therefore, a total of 10,326 NBA matches are included within the final stepwise model.
Table 8.5 provides the forward stepwise regression results for NBA basketball. Six variables (H_AvgPtsFor, H_AvgPtsAg, H_Max5, A_AvgPtsFor, and A_AvgPtsAg) were removed due to multicollinearity concerns. The assumption of collinearity was met within the final stepwise model (Tolerance > 0.1, VIF < 10 for all independent variables), while the assumption of independent errors was also met (Durbin-Watson = 2.024). A histogram of the standardised residuals indicated that the data contained approximately normally distributed errors; while the assumptions of homogeneity of variance and linearity were confirmed through a scatterplot of the standardised residuals.

<table>
<thead>
<tr>
<th>Stepwise models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>ELO_DIFF</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A_ELO</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>H_PS</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A_OS</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>R²</td>
</tr>
<tr>
<td>Adj. R²</td>
</tr>
<tr>
<td>No. observations</td>
</tr>
</tbody>
</table>

Standard error presented in parenthesis
*.*, ***, *** indicates significance at the 90%, 95% and 99% confidence level, respectively

Table 8.5: Forward stepwise coefficients for NBA matches
A statistically significant regression equation was found within the final model \( (F(4, 10322) = 710.798, p > 0.000) \) with an \( R^2 \) of 0.216. While the \( R^2 \) is a slight improvement over the Bundesliga models, the low value indicates that only 21.6\% of the variation in the score difference can be explained by the four significant independent variables (\( ELO\_DIFF, p>0.000 \); \( A\_ELO, p>0.000 \); \( H\_PS, p>0.000 \); \( A\_OS, p>0.000 \)). An interesting result is that two of the weighted-sum importance variables, \( H\_PS \) and \( A\_OS \) were included in the final model, while none of the position-specific importance variables were found to be a significant predictor of the final score difference.

The \( R^2 \) of 0.216 will serve as the benchmark for comparison with the results obtained through classifying matches by their level of importance for competing teams. Figure 8.3 provides a graphical comparison of the \( R^2 \) change observed within each importance category. During calculations, the issue of multicollinearity persisted within the partitioned models, with 13 of the 54 models being affected. For the remaining 41 models, an increase in the \( R^2 \) was observed within only 13 match classifications (36.59\%).
The frequency of inclusion for each variable within the NBA partitioned models is presented in Appendix G. Like the Bundesliga models, ELO_DIFF was the most commonly included variable, appearing in 53 of the partitioned models (98.15%); while it was entered first into the model 44 times (81.48%). The most commonly included match importance variable was H_TS, which was featured in nine of the models (16.67%). The high frequency of the Elo ratings difference compared to the low inclusion rate of the match importance

Figure 8.3: $R^2$ difference for NBA matches classified by importance
variables indicates that the information contained within the former is more critical to modelling the score difference in the NBA than the information contained within the latter; which was also observed within the Bundesliga models. This result is discussed further in the next section of this chapter.

Like the Bundesliga models, the poor results for the NBA model suggest that applying the stepwise regression model to predict the score difference is not suitable for NBA basketball. This is highlighted by the poor $R^2$ result of the final stepwise model, where only 21.6% of the variation in the final score difference was explained by the selected independent variables. However, the use of a stepwise regression model to predict the final score difference in NBA basketball, and the potential significance of match importance, had not been explored in past research. The poor results contained within this chapter suggests that future research should focus on alternative predictive models, especially when investigating the potential significance of match importance on the final score difference.
8.4. Discussion

The primary objective of this chapter was to determine if the match importance generated using the Overtake approach was a significant predictor of the final score difference in both Bundesliga football and NBA basketball. Using a forward stepwise regression model, it was found that certain match importance variables were significant in both first division Bundesliga football and NBA basketball. However, an overall improvement in the performance of the models was observed when matches were classified by their level of importance to both the home and away team. The improvement in the coefficient of determination ($R^2$) suggests that classifying matches by their level of importance can have an effect on the overall performance of the forward stepwise regression model.

In Section 8.3, significant match importance variables were found in the final stepwise model for both first division Bundesliga football ($A_{PS}$ and $H_{EL}$) and NBA basketball ($H_{PS}$ and $A_{OS}$); while no importance variables were significant in the final model for second division Bundesliga football. For the first division Bundesliga model, $A_{PS}$ was found to increase the score difference by 0.658, while $H_{EL}$ was found to decrease the score difference by 0.383. For the NBA model, $H_{PS}$ increased the score difference by 15.073, while $A_{OS}$ decreased the score difference by 17.437. Despite the significant results within both sports, the final models recorded a low coefficient of determination ($R^2 = 0.181$ for first division Bundesliga; $R^2 = 0.216$ for NBA). This means that less than half the variation in the score difference can be explained by the significant importance variables.

Focusing on the remaining independent variables within the final stepwise models, a common inclusion was the Elo ratings difference ($ELO\_DIFF$) between teams. This result indicates that the difference in strength between the teams is a key determinant of the final
score difference within both sports – an intuitive result. For first division Bundesliga, $H_{Max5}$ and $H_{GlsForLast}$ were also found to be significant; meaning that the recent offensive ability of the home team is critical when modelling the final score difference. For second division Bundesliga, $H_{AvgGlsFor}$, $A_{AvgGlsAg}$ and $A_{GlsForLast}$ were all significant. Therefore, the historical offensive ability of the home team, combined with the historical defensive and recent offensive ability of the away team, is a significant determinant of the score difference within the second league. For NBA basketball, only the Elo ratings difference and the Elo rating of the away team ($A_{ELO}$) were found to be significant. Consequently, the recent and historical scoring performance of the teams was not a determinant of the final score difference.

As mentioned, the $R^2$ within each model indicated that less than half the variation in the final score difference could be explained by the significant independent variables. To see if the regression results could be improved, matches were classified according to their level of importance to both the home and away team. Both first and second division Bundesliga models observed an improve $R^2$ in a majority of the partitioned models, which indicates that classifying matches by their level of importance to the competing teams can have an effect on the model’s performance. However, the application within the NBA observed an $R^2$ increase in only 36.59% of the partitioned models. The differing results between the sports suggest that classifying matches by their level of importance has a greater effect in Bundesliga football compared to NBA basketball.

Similar to the final stepwise models, the difference in Elo ratings was the most commonly included variable within the partitioned models in both sports. The Elo ratings difference was also the first variable entered in a majority of the partitioned models during the forward selection procedure. This result indicates that the difference in team strength is more essential when modelling the final score difference than the match importance or score-
based independent variables. The significance of the Elo ratings difference within the stepwise regression models supports similar findings within professional football (Hvattum & Arntzen, 2010; Reade & Akie, 2013).

Focusing on the binned results, the issue of multicollinearity persisted in a greater number of partitioned models within the NBA compared to the Bundesliga. The issue within the NBA may be due to the nature of scoring, where scores are much larger than those in Bundesliga football due to higher scoring rates. As a result of this, match scores more readily and consistently converge to population means, which is detrimental to the usefulness of some of the predictors described previously. When several scores are combined to generate a predictor, such as in the moving average variables, the convergence of scores causes little to no inter-team variation in predictor values, particularly when teams are partitioned by importance. The direct consequence of this is multicollinearity in predictors, rendering them ineffective in many situations. Therefore, partitioning matches into small samples is much less effective in the NBA as compared to Bundesliga football.

While multicollinearity was an issue within a number of the partitioned models, the overall performance of the final stepwise regression model within both sports was poor. For Bundesliga football, less than 20% of the variation in the score difference was explained by the selected independent variables within both the first and second division models. For NBA basketball, only 21.6% of the variation in the final score difference could be explained by the selected independent variables. The low $R^2$ values across both sports suggest that applying a stepwise regression model to predict the final score difference within both sports is a poor choice; especially when investigating the potential significance of match importance on the final score difference. Therefore, future research should avoid this type of model and focus on alternative statistical approaches.
Since the additional independent variables included within this chapter were selected based on an overview of the existing literature, they may not be best suited for use within a stepwise regression model. This was most noticeable within the NBA models, where the issue of multicollinearity persisted within the partitioned models. Future application of the stepwise regression model should assess other possible explanatory variables, such as the average number of shots per match in football; or the total number of assists in basketball, which has been found to be a significant determinant of team success (Ibáñez et al., 2008). While it is difficult to determine what improvement could be achieved by applying alternative variables, it does provide an interesting future research idea that could further enhance the results obtained within this chapter.

While a stepwise regression model was applied within this chapter to explore the match importance generated from the Overtake approach, the poor results for the final stepwise models within both sports suggest that future research should focus on the application of match importance within other statistical models. This includes a bivariate Poisson distribution for the number of goals scored by football teams (Karlis & Ntzoufras, 2003), a generalised linear model (Rue & Salvesen, 2000), or an ordered probit regression for modelling win/draw/loss match results (Goddard, 2005). Since the existing literature has often applied a binary measure of match importance (see Chapter 1), it would be interesting to compare the results when a scale importance value, such as those calculated using the Overtake approach, is applied.

In terms of applying the importance of a match within sports modelling, classifying matches by their level of importance appears to have a greater effect on the model’s performance than including it as an explanatory variable. While some match importance variables were found to be significant predictors of the score difference in both sports, a more enhanced result in terms of the coefficient of determination was obtained by binning matches
by their importance. A similar result was also observed within the Elo ratings system in Chapter 7, where a change in the correct prediction percentage occurred when the level of importance for the pre-match favourite and underdog differed. The promising results from the two modelling techniques suggest that binning matches by their level of importance can, under certain circumstances, have a more pronounced effect on a predictive model’s performance than simply applying match importance as an explanatory variable.

Despite the promising results obtained in this chapter and the previous chapter, the work on assessing the match importance generated from the *Overtake* approach within sports modelling is not complete. While the completed work has focused on modelling the outcome of a sporting contest, there are other aspects of sports research that the *Overtake* match importance can be applied to. This includes applying the match importance to explore the efficiency of betting markets (Forrest & Simmons, 2008); modelling attendance figures at sporting events (Borland & MacDonald, 2003); and assessing the competitive balance of sports competitions (Evans, 2014). However, the work completed within this dissertation does provide an increase in the understanding of how the match importance can affect the predictive model in both a two-result and a three-result sport.
8.5. Summary

In this chapter, the importance of a match generated from the Overtake approach was applied within a forward stepwise regression model. The objective was to determine if the match importance was a significant predictor of the final score difference in both Bundesliga football and NBA basketball. Results from both sports indicated that particular match importance variables, created through a summation of the positional importance, were a significant predictor of the final score difference. However, an overall improvement in the performance of the models was observed when matches were classified by their level of importance to the competing teams. This result indicates that classifying matches based on their level of importance to the competing teams has an effect on the performance of a forward stepwise regression model.
Chapter 9

Conclusions and future research

The work completed within this dissertation was completed to further the knowledge of quantifying and applying the importance of a match within both a two-result (NBA basketball) and a three-result (Bundesliga football) sport. By first completing an evaluation of an existing probabilistic measure of the importance of a match, a new measure called the Overtake approach was introduced and found to be adaptable to sports with differing match outcomes. The results of the new approach were verified through a comparison with a computer simulation procedure, before being applied within both an Elo ratings system and a stepwise regression model. This chapter summarises the key findings from each of the previous five chapters, while also discussing the potential for future research.
9.1. *ParWins* importance

In Chapter 4, an existing probabilistic measure of match importance, called the *ParWins* method, was extended and applied to both a two-result (NBA basketball) and a three-result (Bundesliga football) sport. The measure defined the importance of a match as the difference between success probabilities conditional on a team achieving a win or a loss in their next match. By applying a cumulative binomial distribution function, the method calculated the conditional probabilities with respect to a team achieving a projected wins requirement to finish in different positions within the standings. It was believed that the measure could be adaptable to both sports.

However, application of the method within both sports produced flawed results, such as teams recording an importance value equal to zero despite being in contention for a specific end-of-season outcome. Furthermore, application of the measure within football failed to account for the probability of a draw outcome occurring, with only the win and loss conditional probabilities being included. It was concluded that the measure was not adaptable to both two-result and three-result sports due to the poor season average distribution results and the illogical team-specific importance results.

Despite the *ParWins* method failing to be adaptable to both NBA basketball and Bundesliga football, its ability to quantify the importance of a match in other multi-outcome sports is unclear. Furthermore, the measure assessed in this dissertation applied constant match outcome probabilities within the cumulative binomial distribution to generate the conditional probabilities. However, constant match outcome probabilities ignore crucial factors such as home advantage, which has been found to be prominent within both sports in past research (Schwartz & Barsky, 1977). Future research on the *ParWins* method should focus on applying team-specific match outcome probabilities within NBA basketball and
Bundesliga football, which may lead to an improvement in the results for both sports. Moreover, future studies should focus on applying the measure to other multi-result sports, which would help provide a definitive conclusion regarding its suitability for quantifying the importance of a match.

9.2. **Overtake importance**

In Chapter 5, a new probabilistic measure of match importance, called the *Overtake* approach, was introduced and applied to both Bundesliga football and NBA basketball. The model introduced an adjusted definition of match importance within a three-result sport, where the draw outcome was included alongside the win as a non-negative result. This adjusted definition allowed for the draw outcome to be accounted for during match importance calculations, which had not yet been achieved in past literature. By applying a Markov Chain model to both Bundesliga football and NBA basketball, the model approximated the conditional probabilities by identifying success and failure scenarios such that a lower-ranked team could overtake, but not equal, a higher-ranked team in position $s$. The application of a Markov Chain model in football allowed for the importance of a draw outcome to be quantified, which had not been previously achieved in the literature.

The results for both sports provided an improvement over those observed in Chapter 4. The season average positional importance distributions within both sports peaked at the start of the season before declining towards the conclusion; which corresponded with a decrease in the number of teams in contention for each position as the season progressed. When only the teams in contention for position $s$ were assessed, it was observed that the average importance peaked at the conclusion of the season. A re-evaluation of the case studies from Chapter 4 also revealed an improved result, with teams no longer recording an importance value equal to zero while being in contention for position $s$. 

206
Results also demonstrated that the match importance in Bundesliga football can be broken into win and draw components, where a summation of the two equate to the overall result importance. From the Bundesliga case studies, it was observed that the draw outcome can be important to a team given certain season circumstances, including the close proximity of teams with only one round of matches remaining. The successful results within both sports led to the conclusion that the \textit{Overtake} approach was adaptable to both a two-result and a three-result, with the draw outcome being accounted for in the latter for the first time in the literature.

In terms of future work, like the \textit{ParWins} method, team-specific match outcome probabilities are required within the \textit{Overtake} approach, which may further enhance the results. Furthermore, adjusting the draw outcome between a non-negative and a negative result, dependent on a team’s current position in the standings, may further improve the observed results. Like the \textit{ParWins} method, the \textit{Overtake} approach should be applied to other professional sports, where a comparison between sports leagues would provide a definitive conclusion regarding the use of the \textit{Overtake} approach for quantifying the importance of a match.

\textbf{9.3. Simulation}

In Chapter 6, a comparison between the \textit{Overtake} approach and a Monte Carlo simulation model was completed to validate the application of the former for quantifying the importance of a match. Using the constant match outcome probabilities from Chapter 5, the results from the Monte Carlo simulation were similar to those observed from the \textit{Overtake} approach. This included the average positional importance for both Bundesliga football and NBA basketball reaching a maximum at the start of the season before declining towards the conclusion. Bayes’ Rule was also applied within both models to vary the match outcome
probabilities with respect to the current form of teams; which led to marginal improvement in the results for both models. It was concluded that the reduction in computational complexity of the Overtake approach produced similar results to the Monte Carlo simulation model. Therefore, a reasonable approximation of the importance of a match can be obtained by applying the Overtake approach instead of a computer simulation procedure.

In terms of future research, it would be interesting to compare the two modelling techniques within other professional sports, such as English Premier League football or a European basketball league. A comparison within other sports would further validate the results obtained within Chapter 5 as they can provide further information about when the most important matches, on average, occur throughout a season. Furthermore, future research can compare the Overtake approach to other existing measures of match importance, such as the retrospective approach by Jennett (1984). This could lead to an interesting discussion regarding how the importance of a match varies throughout a season for different finishing positions in the standings. However, for the sake of validating the Overtake approach in Chapter 6, the Monte Carlo simulation procedure was viewed as the best method for comparison as it accounts for the results of all teams, and thus provides a comprehensive perspective on how the importance of a match varies throughout a season.

9.4. Elo ratings

In Chapter 7, the importance of a match was applied to three variations of the Elo ratings system in both Bundesliga football and NBA basketball. It was hypothesised that a change in the correct prediction percentage and the Brier score of the Elo ratings system could be observed by classifying matches based on their level of importance to both the pre-match favourite and underdog. Using different summations of the positional importance from Chapter 5, it was observed that an effect on the correct prediction percentage and the Brier
score occurs when matches are classified by their level of importance to the competing teams. The most noticeable results were observed when solely focusing on the importance of finishing in first position; where the correct prediction percentage increased when the match was of higher importance to the favourite, and decreased when the match was of higher importance to the underdog. It was concluded that, under certain circumstances, the match importance can have an effect on the correct prediction percentage and the Brier score of the Elo ratings system within both sports.

For future research, enhancing the Elo ratings systems’ ability to predict draw outcomes in football would lead to an improvement in the overall results in addition to providing a model that aligns more closely with the nature of football. Furthermore, other measures of model performance, such as the absolute difference between predicted and actual margins (Bailey & Clarke, 2004), may lead to an alternative result regarding the application of match importance within the Elo ratings system. Future research could also explore adjusting the features within the model, such as applying a larger quantity of test data to configure the model parameters; or introducing a seasonal decay to smooth the results (Ryall & Bedford, 2010). The application of match importance within the Elo ratings system should also be assessed in other professional sports, which would help provide a well-rounded perspective regarding the observed results in this dissertation.

9.5. Regression analysis

In Chapter 8, the match importance generated from the Overtake approach was applied within a stepwise regression model for both Bundesliga football and NBA basketball. The objective was to determine if the importance of a match was a significant determinant of the final score difference within each sport. To complete the models, additional independent variables were included based on their application within past literature. Within both first
division Bundesliga and NBA basketball, it was observed that particular match importance variables were significant, such as the home ‘Europa League’ for Bundesliga football, and the away ‘Outcome Sum’ for NBA basketball. However, the overall performance of each model was poor, with less than half the variation of the score difference being explained by the independent variables.

An overall improvement in the stepwise regression models was observed when the matches were classified according to their level of importance to the competing teams. The improvement was observed through an increase in the coefficient of determination across a majority of the partitioned models within both first and second division Bundesliga football. However, an increase in the majority of the partitioned models was not observed within NBA basketball, where the issue of multicollinearity persisted within the partitioned models. This led to the conclusion that classifying matches based on their level of importance to the competing teams had a more pronounced effect within Bundesliga football than NBA basketball.

To further explore the results in future research, alternative independent variables should be applied within the stepwise regression models. This includes game-specific variables, such as the total number of assists in basketball; or the average number of shots on target in football. The application of match importance from the Overtake approach could also be explored within other statistical models, such as an ordered probit regression model or a bivariate Poisson distribution model; both which have been assessed frequently in past literature. Furthermore, the match importance generated from the Overtake approach can be applied in future research to model other aspects of a sporting contest besides the match outcomes, such as attendance figures, competitive balance in sporting competitions and the efficiency of betting markets.
9.6. Summary

The work completed within this dissertation has helped further the knowledge of quantifying the importance of a match in professional sports using probabilistic models. By building off the drawbacks of an existing method, the new *Overtake* approach provides a measure of match importance that is adaptable to both a two-result and a three-result sport; where the importance of the draw outcome is quantified for the first time in the latter. The *Overtake* approach also provides similar results to a Monte Carlo simulation procedure while reducing the computational complexity and total runtime. The observed results within both the Elo ratings system and forward stepwise regression models also indicate that there are circumstances where the match outcome predictions can be affected by the level of match importance to the competing teams. While there is still research to be completed within the literature, the findings contained within this dissertation help increase the knowledge of quantifying and applying match importance within sports modelling.
Bibliography


212


Appendix

Appendix A: *Overtake* approach – Markov Chain VBA program

for Bundesliga football

'Function to recursively calculate the probability Team A defeats Team B

'Function argument "RelativePosition" is either "Higher" or "Lower", we assume goal
difference is maintained, so a "Higher" team needs only have end score =0

Public Function fWinProb(TeamAScore, TeamBScore, RelativePosition, GamesRemaining, PWin, PDraw, PLose)

Dim RowCounter, ColumnCounter, ColumnCounter2, Deficit, RowRef As Integer

Dim SumProb

Deficit = TeamAScore - TeamBScore

ReDim TransProb(1 To 1, 1 To 7)

TransProb(1, 1) = PLose * PWin

TransProb(1, 2) = PDraw * PWin

TransProb(1, 3) = PLose * PDraw

TransProb(1, 4) = PWin * PWin + PDraw * PDraw + PLose * PLose

TransProb(1, 5) = PDraw * PLose

TransProb(1, 6) = PWin * PDraw

TransProb(1, 7) = PWin * PLose

ReDim ProbDist(0 To GamesRemaining, -105 To 105)

For ColumnCounter = -105 To 105

ProbDist(0, ColumnCounter) = 0

Next ColumnCounter

ProbDist(0, Deficit) = 1

For RowCounter = 1 To GamesRemaining

    For ColumnCounter = -102 To 102

        For ColumnCounter2 = -105 To 105

            SumProb = SumProb + ProbDist(RowCounter - 1, ColumnCounter2) * TransProb(ColumnCounter2, Transmit) * ProbDist(RowCounter - 1, Deficit)

        Next ColumnCounter2

    Next ColumnCounter

Next RowCounter

ProbDist(RowCounter, ColumnCounter) = SumProb

End Function
RowRef = RowCounter - 1

SumProb = 0

For ColumnCounter2 = 1 To 7
    SumProb = SumProb + TransProb(1, ColumnCounter2) * ProbDist(RowRef, ColumnCounter - 4 + ColumnCounter2)
    Next ColumnCounter2

ProbDist(RowCounter, ColumnCounter) = SumProb

Next ColumnCounter

Next RowCounter

'Calculate probability Team A finishes higher than Team B

SumProb = 0

For ColumnCounter = 1 To 102
    SumProb = SumProb + ProbDist(GamesRemaining, ColumnCounter)
    Next ColumnCounter

If RelativePosition = "Higher" Then SumProb = SumProb + ProbDist(GamesRemaining, 0)

fWinProb = SumProb

End Function

'Function to determine Win or Draw Importance for Team A

'ImpType is either "Win" or "Draw"

Public Function fImp(TeamAScore, TeamBScore, RelativePosition, GamesRemaining, PWin, PDraw, PLose, ImpType)
    Dim pw1, pw2, pw3, pd1, pd2, pd3, pl1, pl2, pl3
    Dim pw, pd, pl

    pw1 = PWin * fWinProb(TeamAScore + 3, TeamBScore + 3, RelativePosition, GamesRemaining - 1, PWin, PDraw, PLose)
    pw2 = PDraw * fWinProb(TeamAScore + 3, TeamBScore + 1, RelativePosition, GamesRemaining - 1, PWin, PDraw, PLose)
pw3 = P Lose * fWinProb(TeamAScore + 3, TeamBScore, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pw = PWin * (pw1 + pw2 + pw3)

pd1 = PWin * fWinProb(TeamAScore + 1, TeamBScore + 3, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pd2 = PDraw * fWinProb(TeamAScore + 1, TeamBScore + 1, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pd3 = P Lose * fWinProb(TeamAScore + 1, TeamBScore, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pd = PDraw * (pd1 + pd2 + pd3)

pl1 = PWin * fWinProb(TeamAScore, TeamBScore + 3, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pl2 = PDraw * fWinProb(TeamAScore, TeamBScore + 1, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pl3 = P Lose * fWinProb(TeamAScore, TeamBScore, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pl = P Lose * (pl1 + pl2 + pl3)

If pw + pd = 0 Then fImp = 0

If ImpType = "Win" And pw + pd > 0 And pw * P Lose > pl * PWin Then
fImp = pw - pl * (pw / (pw + pd))

ElseIf ImpType = "Draw" And pw + pd > 0 And pd * P Lose > pl * PDraw Then
fImp = pd - pl * (pd / (pw + pd))

Else
fImp = 0

End If

Public Function fPW(TeamAScore, TeamBScore, RelativePosition, GamesRemaining, PWin, PDraw, P Lose, ImpType)

Dim pw1, pw2, pw3

pw1 = PWin * fWinProb(TeamAScore + 3, TeamBScore + 3, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pw2 = PDraw * fWinProb(TeamAScore + 3, TeamBScore + 1, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pw3 = P Lose * fWinProb(TeamAScore + 3, TeamBScore, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

fPW = PWin * (pw1 + pw2 + pw3)
End Function

Public Function fPD(TeamAScore, TeamBScore, RelativePosition, GamesRemaining, PWin, PDraw, P Lose, ImpType)
Dim pd1, pd2, pd3
pd1 = PWin * fWinProb(TeamAScore + 1, TeamBScore + 3, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pd2 = PDraw * fWinProb(TeamAScore + 1, TeamBScore + 1, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pd3 = P Lose * fWinProb(TeamAScore + 1, TeamBScore, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

fPD = PDraw * (pd1 + pd2 + pd3)
End Function

Public Function fPL(TeamAScore, TeamBScore, RelativePosition, GamesRemaining, PWin, PDraw, P Lose, ImpType)
Dim pl1, pl2, pl3
pl1 = PWin * fWinProb(TeamAScore, TeamBScore + 3, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pl2 = PDraw * fWinProb(TeamAScore, TeamBScore + 1, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

pl3 = P Lose * fWinProb(TeamAScore, TeamBScore, RelativePosition, GamesRemaining - 1, PWin, PDraw, P Lose)

fPL = P Lose * (pl1 + pl2 + pl3)
End Function
' Copy and Paste Importance Matrix Code

Public Sub pResultImportance()
Application.ScreenUpdating = False

Dim ImpCounter, RankCounter, Teams, Rounds, Imps, PWin, PDraw, P Lose, ImpType, ImpFirstRow,
GamesRemaining

' Dim SimTime

' Dim StartTime

' StartTime = Now()

Teams = 18

Rounds = 34

Imps = 17

If Cells(1, 2) = 1 Then

    PWin = 0.375
    PDraw = 0.25
    P Lose = 0.375

Else

    PWin = 0.36
    PDraw = 0.28
    P Lose = 0.36

End If

' For All Divisions: PWin/ PDraw/ P Lose = 0.37/0.26/0.37

' For Division One: PWin/ PDraw/ P Lose = 0.375/0.25/0.375

' For Division Two: PWin/ PDraw/ P Lose = 0.36/0.28/0.36

Dim PasteStartingRow, PasteStartingColumn

Dim ScoreStartingRow, ScoreStartingColumn

ScoreStartingRow = 2

ScoreStartingColumn = 3

PasteStartingColumn = 38

Dim CurrentScoreStartingRow, CurrentScoreStartingColumn

CurrentScoreStartingRow = 2
CurrentScoreStartingColumn = 38

' Current Position in Score Matrix
ReDim CurrentScoreMatrix(1 To Teams, 1 To Rounds)

Dim CurrentScoreRange As Range

Set CurrentScoreRange = Range(Cells(CurrentScoreStartingRow, CurrentScoreStartingColumn),
  Cells(CurrentScoreStartingRow + Teams - 1, CurrentScoreStartingColumn + Rounds - 1))

CurrentScoreMatrix = CurrentScoreRange

For ImpCounter = 1 To 2
  If ImpCounter = 1 Then
    ImpType = "Win"
    ImpFirstRow = 162
  Else
    ImpType = "Draw"
    ImpFirstRow = 502
  End If

  For RankCounter = 1 To Imps ' Number of Importance Positions
    PasteStartingRow = ImpFirstRow + 20 * (RankCounter - 1)
    ' Retrieve Scores
    ReDim ScoreMatrix(1 To Teams, 1 To Rounds)
    Dim ScoreRange As Range
    Set ScoreRange = Range(Cells(ScoreStartingRow, ScoreStartingColumn),
      Cells(ScoreStartingRow + Teams - 1, ScoreStartingColumn + Rounds - 1))
    ScoreMatrix = ScoreRange
    ' Importance Calculations
    Dim RowCounter, ColumnCounter
    Dim RelativePosition, PositionScore, TeamScore, Score1, Score2
    ReDim ImportanceMatrix(1 To Teams, 1 To Rounds)
    For ColumnCounter = 1 To Rounds - 1
GamesRemaining = Rounds - ColumnCounter

PositionScore = CurrentScoreMatrix(RankCounter, ColumnCounter) ' Current IMP leader score

For RowCounter = 1 To Teams

   TeamScore = ScoreMatrix(RowCounter, ColumnCounter) ' Current team of interest

   Score1 = TeamScore

   If TeamScore > PositionScore Then

      Score2 = PositionScore

      RelativePosition = "Higher"

   ElseIf TeamScore = PositionScore Then

      Score2 = CurrentScoreMatrix(RankCounter + 1, ColumnCounter)

      RelativePosition = "Higher"

   Else

      Score2 = PositionScore

      RelativePosition = "Lower"

   End If

   ImportanceMatrix(RowCounter, ColumnCounter) = fImp(Score1, Score2, RelativePosition, GamesRemaining, PWin, PDraw, PLose, ImpType)

Next RowCounter

Next ColumnCounter

' Paste Importance Values into Excel Sheet

Dim PasteRange As Range

Set PasteRange = Range(Cells(PasteStartingRow, PasteStartingColumn), Cells(PasteStartingRow + Teams - 1, PasteStartingColumn + Rounds - 1))

PasteRange = ImportanceMatrix

Next RankCounter

Next ImpCounter

'SimTime = Format(86440 * (Now() - StartTime), "0.0")
'MsgBox ("Complete!"

End Sub
Appendix B: *Overtake* approach – Markov Chain VBA program for NBA basketball

Public Function fWinProb(TeamAScore, TeamBScore, RelativePosition, GamesRemaining, PWin, P Lose)
Dim RowCounter, ColumnCounter, ColumnCounter2, Deficit, RowRef As Integer

Dim SumProb
Deficit = TeamAScore - TeamBScore
ReDim TransProb(1 To 1, 1 To 3)
TransProb(1, 1) = PLose * PWin
TransProb(1, 2) = 2 * PWin * PLose
TransProb(1, 3) = PLose * PWin
ReDim ProbDist(0 To GamesRemaining, -90 To 90)
For ColumnCounter = -90 To 90
ProbDist(0, ColumnCounter) = 0
Next ColumnCounter
ProbDist(0, Deficit) = 1
For RowCounter = 1 To GamesRemaining
For ColumnCounter = -87 To 87
RowRef = RowCounter - 1
SumProb = 0
For ColumnCounter2 = 1 To 3
SumProb = SumProb + TransProb(1, ColumnCounter2) * ProbDist(RowRef, ColumnCounter - 2 + ColumnCounter2)
Next ColumnCounter2
ProbDist(RowCounter, ColumnCounter) = SumProb
Next ColumnCounter
Next RowCounter
'Calculate probability Team A finishes higher than Team B

SumProb = 0

For ColumnCounter = 1 To 90
    SumProb = SumProb + ProbDist(GamesRemaining, ColumnCounter)
Next ColumnCounter

If RelativePosition = "Higher" Then SumProb = SumProb + ProbDist(GamesRemaining, 0)

fWinProb = SumProb

End Function

Public Function fImp(TeamAScore, TeamBScore, RelativePosition, GamesRemaining, PWin, P Lose)

If GamesRemaining = 0 Then
    fImp = 0
Else
    Dim pw1, pw2, pw3, pl1, pl2
    Dim pw, pl
    pw1 = PWin * fWinProb(TeamAScore + 1, TeamBScore + 1, RelativePosition, GamesRemaining - 1, PWin, P Lose)
    pw2 = P Lose * fWinProb(TeamAScore + 1, TeamBScore + 0, RelativePosition, GamesRemaining - 1, PWin, P Lose)
    pw = PWin * (pw1 + pw2)
    pl1 = PWin * fWinProb(TeamAScore, TeamBScore + 1, RelativePosition, GamesRemaining - 1, PWin, P Lose)
    pl2 = P Lose * fWinProb(TeamAScore, TeamBScore + 0, RelativePosition, GamesRemaining - 1, PWin, P Lose)
    pl = P Lose * (pl1 + pl2)
    fImp = pw - pl
End If

End Function

' Copy and paste values into Importance Matrix

Public Sub pResultImportance()
Application.ScreenUpdating = False

' Preliminaries
Dim RankCounter, Teams, Days, Imps, PWin, Plose, ImpFirstRow

Teams = 15
Days = Cells(1, 3)
Imps = 8
PWin = 0.5
Plose = 0.5

Dim PasteStartingRow, PasteStartingColumn
Dim ScoreStartingRow, ScoreStartingColumn
Dim CurrentScoreStartingRow, CurrentScoreStartingColumn
Dim GameStartingRow, GameStartingColumn
Dim ConferenceCounter

For ConferenceCounter = 1 To 2 ' East = 1, West = 2
    If ConferenceCounter = 1 Then
        PasteStartingColumn = 169
        ScoreStartingRow = 34
        ScoreStartingColumn = 4
        CurrentScoreStartingRow = 194
        CurrentScoreStartingColumn = 4
        GamesStartingRow = 130
        GamesStartingColumn = 4
    Else
        PasteStartingColumn = 169
        ScoreStartingRow = 49
        ScoreStartingColumn = 4
    End If
CurrentScoreStartingRow = 209
CurrentScoreStartingColumn = 4
GamesStartingRow = 145
GamesStartingColumn = 4
End If
' Current Position in Score Matrix
ReDim CurrentScoreMatrix(1 To Teams, 1 To Days)
Dim CurrentScoreRange As Range
Set CurrentScoreRange = Range(Cells(CurrentScoreStartingRow, CurrentScoreStartingColumn),
Cells(CurrentScoreStartingRow + Teams - 1, CurrentScoreStartingColumn + Days - 1))
CurrentScoreMatrix = CurrentScoreRange
If ConferenceCounter = 1 Then
  ImpFirstRow = 226
Else
  ImpFirstRow = 241
End If
For RankCounter = 1 To Imps
PasteStartingRow = ImpFirstRow + 32 * (RankCounter - 1)
' Retrieve Scores
ReDim ScoreMatrix(1 To Teams, 1 To Days)
Dim ScoreRange As Range
Set ScoreRange = Range(Cells(ScoreStartingRow, ScoreStartingColumn), Cells(ScoreStartingRow +
Teams - 1, ScoreStartingColumn + Days - 1))
ScoreMatrix = ScoreRange
' Retrieve Games
ReDim GamesMatrix(1 To Teams, 1 To Days)
Dim GamesRange As Range
Set GamesRange = Range(Cells(GamesStartingRow, GamesStartingColumn), Cells(GamesStartingRow + Teams - 1, GamesStartingColumn + Days - 1))

GamesMatrix = GamesRange

' Result Importance Calculations

Dim RowCounter, ColumnCounter

Dim RelativePosition, PositionScore, Score1, Score2, TeamGames, GamesRemaining, Games1

ReDim ImportanceMatrix(1 To Teams, 1 To Days)

For ColumnCounter = 1 To Days - 1

    PositionScore = CurrentScoreMatrix(RankCounter, ColumnCounter)

    For RowCounter = 1 To Teams

        TeamScore = ScoreMatrix(RowCounter, ColumnCounter)
        TeamGames = GamesMatrix(RowCounter, ColumnCounter)
        Score1 = TeamScore
        Games1 = TeamGames
        GamesRemaining = 82 - Games1

        If TeamScore > PositionScore Then
            Score2 = PositionScore
            RelativePosition = "Higher"
        ElseIf TeamScore = PositionScore Then
            Score2 = CurrentScoreMatrix(RankCounter + 1, ColumnCounter)
            RelativePosition = "Higher"
        Else
            Score2 = PositionScore
            RelativePosition = "Lower"
        End If

        ImportanceMatrix(RowCounter, ColumnCounter) = fImp(Score1, Score2, RelativePosition, GamesRemaining, PWin, PLose)

    Next RowCounter

Next ColumnCounter
Next ColumnCounter

' Paste values into Importance Matrix

Dim PasteRange As Range

Set PasteRange = Range(Cells(PasteStartingRow, PasteStartingColumn), Cells(PasteStartingRow + Teams - 1, PasteStartingColumn + Days - 1))

PasteRange = ImportanceMatrix

Next RankCounter

Next ConferenceCounter

End Sub
Appendix C: Monte Carlo simulation VBA program for Bundesliga football

Dim TPositions, TRounds, TOutcomes, TFinalOutcomes
TPositions = 18
TRounds = 33
TOutcomes = 3
TFinalOutcomes = 2
Dim RC1, RC2, RC3, CC1, CC2, CC3
ReDim CurrentTeamRankMatrix(1 To 18, 1 To 34)
ReDim CurrentTeamMatchOutcomeMatrix(1 To 18, 1 To 34)
ReDim FinalTeamPointsMatrix(1 To 18, 1 To 1)
Dim CTeamRankRange As Range
Dim CTeamMatchOutcomeRange As Range
Dim FinalPointsRange As Range
Dim WinningTeam, RequiredPoints
Dim CPosition, CRound, COutcome, CFinalOutcome
Dim RoundCat, TO1, TO2, Round

'Preliminaries
Dim SimTime
Dim StartTime

StartTime = Now()
Dim IterationCounter
Dim Iterations
Dim SecondIterations
Dim C1, C2, C3, C4
Dim PositionNumber, PositionCat
Iterations = Application.InputBox("Iterations", Type:=1)

SecondIterations = Application.InputBox("Second Level Iterations", Type:=1)

' Dim MaxRunTime

' MaxRunTime = Application.InputBox("Max run time (minutes)", Type:=1)

Dim DivisionCounter

For DivisionCounter = 1 To 2

ReDim SheetCounter(1 To 10, 1 To 1)

If DivisionCounter = 1 Then

   SheetCounter(1, 1) = "Div1Position1"
   SheetCounter(2, 1) = "Div1Position2"
   SheetCounter(3, 1) = "Div1Position3"
   SheetCounter(4, 1) = "Div1Position4"
   SheetCounter(5, 1) = "Div1Position5"
   SheetCounter(6, 1) = "Div1Position6"
   SheetCounter(7, 1) = "Div1Position7"
   SheetCounter(8, 1) = "Div1Position15"
   SheetCounter(9, 1) = "Div1Position16"
   SheetCounter(10, 1) = "Div1Position17"

Else

   SheetCounter(1, 1) = "Div2Position1"
   SheetCounter(2, 1) = "Div2Position2"
   SheetCounter(3, 1) = "Div2Position3"
   SheetCounter(4, 1) = "Div2Position4"
   SheetCounter(5, 1) = "Div2Position5"
   SheetCounter(6, 1) = "Div2Position6"
   SheetCounter(7, 1) = "Div2Position7"
   SheetCounter(8, 1) = "Div2Position15"

End If
SheetCounter(9, 1) = "Div2Position16"

SheetCounter(10, 1) = "Div2Position17"

End If

Dim PositionCounter

For PositionCounter = 1 To 10
    If PositionCounter < 8 Then
        PositionNumber = PositionCounter
        PositionCat = 0
    ElseIf PositionCounter = 8 Then
        PositionNumber = 15
        PositionCat = 1
    ElseIf PositionCounter = 9 Then
        PositionNumber = 16
        PositionCat = 1
    ElseIf PositionCounter = 10 Then
        PositionNumber = 17
        PositionCat = 1
    End If

Sheets("SimulationTemplate").Activate

Range("CE5:ND1048576").ClearContents

Dim RowCounter

Dim RowCounter2

Dim ColumnCounter

Dim ColumnCounter2

Dim ColumnCounter3

Dim MatMin

Dim MatMax
Dim TempVal, TempVal2
Dim SelectionToggle
Dim Team1, Team2
Dim P_XY_W, P_XY_D, P_XY_L
Dim SDSum, SDMean
Dim RoundCatCounter, OutcomeCounter, PrevOutcomeCounter, OppPrevOutcomeCounter

'Bayesian match outcome probability adjustments
Sheets("Contingency Table").Activate
ReDim ProbMatrix(1 To 3, 1 To 3, 1 To 3, 1 To 3)
For RoundCatCounter = 1 To 3
    For OutcomeCounter = 1 To 3
        For PrevOutcomeCounter = 1 To 3
            For OppPrevOutcomeCounter = 1 To 3
                If DivisionCounter = 1 Then
                    ProbMatrix(RoundCatCounter, OutcomeCounter, PrevOutcomeCounter, OppPrevOutcomeCounter) = Cells(33 + 3 * (OppPrevOutcomeCounter - 1) + OutcomeCounter, 31 + 3 * (RoundCatCounter - 1) + PrevOutcomeCounter)
                Else
                    ProbMatrix(RoundCatCounter, OutcomeCounter, PrevOutcomeCounter, OppPrevOutcomeCounter) = Cells(33 + 3 * (OppPrevOutcomeCounter - 1) + OutcomeCounter, 72 + 3 * (RoundCatCounter - 1) + PrevOutcomeCounter)
                End If
            Next OppPrevOutcomeCounter
        Next PrevOutcomeCounter
    Next OutcomeCounter
Next RoundCatCounter

Sheets("SimulationTemplate").Activate

'Create starting sequences
ReDim RandMatrix1(1 To 18, 1 To 2)
ReDim RandMatrix2(1 To 18, 1 To 2)
ReDim SelectionMatrix(1 To 34, 1 To 2)
ReDim PointSD(1 To SecondIterations, 1 To 34)
ReDim FinalPointSD(1 To Iterations, 1 To 34)
ReDim SimSummaryMatrix(1 To Iterations, 1 To TPositions, 1 To TRounds, 1 To 6)
For IterationCounter = 1 To Iterations
ReDim SimMatrix(1 To TPositions, 1 To TRounds, 1 To TOutcomes, 1 To TFinalOutcomes)
For C1 = 1 To TPositions
  For C2 = 1 To TRounds
    For C3 = 1 To TOutcomes
      For C4 = 1 To TFinalOutcomes
        SimMatrix(C1, C2, C3, C4) = 0
      Next C4
    Next C3
  Next C2
Next C1
For RowCounter = 1 To 18
  RandMatrix1(RowCounter, 2) = Rnd()
  RandMatrix2(RowCounter, 2) = Rnd()
Next RowCounter
RandMatrix1(1, 1) = "A"
RandMatrix2(1, 1) = "A"
RandMatrix1(2, 1) = "B"
RandMatrix2(2, 1) = "B"
RandMatrix1(3, 1) = "C"
RandMatrix2(3, 1) = "C"
RandMatrix1(4, 1) = "D"
RandMatrix2(4, 1) = "D"
RandMatrix1(5, 1) = "E"
RandMatrix2(5, 1) = "E"
RandMatrix1(6, 1) = "F"
RandMatrix2(6, 1) = "F"
RandMatrix1(7, 1) = "G"
RandMatrix2(7, 1) = "G"
RandMatrix1(8, 1) = "H"
RandMatrix2(8, 1) = "H"
RandMatrix1(9, 1) = "I"
RandMatrix2(9, 1) = "I"
RandMatrix1(10, 1) = "J"
RandMatrix2(10, 1) = "J"
RandMatrix1(11, 1) = "K"
RandMatrix2(11, 1) = "K"
RandMatrix1(12, 1) = "L"
RandMatrix2(12, 1) = "L"
RandMatrix1(13, 1) = "M"
RandMatrix2(13, 1) = "M"
RandMatrix1(14, 1) = "N"
RandMatrix2(14, 1) = "N"
RandMatrix1(15, 1) = "O"
RandMatrix2(15, 1) = "O"
RandMatrix1(16, 1) = "P"
RandMatrix2(16, 1) = "P"
RandMatrix1(17, 1) = "Q"
RandMatrix2(17, 1) = "Q"
RandMatrix1(18, 1) = "R"
RandMatrix2(18, 1) = "R"

For RowCounter = 1 To 17
    SelectionMatrix(RowCounter, 1) = 1
    SelectionMatrix(RowCounter, 2) = Rnd()
    SelectionMatrix(RowCounter + 17, 1) = 2
    SelectionMatrix(RowCounter + 17, 2) = Rnd()
Next RowCounter

'Sort matrices
MatMin = LBound(RandMatrix1)
MatMax = UBound(RandMatrix1)

For RowCounter = MatMin To MatMax - 1
    For RowCounter2 = RowCounter + 1 To MatMax
        If RandMatrix1(RowCounter, 2) > RandMatrix1(RowCounter2, 2) Then
            TempVal = RandMatrix1(RowCounter, 2)
            RandMatrix1(RowCounter, 2) = RandMatrix1(RowCounter2, 2)
            RandMatrix1(RowCounter2, 2) = TempVal
            TempVal = RandMatrix1(RowCounter, 1)
            RandMatrix1(RowCounter, 1) = RandMatrix1(RowCounter2, 1)
            RandMatrix1(RowCounter2, 1) = TempVal
        End If
    Next RowCounter2
Next RowCounter

MatMin = LBound(RandMatrix2)
MatMax = UBound(RandMatrix2)

For RowCounter = MatMin To MatMax - 1
For RowCounter2 = RowCounter + 1 To MatMax

    If RandMatrix2(RowCounter, 2) > RandMatrix2(RowCounter2, 2) Then
        TempVal = RandMatrix2(RowCounter, 2)
        RandMatrix2(RowCounter, 2) = RandMatrix2(RowCounter2, 2)
        RandMatrix2(RowCounter2, 2) = TempVal
        TempVal = RandMatrix2(RowCounter, 1)
        RandMatrix2(RowCounter, 1) = RandMatrix2(RowCounter2, 1)
        RandMatrix2(RowCounter2, 1) = TempVal
    End If

Next RowCounter2

Next RowCounter

MatMin = LBound(SelectionMatrix)
MatMax = UBound(SelectionMatrix)

For RowCounter = MatMin To MatMax - 1

    For RowCounter2 = RowCounter + 1 To MatMax

        If SelectionMatrix(RowCounter, 2) > SelectionMatrix(RowCounter2, 2) Then
            TempVal = SelectionMatrix(RowCounter, 2)
            SelectionMatrix(RowCounter, 2) = SelectionMatrix(RowCounter2, 2)
            SelectionMatrix(RowCounter2, 2) = TempVal
            TempVal = SelectionMatrix(RowCounter, 1)
            SelectionMatrix(RowCounter, 1) = SelectionMatrix(RowCounter2, 1)
            SelectionMatrix(RowCounter2, 1) = TempVal
        End If

    Next RowCounter2

Next RowCounter

'Order sequences

ReDim OddMatrix(1 To 2, 1 To 9)
ReDim EvenMatrix(1 To 2, 1 To 9)
For ColumnCounter = 1 To 9
    OddMatrix(1, ColumnCounter) = RandMatrix1(ColumnCounter, 1)
    OddMatrix(2, ColumnCounter) = RandMatrix1(19 - ColumnCounter, 1)
    EvenMatrix(1, ColumnCounter) = RandMatrix2(ColumnCounter, 1)
    EvenMatrix(2, ColumnCounter) = RandMatrix2(19 - ColumnCounter, 1)
Next ColumnCounter
'ReCreate schedule
ReDim CurrentRoundMatrix(1 To 9, 1 To 2)
Dim RoundPasteRange As Range
For ColumnCounter = 1 To 34
    Set RoundPasteRange = Range(Cells(3 + (ColumnCounter - 1) * 9, 4), Cells(11 + (ColumnCounter - 1) * 9, 5))
    If SelectionMatrix(ColumnCounter, 1) = 1 Then
        CurrentRoundMatrix = Application.Transpose(OddMatrix)
        RoundPasteRange = CurrentRoundMatrix
        TempVal = OddMatrix(2, 1)
        For ColumnCounter2 = 1 To 8
            OddMatrix(2, ColumnCounter2) = OddMatrix(2, ColumnCounter2 + 1)
        Next ColumnCounter2
        OddMatrix(2, 9) = OddMatrix(1, 9)
        For ColumnCounter2 = 2 To 8
            OddMatrix(1, 11 - ColumnCounter2) = OddMatrix(1, 10 - ColumnCounter2)
        Next ColumnCounter2
        OddMatrix(1, 2) = TempVal
    ElseIf SelectionMatrix(ColumnCounter, 1) = 2 Then
        CurrentRoundMatrix = Application.Transpose(EvenMatrix)
RoundPasteRange = CurrentRoundMatrix

    TempVal = EvenMatrix(2, 1)

For ColumnCounter2 = 1 To 8

    EvenMatrix(2, ColumnCounter2) = EvenMatrix(2, ColumnCounter2 + 1)

Next ColumnCounter2

EvenMatrix(2, 9) = EvenMatrix(1, 9)

For ColumnCounter2 = 2 To 8

    EvenMatrix(1, 11 - ColumnCounter2) = EvenMatrix(1, 10 - ColumnCounter2)

Next ColumnCounter2

    EvenMatrix(1, 2) = TempVal

End If

Next ColumnCounter

'Code to Simulate Current Season

ReDim ResultMatrix(1 To 306, 1 To 3)

ReDim LastOutcomeMatrix(1 To 18, 1 To 1)

ReDim LineUpMatrix(1 To 306, 1 To 2)

Dim LineUpRange As Range

Set LineUpRange = Range(Cells(3, 4), Cells(308, 5))

LineUpMatrix = LineUpRange

For RowCounter = 1 To 306

    LineUpMatrix(RowCounter, 1) = fTeamNumber(LineUpMatrix(RowCounter, 1))

    LineUpMatrix(RowCounter, 2) = fTeamNumber(LineUpMatrix(RowCounter, 2))

Next RowCounter

For SecondIterationCounter = 1 To SecondIterations

For RowCounter = 1 To 306

    Round = WorksheetFunction.RoundUp(RowCounter / 34, 0)

    RoundCat = fRoundCat(Round)

Next RowCounter

For SecondIterationCounter = 1 To SecondIterations
Team1 = LineUpMatrix(RowCounter, 1)
Team2 = LineUpMatrix(RowCounter, 2)
If Round > 1 Then
  TO1 = LastOutcomeMatrix(Team1, 1)
  TO2 = LastOutcomeMatrix(Team2, 1)
  P_XY_W = ProbMatrix(RoundCat, 1, TO1, TO2)
  P_XY_D = ProbMatrix(RoundCat, 2, TO1, TO2)
  P_XY_L = ProbMatrix(RoundCat, 3, TO1, TO2)
End If

ResultMatrix(RowCounter, 1) = Rnd()
If ResultMatrix(RowCounter, 1) <= fVaryProb1(RoundCat, P_XY_W, P_XY_D, P_XY_L) Then
  ResultMatrix(RowCounter, 2) = 3
  ResultMatrix(RowCounter, 3) = 0
  LastOutcomeMatrix(Team1, 1) = 1
  LastOutcomeMatrix(Team2, 1) = 3
ElseIf ResultMatrix(RowCounter, 1) <= fVaryProb2(RoundCat, P_XY_W, P_XY_D, P_XY_L) Then
  ResultMatrix(RowCounter, 2) = 1
  ResultMatrix(RowCounter, 3) = 1
  LastOutcomeMatrix(Team1, 1) = 2
  LastOutcomeMatrix(Team2, 1) = 2
Else
  ResultMatrix(RowCounter, 2) = 0
  ResultMatrix(RowCounter, 3) = 3
  LastOutcomeMatrix(Team1, 1) = 3
  LastOutcomeMatrix(Team2, 1) = 1
End If

Next RowCounter
Dim ResultPasteRange As Range
Set ResultPasteRange = Range(Cells(3, 6), Cells(308, 8))
ResultPasteRange = ResultMatrix
'Sort Points from Highest to Lowest
ActiveSheet.Calculate
Dim rngFirstRow As Range
Dim rng As Range
Dim ws As Worksheet
Range("K3:AR20").Copy
Range("AU3").Select
Selection.PasteSpecial Paste:=xlPasteValues
Application.CutCopyMode = False
Set ws = ActiveSheet
Set rngFirstRow = ws.Range("AU3:CB3")
For Each rng In rngFirstRow
    With ws.Sort
        .SortFields.Clear
        .SortFields.Add Key:=rng, Order:=xlDescending
        .SetRange ws.Range(rng, rng.End(xlDown))
        .Header = xlNo
        .MatchCase = False
        .Apply
    End With
Next rng
Set CTeamRankRange = Range(Cells(43, 11), Cells(60, 44))
Set CTeamMatchOutcomeRange = Range(Cells(23, 11), Cells(40, 44))
Set FinalPointsRange = Range(Cells(3, 44), Cells(20, 44))
CurrentTeamRankMatrix = CTeamRankRange

CurrentTeamMatchOutcomeMatrix = CTeamMatchOutcomeRange

FinalTeamPointsMatrix = FinalPointsRange

RequiredPoints = Cells((3 + (7 * PositionCat)) + (PositionCounter - 1), 80)

' Create Frequencies matrix

For RC1 = 1 To 18
    For CC1 = 1 To 33
        CPosition = CurrentTeamRankMatrix(RC1, CC1)
        CRound = CC1
        COutcome = CurrentTeamMatchOutcomeMatrix(RC1, CC1)
        TeamFinalOutcome = FinalTeamPointsMatrix(RC1, 1)
        If TeamFinalOutcome >= RequiredPoints Then
            CFinalOutcome = 1
        Else
            CFinalOutcome = 2
        End If
        SimMatrix(CPosition, CRound, COutcome, CFinalOutcome) = SimMatrix(CPosition, CRound, COutcome, CFinalOutcome) + 1
    Next CC1
Next RC1

' Standard Deviation of Points for Second Iterations

ReDim PointSDMatrix(1 To 1, 1 To 34)
Dim PointSDRange As Range
Set PointSDRange = Range(Cells(123, 11), Cells(123, 44))
PointSDMatrix = PointSDRange
For ColumnCounter = 1 To 34
    PointSD(SecondIterationCounter, ColumnCounter) = PointSDMatrix(1, ColumnCounter)
Next ColumnCounter
Next SecondIterationCounter

'Summarise Simmatrix

For RC1 = 1 To 18

For CC1 = 1 To 33

    Dim Win1, Win2, Draw1, Draw2, Loss1, Loss2, PWin, PDraw, PLoss, WinImp, DrawImp,
    ResultImp

    Win1 = SimMatrix(RC1, CC1, 1, 1)

    Win2 = SimMatrix(RC1, CC1, 1, 2)

    Draw1 = SimMatrix(RC1, CC1, 2, 1)

    Draw2 = SimMatrix(RC1, CC1, 2, 2)

    Loss1 = SimMatrix(RC1, CC1, 3, 1)

    Loss2 = SimMatrix(RC1, CC1, 3, 2)

    If Win1 + Draw1 = 0 Then
        PResult = 0
    Else
        PResult = (Win1 + Draw1) / (Win1 + Win2 + Draw1 + Draw2)
    End If

    If Loss1 = 0 Then
        PLoss = 0
    Else
        PLoss = Loss1 / (Loss1 + Loss2)
    End If

    If RC1 < PositionNumber Then
        ResultImp = 0
    ElseIf PResult < PLoss Then
        ResultImp = 0
    Else
        ResultImp = PResult - PLoss
    End If

    Next CC1

Next RC1
End If

SimSummaryMatrix(IterationCounter, RC1, CC1, 1) = PWin
SimSummaryMatrix(IterationCounter, RC1, CC1, 2) = PDraw
SimSummaryMatrix(IterationCounter, RC1, CC1, 3) = PLoss
SimSummaryMatrix(IterationCounter, RC1, CC1, 4) = WinImp
SimSummaryMatrix(IterationCounter, RC1, CC1, 5) = DrawImp
SimSummaryMatrix(IterationCounter, RC1, CC1, 6) = ResultImp

Next CC1
Next RC1

' Summarising Standard Deviation of Points

For CC1 = 1 To 34
    SDSum = 0
    SDMean = 0
    For RC1 = 1 To SecondIterations
        SDSum = SDSum + PointSD(RC1, CC1)
    Next RC1
    SDMean = SDSum / SecondIterations
    FinalPointSD(IterationCounter, CC1) = SDMean
Next CC1
Next IterationCounter

' Summarising Average Importance

Dim CuSum, CuSum2
ReDim FinalPasteWinMatrix(1 To Iterations, 1 To 33)
' ReDim FinalPasteDrawMatrix(1 To Iterations, 1 To 33)
ReDim FinalPasteResultMatrix(1 To Iterations, 1 To 33)
ReDim ImpPositionMatrix(1 To Iterations, 1 To 33, 1 To TPositions)
ReDim FinalPositionMatrix(1 To Iterations, 1 To 594)
For RC1 = 1 To Iterations
    For CC1 = 1 To 33
        CuSum = 0
        CuSum2 = 0
        For CC2 = 1 To 18
            CuSum = CuSum + SimSummaryMatrix(RC1, CC2, CC1, 6)
        Next CC2
        FinalPasteWinMatrix(RC1, CC1) = CDec(CuSum / 18)
        ' Importance for each position to finish in PositionCounter
        For CC2 = 1 To 18
            ImpPositionMatrix(RC1, CC1, CC2) = SimSummaryMatrix(RC1, CC2, CC1, 6)
        Next CC2
        FinalPasteResultMatrix(RC1, CC1) = FinalPasteWinMatrix(RC1, CC1)
    Next CC1
Next RC1

' Summarising Importance by Position to finish in First
For RC1 = 1 To Iterations
    For CC1 = 1 To 33
        For CC2 = 1 To 18
            CuSum2 = (CC1 * CC2) + ((CC2 - 1) * (33 - CC1))
            FinalPositionMatrix(RC1, CuSum2) = ImpPositionMatrix(RC1, CC1, CC2)
        Next CC2
    Next CC1
Next RC1

' Paste Values into Cells
Dim WinImpPasteRange As Range
Dim DrawImpPasteRange As Range
Dim ResultImpPasteRange As Range
Dim ImpPositionPasteRange As Range
Sheets(SheetCounter(PositionCounter, 1)).Activate
Range("B4:AZ1048576").ClearContents
'Set WinImpPasteRange = Range(Cells(4, 2), Cells(Iterations + 3, 34))
'Set DrawImpPasteRange = Range(Cells(4, 35), Cells(Iterations + 3, 67))
Set ResultImpPasteRange = Range(Cells(4, 68), Cells(Iterations + 3, 100))
Set PointSDRange = Range(Cells(4, 101), Cells(Iterations + 3, 134))
Set ImpPositionPasteRange = Range(Cells(4, 135), Cells(Iterations + 3, 728))
'WinImpPasteRange = FinalPasteWinMatrix
'DrawImpPasteRange = FinalPasteDrawMatrix
ResultImpPasteRange = FinalPasteResultMatrix
PointSDRange = FinalPointSD
ImpPositionPasteRange = FinalPositionMatrix
ActiveSheet.Calculate
Next PositionCounter
Next DivisionCounter
SimTime = Format(86440 * (Now() - StartTime), "0.0")
MsgBox ("Simulation time was " & SimTime & " seconds.")
End Sub
Appendix D: Monte Carlo simulation VBA program for NBA basketball

' Importance Simulation Procedure
Sub NBASimulation()
    Application.ScreenUpdating = False

' Preliminaries
Dim Iterations, SecondIterations
Iterations = Application.InputBox("Iterations", Type:=1)
SecondIterations = Application.InputBox("Second Level Iterations", Type:=1)
Dim StartTime
Dim SimTime
StartTime = Now()
Sheets("SimulationTemplate").Activate
Dim TPositions, TMatches, TOutcomes, TFinalOutcomes
TPositions = 15
TMatches = 81
TOutcomes = 2 ' Match Outcomes (Win = 1, Loss = 2)
TFinalOutcomes = 2 ' Achieve Required Wins Yes/No (1/2)
Dim RC1, RC2, RC3, CC1, CC2, CC3
Dim C1, C2, C3, C4
Dim IterationCounter, SecondIterationCounter
Dim CurrentRand
Dim RequiredWins
Dim CPosition, CMatch, COutcome, TeamFinalOutcome, CFinalOutcome

' Variables for Bayes' Rule application
Dim Team1, Team2

Dim P_XY_W, P_XY_L

Dim RoundCatCounter, OutcomeCounter, PrevOutcomeCounter, OppPrevOutcomeCounter

Dim RowCounter, GameT1, GameT2, GameCat, TO1, TO2

ReDim CurrentScheduleMatrix(1 To 1230, 1 To 2)

Dim PasteRange As Range

ReDim CurrentTeamRankMatrix(1 To 15, 1 To 82)

ReDim CurrentMatchOutcomeMatrix(1 To 15, 1 To 82)

ReDim FinalWinsMatrix(1 To 15, 1 To 1)

Dim TeamRankRange As Range

Dim TeamMatchOutcomeRange As Range

Dim FinalWinsRange As Range

ReDim SimSummaryMatrix(1 To Iterations, 1 To TPositions, 1 To TMatches, 1 To 3)

ReDim SheetCounter(1 To 8, 1 To 1)

SheetCounter(1, 1) = "Position1"

SheetCounter(2, 1) = "Position2"

SheetCounter(3, 1) = "Position3"

SheetCounter(4, 1) = "Position4"

SheetCounter(5, 1) = "Position5"

SheetCounter(6, 1) = "Position6"

SheetCounter(7, 1) = "Position7"

SheetCounter(8, 1) = "Position8"

'Bayesian match outcome probability adjustments

Sheets("Contingency Table").Activate

ReDim ProbMatrix(1 To 2, 1 To 2, 1 To 2, 1 To 2)

For RoundCatCounter = 1 To 2

For OutcomeCounter = 1 To 2

Next OutcomeCounter

Next RoundCatCounter
For PrevOutcomeCounter = 1 To 2
For OppPrevOutcomeCounter = 1 To 2
ProbMatrix(RoundCatCounter, OutcomeCounter, PrevOutcomeCounter,
OppPrevOutcomeCounter) = Cells(21 + 2 * (OppPrevOutcomeCounter - 1) + OutcomeCounter, 18 +
2 * (RoundCatCounter - 1) + PrevOutcomeCounter)
Next OppPrevOutcomeCounter
Next PrevOutcomeCounter
Next OutcomeCounter
Next RoundCatCounter
Sheets("SimulationTemplate").Activate
' Collect Schedules
Sheets("Schedules").Activate
ReDim ScheduleOneMatrix(1 To 1230, 1 To 2)
ReDim ScheduleTwoMatrix(1 To 1230, 1 To 2)
ReDim ScheduleThreeMatrix(1 To 1230, 1 To 2)
ReDim ScheduleFourMatrix(1 To 1230, 1 To 2)
ReDim ScheduleFiveMatrix(1 To 1230, 1 To 2)
ReDim ScheduleSixMatrix(1 To 1230, 1 To 2)
ReDim ScheduleSevenMatrix(1 To 1230, 1 To 2)
ReDim ScheduleEightMatrix(1 To 1230, 1 To 2)
ReDim ScheduleNineMatrix(1 To 1230, 1 To 2)
Dim ScheduleOneRange As Range
Dim ScheduleTwoRange As Range
Dim ScheduleThreeRange As Range
Dim ScheduleFourRange As Range
Dim ScheduleFiveRange As Range
Dim ScheduleSixRange As Range
Dim ScheduleSevenRange As Range
Dim ScheduleEightRange As Range
Dim ScheduleNineRange As Range

Set ScheduleOneRange = Range(Cells(2, 3), Cells(1231, 4))
Set ScheduleTwoRange = Range(Cells(1232, 3), Cells(2461, 4))
Set ScheduleThreeRange = Range(Cells(2462, 3), Cells(3691, 4))
Set ScheduleFourRange = Range(Cells(3692, 3), Cells(4921, 4))
Set ScheduleFiveRange = Range(Cells(4922, 3), Cells(6151, 4))
Set ScheduleSixRange = Range(Cells(6152, 3), Cells(7381, 4))
Set ScheduleSevenRange = Range(Cells(7382, 3), Cells(8611, 4))
Set ScheduleEightRange = Range(Cells(8612, 3), Cells(9841, 4))
Set ScheduleNineRange = Range(Cells(9842, 3), Cells(11071, 4))

ScheduleOneMatrix = ScheduleOneRange
ScheduleTwoMatrix = ScheduleTwoRange
ScheduleThreeMatrix = ScheduleThreeRange
ScheduleFourMatrix = ScheduleFourRange
ScheduleFiveMatrix = ScheduleFiveRange
ScheduleSixMatrix = ScheduleSixRange
ScheduleSevenMatrix = ScheduleSevenRange
ScheduleEightMatrix = ScheduleEightRange
ScheduleNineMatrix = ScheduleNineRange

' Determine Current Schedule
Dim PositionCounter
For PositionCounter = 1 To 4
Sheets("SimulationTemplate").Activate
Set PasteRange = Range(Cells(3, 4), Cells(1232, 5))
For IterationCounter = 1 To Iterations ' Schedule Iteration Counter
ReDim SimMatrix(1 To TPositions, 1 To TMatches, 1 To TOutcomes, 1 To TFinalOutcomes)

For C1 = 1 To TPositions
    For C2 = 1 To TMatches
        For C3 = 1 To TOutcomes
            For C4 = 1 To TFinalOutcomes
                SimMatrix(C1, C2, C3, C4) = 0 ' Count number of successes/failures
            Next C4
        Next C3
    Next C2
Next C1

CurrentRand = WorksheetFunction.RandBetween(1, 9) ' Determine which schedule we are using for current first level iteration
If CurrentRand = 1 Then CurrentScheduleMatrix = ScheduleOneMatrix
If CurrentRand = 2 Then CurrentScheduleMatrix = ScheduleTwoMatrix
If CurrentRand = 3 Then CurrentScheduleMatrix = ScheduleThreeMatrix
If CurrentRand = 4 Then CurrentScheduleMatrix = ScheduleFourMatrix
If CurrentRand = 5 Then CurrentScheduleMatrix = ScheduleFiveMatrix
If CurrentRand = 6 Then CurrentScheduleMatrix = ScheduleSixMatrix
If CurrentRand = 7 Then CurrentScheduleMatrix = ScheduleSevenMatrix
If CurrentRand = 8 Then CurrentScheduleMatrix = ScheduleEightMatrix
If CurrentRand = 9 Then CurrentScheduleMatrix = ScheduleNineMatrix
    PasteRange = CurrentScheduleMatrix
    ActiveSheet.Calculate
    ' Determine match outcome probabilities using Bayes' Rule
ReDim ResultMatrix(1 To 1230, 1 To 3)
ReDim LastOutcomeMatrix(1 To 30, 1 To 1)
ReDim LineUpMatrix(1 To 1230, 1 To 2)
Dim LineUpRange As Range

Set LineUpRange = Range(Cells(3, 4), Cells(1232, 5))

LineUpMatrix = LineUpRange

For RowCounter = 1 To 1230
    LineUpMatrix(RowCounter, 1) = fTeamNumber(LineUpMatrix(RowCounter, 1))
    LineUpMatrix(RowCounter, 2) = fTeamNumber(LineUpMatrix(RowCounter, 2))
Next RowCounter

For SecondIterationCounter = 1 To SecondIterations ' Results Iteration Counter
    For RowCounter = 1 To 1230
        GameT1 = Cells(RowCounter + 2, 1)
        GameT2 = Cells(RowCounter + 2, 9)
        GameCat = fGameCat(GameT1, GameT2)
        Team1 = LineUpMatrix(RowCounter, 1)
        Team2 = LineUpMatrix(RowCounter, 2)
        If GameT1 >= 2 And GameT2 >= 2 Then
            TO1 = LastOutcomeMatrix(Team1, 1)
            TO2 = LastOutcomeMatrix(Team2, 1)
            P_XY_W = ProbMatrix(GameCat, 1, TO1, TO2)
            P_XY_L = ProbMatrix(GameCat, 2, TO1, TO2)
        End If
        ' Produce match outcomes
        ResultMatrix(RowCounter, 1) = Rnd()
        If ResultMatrix(RowCounter, 1) <= fVaryProbWin(GameCat, P_XY_W, P_XY_L) Then
            ResultMatrix(RowCounter, 2) = 1
            ResultMatrix(RowCounter, 3) = 0
            LastOutcomeMatrix(Team1, 1) = 1
            LastOutcomeMatrix(Team2, 1) = 2
        End If
    Next RowCounter
Next SecondIterationCounter
Else
    ResultMatrix(RowCounter, 2) = 0
    ResultMatrix(RowCounter, 3) = 1
    LastOutcomeMatrix(Team1, 1) = 2
    LastOutcomeMatrix(Team2, 1) = 1
End If
Next RowCounter

Dim ResultPasteRange As Range
Set ResultPasteRange = Range(Cells(3, 6), Cells(1232, 8))
ResultPasteRange = ResultMatrix
' Determine Results for Selected Schedule and Sort Wins Descending
ActiveSheet.Calculate

Dim rngFirstRow As Range
Dim rng As Range
Dim ws As Worksheet
Range("K3:CN17").Copy
Range("CQ3").Select
Selection.PasteSpecial Paste:=xlPasteValues
Application.CutCopyMode = False
Set ws = ActiveSheet
Set rngFirstRow = ws.Range("CQ3:FT3")
For Each rng In rngFirstRow
    With ws.Sort
        .SortFields.Clear
        .SortFields.Add Key:=rng, Order:=xlDescending
        .SetRange ws.Range(rng, rng.End(xlDown))
        .Header = xlNo
.MatchCase = False

.Apply

End With

Next rng

Set TeamRankRange = Range(Cells(37, 11), Cells(51, 92))

Set TeamMatchOutcomeRange = Range(Cells(20, 11), Cells(34, 92))

Set FinalWinsRange = Range(Cells(3, 92), Cells(17, 92))

CurrentTeamRankMatrix = TeamRankRange

CurrentMatchOutcomeMatrix = TeamMatchOutcomeRange

FinalWinsMatrix = FinalWinsRange

RequiredWins = Cells(3 + (PositionCounter - 1), 176) ' Final Wins Total for current Position Counter

' Create Frequencies

For RC1 = 1 To 15

    For CC1 = 1 To 81

        CPosition = CurrentTeamRankMatrix(RC1, CC1)
        CMatch = CC1
        COutcome = CurrentMatchOutcomeMatrix(RC1, CC1)
        TeamFinalOutcome = FinalWinsMatrix(RC1, 1)

        If TeamFinalOutcome >= RequiredWins Then

            CFinalOutcome = 1

        Else

            CFinalOutcome = 2

        End If

        SimMatrix(CPosition, CMatch, COutcome, CFinalOutcome) = SimMatrix(CPosition, CMatch, COutcome, CFinalOutcome) + 1

    Next CC1

Next RC1
Next SecondIterationCounter

' Summarise SimMatrix

For RC1 = 1 To 15
    For CC1 = 1 To 81
        Dim Win1, Win2, Loss1, Loss2, PWin, PLoss, ResultImp
        Win1 = SimMatrix(RC1, CC1, 1, 1)
        Win2 = SimMatrix(RC1, CC1, 1, 2)
        Loss1 = SimMatrix(RC1, CC1, 2, 1)
        Loss2 = SimMatrix(RC1, CC1, 2, 2)
        If Win1 = 0 Then
            PWin = 0
        Else
            PWin = Win1 / (Win1 + Win2)
        End If
        If Loss1 = 0 Then
            PLoss = 0
        Else
            PLoss = Loss1 / (Loss1 + Loss2)
        End If
        If RC1 < PositionCounter Then
            ResultImp = 0
        ElseIf PWin < PLoss Then
            ResultImp = 0
        Else
            ResultImp = PWin - PLoss
        End If
        SimSummaryMatrix(IterationCounter, RC1, CC1, 1) = PWin
SimSummaryMatrix(IterationCounter, RC1, CC1, 2) = PLoss
SimSummaryMatrix(IterationCounter, RC1, CC1, 3) = ResultImp

Next CC1
Next RC1
Next IterationCounter
' Calculate Average Importance across all teams
Dim CuSum
ReDim FinalPasteImportanceMatrix(1 To Iterations, 1 To 81)
For RC1 = 1 To Iterations
    For CC1 = 1 To 81
        CuSum = 0
        For CC2 = 1 To 15
            CuSum = CuSum + SimSummaryMatrix(RC1, CC2, CC1, 3)
        Next CC2
    Next CC1
    Next RC1
' Paste values into Cells
Dim ImportancePasteRange As Range
Sheets(SheetCounter(PositionCounter, 1)).Activate
Range("B4:AWW1048576").ClearContents
Set ImportancePasteRange = Range(Cells(4, 2), Cells(Iterations + 3, 82))
ImportancePasteRange = FinalPasteImportanceMatrix
ActiveSheet.Calculate
Next PositionCounter

SimTime = Format(86440 * (Now() - StartTime), "0.0")
MsgBox ("Simulation time was " & SimTime & " seconds.")
End Sub

' Determine round category for NBA match

Public Function fGameCat(GameT1, GameT2) As Integer
    If GameT1 + GameT2 <= 3 Then
        fGameCat = 0
    ElseIf GameT1 < 42 Then
        fGameCat = 1
    Else
        fGameCat = 2
    End If
End Function

' Cumulative probability of win

Public Function fVaryProbWin(GameCat, P_XY_W, P_XY_L)
    Dim P_W, P_L
    P_W = 0.5
    P_L = 0.5
    If GameCat = 0 Then
        fVaryProbWin = P_W
    Else
        Dim Numerator, Denominator
        Numerator = P_W * P_XY_W
        Denominator = Numerator + (P_L * P_XY_L)
        fVaryProbWin = CDec(Numerator / Denominator)
    End If
End Function
'Cumulative probability of loss

Public Function fVaryProbLose(GameCat, P_XY_W, P_XY_L)

Dim P_W, P_L
P_W = 0.5
P_L = 0.5
If GameCat = 0 Then
fVaryProbLose = 1 - P_L
Else
Dim Numerator, Denominator
Numerator = P_L * P_XY_L
Denominator = Numerator + (P_W * P_XY_W)
fVaryProbDiv1Lose = CDec(Numerator / Denominator)
End If
End Function

Public Function fTeamNumber(Letter) As Integer
If Letter = "AAA" Then
    fTeamNumber = 1
ElseIf Letter = "AAB" Then
    fTeamNumber = 2
ElseIf Letter = "AAC" Then
    fTeamNumber = 3
ElseIf Letter = "AAD" Then
    fTeamNumber = 4
ElseIf Letter = "AAE" Then
    fTeamNumber = 5
ElseIf Letter = "ABA" Then
    fTeamNumber = 6

261
ElseIf Letter = "ABB" Then
    fTeamNumber = 7
ElseIf Letter = "ABC" Then
    fTeamNumber = 8
ElseIf Letter = "ABD" Then
    fTeamNumber = 9
ElseIf Letter = "ABE" Then
    fTeamNumber = 10
ElseIf Letter = "ACA" Then
    fTeamNumber = 11
ElseIf Letter = "ACB" Then
    fTeamNumber = 12
ElseIf Letter = "ACC" Then
    fTeamNumber = 13
ElseIf Letter = "ACD" Then
    fTeamNumber = 14
ElseIf Letter = "ACE" Then
    fTeamNumber = 15
ElseIf Letter = "BAA" Then
    fTeamNumber = 16
ElseIf Letter = "BAB" Then
    fTeamNumber = 17
ElseIf Letter = "BAC" Then
    fTeamNumber = 18
ElseIf Letter = "BAD" Then
    fTeamNumber = 19
ElseIf Letter = "BAE" Then
fTeamNumber = 20
ElseIf Letter = "BBA" Then
    fTeamNumber = 21
ElseIf Letter = "BBB" Then
    fTeamNumber = 22
ElseIf Letter = "BBC" Then
    fTeamNumber = 23
ElseIf Letter = "BBD" Then
    fTeamNumber = 24
ElseIf Letter = "BBE" Then
    fTeamNumber = 25
ElseIf Letter = "BCA" Then
    fTeamNumber = 26
ElseIf Letter = "BCB" Then
    fTeamNumber = 27
ElseIf Letter = "BCC" Then
    fTeamNumber = 28
ElseIf Letter = "BCD" Then
    fTeamNumber = 29
ElseIf Letter = "BCE" Then
    fTeamNumber = 30
End If
End Function
### Appendix E: Binning cut-off values for Elo models

#### Appendix E-1: Bundesliga cut-off values

<table>
<thead>
<tr>
<th>League Champions</th>
<th>Championship League</th>
<th>Europa League</th>
<th>Positioning</th>
<th>Religations</th>
<th>Positional Stats</th>
<th>Outcome Stats</th>
<th>Total Stats</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Div. 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favourite</td>
<td>≥ 0.0962</td>
<td>≥ 0.3142</td>
<td>≥ 0.3076</td>
<td>≥ 0.0001</td>
<td>≥ 0.2875</td>
<td>≥ 0.0005</td>
<td>≥ 0.0072</td>
</tr>
<tr>
<td>M</td>
<td>0.0033:0.0961</td>
<td>0.0772:0.3141</td>
<td>0.0001:0.3075</td>
<td>0.1644:0.2874</td>
<td>0.0002:0.0002</td>
<td>0.3289:0.0071</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>≤ 0.0032</td>
<td>≤ 0.0771</td>
<td>≤ 0.0000</td>
<td>≤ 0.0000</td>
<td>≤ 0.1613</td>
<td>≤ 0.0091</td>
<td>≤ 0.3298</td>
</tr>
<tr>
<td>Unlucky</td>
<td>≥ 0.3560</td>
<td>≥ 0.2301</td>
<td>≥ 0.3021</td>
<td>≥ 0.0001</td>
<td>≥ 0.2403</td>
<td>≥ 0.0072</td>
<td>≥ 0.6127</td>
</tr>
</tbody>
</table>

| **Div. 2** | | | | | | | |
| Favourite | ≥ 0.2119 | ≥ 0.1318 | ≥ 0.0001 | ≥ 0.0001 | ≥ 0.2974 | ≥ 0.0005 | ≥ 0.0069 |
| M | 0.0358:0.2118 | 0.0551:0.1317 | 0.0002:0.0002 | 0.0002:0.0002 | 0.3614:0.0088 |
| L | ≤ 0.4307 | ≤ 0.0090 | ≤ 0.0000 | ≤ 0.0000 | ≤ 0.3283 | ≤ 0.0001 | ≤ 0.5613 |
| Unlucky | ≥ 0.1576 | ≥ 0.0686 | ≥ 0.0001 | ≥ 0.0147 | ≤ 0.0254 | ≤ 0.0063 | ≤ 0.0463 |

| **Position** | | | | | | | |
| **Div. 1** | | | | | | | |
| Favourite | ≥ 0.0968 | ≥ 0.03142 | ≥ 0.3031 | ≥ 0.0001 | ≥ 0.2974 | ≥ 0.0005 | ≥ 0.0069 |
| M | 0.0033:0.0968 | 0.0414:0.3141 | 0.0001:0.3031 | 0.1640:0.2974 | 0.0004:0.0002 | 0.3289:0.0068 |
| L | ≤ 0.0032 | ≤ 0.0013 | ≤ 0.0000 | ≤ 0.0000 | ≤ 0.1619 | ≤ 0.0095 | ≤ 0.3288 |
| Unlucky | ≥ 0.3000 | ≥ 0.2301 | ≥ 0.3050 | ≥ 0.0001 | ≥ 0.2578 | ≥ 0.0075 | ≤ 0.6133 |

| **Div. 2** | | | | | | | |
| Favourite | ≥ 0.0968 | ≥ 0.03142 | ≥ 0.3031 | ≥ 0.0001 | ≥ 0.2974 | ≥ 0.0005 | ≥ 0.0069 |
| M | 0.0358:0.2118 | 0.0551:0.1317 | 0.0002:0.0002 | 0.0002:0.0002 | 0.3614:0.0088 |
| L | ≤ 0.4307 | ≤ 0.0090 | ≤ 0.0000 | ≤ 0.0000 | ≤ 0.3283 | ≤ 0.0001 | ≤ 0.5613 |

| **Table** | | | | | | | |
| **Div. 1** | | | | | | | |
| Favourite | ≥ 0.0962 | ≥ 0.3801 | ≥ 0.3011 | ≥ 0.0001 | ≥ 0.2948 | ≥ 0.0002 | ≥ 0.0064 |
| M | 0.0037:0.0961 | 0.0700:0.3800 | 0.0001:0.3010 | 0.1610:0.2948 | 0.0006:0.0001 | 0.3279:0.0063 |
| L | ≤ 0.0026 | ≤ 0.0702 | ≤ 0.0000 | ≤ 0.0000 | ≤ 0.1609 | ≤ 0.0085 | ≤ 0.3278 |
| Unlucky | ≥ 0.3075 | ≥ 0.2305 | ≥ 0.3008 | ≥ 0.0001 | ≥ 0.2500 | ≥ 0.0076 | ≥ 0.6138 |

| **Div. 2** | | | | | | | |
| Favourite | ≥ 0.0962 | ≥ 0.3801 | ≥ 0.3011 | ≥ 0.0001 | ≥ 0.2948 | ≥ 0.0002 | ≥ 0.0064 |
| M | 0.0358:0.2118 | 0.0551:0.1317 | 0.0002:0.0002 | 0.0002:0.0002 | 0.3614:0.0088 |
| L | ≤ 0.4307 | ≤ 0.0090 | ≤ 0.0000 | ≤ 0.0000 | ≤ 0.3283 | ≤ 0.0001 | ≤ 0.5613 |

| **Total** | | | | | | | |
| **Div. 1** | | | | | | | |
| Favourite | ≥ 0.0962 | ≥ 0.3801 | ≥ 0.3011 | ≥ 0.0001 | ≥ 0.2948 | ≥ 0.0002 | ≥ 0.0064 |
| M | 0.0037:0.0961 | 0.0700:0.3800 | 0.0001:0.3010 | 0.1610:0.2948 | 0.0006:0.0001 | 0.3279:0.0063 |
| L | ≤ 0.0026 | ≤ 0.0702 | ≤ 0.0000 | ≤ 0.0000 | ≤ 0.1609 | ≤ 0.0085 | ≤ 0.3278 |

| **Div. 2** | | | | | | | |
| Favourite | ≥ 0.0962 | ≥ 0.3801 | ≥ 0.3011 | ≥ 0.0001 | ≥ 0.2948 | ≥ 0.0002 | ≥ 0.0064 |
| M | 0.0358:0.2118 | 0.0551:0.1317 | 0.0002:0.0002 | 0.0002:0.0002 | 0.3614:0.0088 |
| L | ≤ 0.4307 | ≤ 0.0090 | ≤ 0.0000 | ≤ 0.0000 | ≤ 0.3283 | ≤ 0.0001 | ≤ 0.5613 |

| **264** | | | | | | | |
## Appendix E-2: NBA cut-off values

<table>
<thead>
<tr>
<th>Region</th>
<th>Outcome</th>
<th>Top 1</th>
<th>Positions 2-5</th>
<th>Positional Sum</th>
<th>Outcome Sum</th>
<th>Total Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>Favourite</td>
<td>H</td>
<td>≥ 0.0255</td>
<td>≥ 0.0861</td>
<td>≥ 0.8773</td>
<td>≥ 0.6945</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0003-0.0274</td>
<td>0.0312-0.0960</td>
<td>0.0001-0.0772</td>
<td>0.335-0.0644</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0002</td>
<td>≤ 0.0311</td>
<td>≤ 0.8000</td>
<td>≤ 0.6334</td>
</tr>
<tr>
<td></td>
<td>Underdog</td>
<td>H</td>
<td>≥ 0.067</td>
<td>≥ 0.0767</td>
<td>≥ 0.9063</td>
<td>≥ 0.8308</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0001-0.0096</td>
<td>0.0028-0.0069</td>
<td>0.0111-0.0062</td>
<td>0.0116-0.0049</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0000</td>
<td>≤ 0.0029</td>
<td>≤ 0.5110</td>
<td>≤ 0.4115</td>
</tr>
<tr>
<td>West</td>
<td>Favourite</td>
<td>H</td>
<td>≥ 0.0206</td>
<td>≥ 0.0199</td>
<td>≥ 0.8000</td>
<td>≥ 0.8771</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0016-0.0208</td>
<td>0.0302-0.0198</td>
<td>0.0000-0.0096</td>
<td>0.303-0.0758</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0017</td>
<td>≤ 0.0042</td>
<td>≤ 0.8000</td>
<td>≤ 0.6334</td>
</tr>
<tr>
<td></td>
<td>Underdog</td>
<td>H</td>
<td>≥ 0.085</td>
<td>≥ 0.0845</td>
<td>≥ 0.9068</td>
<td>≥ 0.8428</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0001-0.0092</td>
<td>0.0020-0.0094</td>
<td>0.0007-0.0007</td>
<td>0.0002-0.0043</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0000</td>
<td>≤ 0.0009</td>
<td>≤ 0.8006</td>
<td>≤ 0.6041</td>
</tr>
<tr>
<td>F3/D3</td>
<td>Favourite</td>
<td>H</td>
<td>≥ 0.0252</td>
<td>≥ 0.0944</td>
<td>≥ 0.9705</td>
<td>≥ 0.6038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0002-0.0271</td>
<td>0.0201-0.0403</td>
<td>0.0001-0.0784</td>
<td>0.3528-0.0072</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0001</td>
<td>≤ 0.0029</td>
<td>≤ 0.8000</td>
<td>≤ 0.6375</td>
</tr>
<tr>
<td></td>
<td>Underdog</td>
<td>H</td>
<td>≥ 0.0075</td>
<td>≥ 0.0085</td>
<td>≥ 0.9068</td>
<td>≥ 0.8522</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0001-0.0074</td>
<td>0.0028-0.0084</td>
<td>0.0088-0.0067</td>
<td>0.0123-0.0071</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0000</td>
<td>≤ 0.0028</td>
<td>≤ 0.8017</td>
<td>≤ 0.6122</td>
</tr>
<tr>
<td>F3/D3</td>
<td>Favourite</td>
<td>H</td>
<td>≥ 0.0408</td>
<td>≥ 0.0138</td>
<td>≥ 0.8005</td>
<td>≥ 0.9742</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0016-0.0267</td>
<td>0.0306-0.0157</td>
<td>0.0000-0.0004</td>
<td>0.3732-0.0072</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0015</td>
<td>≤ 0.0039</td>
<td>≤ 0.8000</td>
<td>≤ 0.6371</td>
</tr>
<tr>
<td></td>
<td>Underdog</td>
<td>H</td>
<td>≥ 0.0003</td>
<td>≥ 0.0072</td>
<td>≥ 0.9005</td>
<td>≥ 0.8446</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0001-0.0032</td>
<td>0.0001-0.0001</td>
<td>0.0006-0.0004</td>
<td>0.0032-0.0045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0000</td>
<td>≤ 0.0003</td>
<td>≤ 0.8003</td>
<td>≤ 0.6005</td>
</tr>
<tr>
<td>F3/D3</td>
<td>Favourite</td>
<td>H</td>
<td>≥ 0.0255</td>
<td>≥ 0.0900</td>
<td>≥ 0.9705</td>
<td>≥ 0.6045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0002-0.0274</td>
<td>0.0312-0.0099</td>
<td>0.0001-0.0764</td>
<td>0.3355-0.0064</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0002</td>
<td>≤ 0.0031</td>
<td>≤ 0.8000</td>
<td>≤ 0.6374</td>
</tr>
<tr>
<td></td>
<td>Underdog</td>
<td>H</td>
<td>≥ 0.0067</td>
<td>≥ 0.0071</td>
<td>≥ 0.9005</td>
<td>≥ 0.8432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0001-0.0066</td>
<td>0.0027-0.0078</td>
<td>0.0115-0.0064</td>
<td>0.0114-0.0058</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0000</td>
<td>≤ 0.0026</td>
<td>≤ 0.8114</td>
<td>≤ 0.6113</td>
</tr>
<tr>
<td>F3/D3</td>
<td>Favourite</td>
<td>H</td>
<td>≥ 0.0600</td>
<td>≥ 0.0186</td>
<td>≥ 0.9013</td>
<td>≥ 0.8947</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0016-0.0268</td>
<td>0.0313-0.0109</td>
<td>0.0000-0.0011</td>
<td>0.3603-0.0078</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0017</td>
<td>≤ 0.0048</td>
<td>≤ 0.8000</td>
<td>≤ 0.6382</td>
</tr>
<tr>
<td></td>
<td>Underdog</td>
<td>H</td>
<td>≥ 0.0004</td>
<td>≥ 0.0082</td>
<td>≥ 0.9000</td>
<td>≥ 0.8428</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>0.0001-0.0033</td>
<td>0.0010-0.0031</td>
<td>0.0005-0.0009</td>
<td>0.0042-0.0027</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>≤ 0.0000</td>
<td>≤ 0.0009</td>
<td>≤ 0.8004</td>
<td>≤ 0.6041</td>
</tr>
</tbody>
</table>
Appendix F: NBA additional Elo models

This section provides the change in correct prediction percentage results for the NBA Pythagorean and FiveThirtyEight Elo models.
## Appendix F-1: NBA Pythagorean Elo – “Conference Champion” and “Top 4”

### Conference Champions vs. Top 4

<table>
<thead>
<tr>
<th>Team</th>
<th>East 1</th>
<th>East 2</th>
<th>West 1</th>
<th>West 2</th>
<th>Overall</th>
<th>East 1</th>
<th>East 2</th>
<th>West 1</th>
<th>West 2</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>80.27 (61)</td>
<td>56.56 (153)</td>
<td>72.47 (135)</td>
<td>55.78 (162)</td>
<td>66.67 (34)</td>
<td>80.27 (61)</td>
<td>56.56 (153)</td>
<td>72.47 (135)</td>
<td>55.78 (162)</td>
<td>66.67 (34)</td>
</tr>
<tr>
<td>2.</td>
<td>68.34 (89)</td>
<td>72.15 (126)</td>
<td>72.52 (127)</td>
<td>72.26 (126)</td>
<td>72.34 (58)</td>
<td>68.34 (89)</td>
<td>72.15 (126)</td>
<td>72.52 (127)</td>
<td>72.26 (126)</td>
<td>72.34 (58)</td>
</tr>
<tr>
<td>3.</td>
<td>73.91 (23)</td>
<td>56.36 (130)</td>
<td>75.80 (20)</td>
<td>57.15 (42)</td>
<td>69.30 (23)</td>
<td>73.91 (23)</td>
<td>56.36 (130)</td>
<td>75.80 (20)</td>
<td>57.15 (42)</td>
<td>69.30 (23)</td>
</tr>
<tr>
<td>4.</td>
<td>80.81 (9)</td>
<td>83.72 (12)</td>
<td>83.72 (12)</td>
<td>83.72 (12)</td>
<td>83.72 (12)</td>
<td>80.81 (9)</td>
<td>83.72 (12)</td>
<td>83.72 (12)</td>
<td>83.72 (12)</td>
<td>83.72 (12)</td>
</tr>
</tbody>
</table>

### Separations

<table>
<thead>
<tr>
<th>Team</th>
<th>Separation</th>
<th>2018/19</th>
<th>2019/20</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>26.15 (26)</td>
<td>56.56 (153)</td>
<td>56.56 (153)</td>
<td>56.56 (153)</td>
</tr>
<tr>
<td>2.</td>
<td>24.82 (24)</td>
<td>72.47 (135)</td>
<td>72.47 (135)</td>
<td>72.47 (135)</td>
</tr>
<tr>
<td>3.</td>
<td>23.26 (23)</td>
<td>55.78 (162)</td>
<td>55.78 (162)</td>
<td>55.78 (162)</td>
</tr>
<tr>
<td>4.</td>
<td>21.17 (21)</td>
<td>66.67 (34)</td>
<td>66.67 (34)</td>
<td>66.67 (34)</td>
</tr>
</tbody>
</table>

Sample size presented in parentheses.
## Appendix F-2: NBA Pythagorean Elo – “Positions 5-8” and “Positional Sum”

<table>
<thead>
<tr>
<th>East</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>61.49 (4)</td>
<td>25.08 (16)</td>
<td>40.01 (16)</td>
<td>66.14 (15)</td>
<td>77.13 (15)</td>
<td>60.01 (16)</td>
<td>50.08 (16)</td>
<td>53.77 (15)</td>
</tr>
<tr>
<td>NM</td>
<td>71.48 (6)</td>
<td>65.22 (40)</td>
<td>57.87 (64)</td>
<td>63.68 (52)</td>
<td>71.43 (49)</td>
<td>70.02 (36)</td>
<td>71.41 (36)</td>
<td>70.59 (17)</td>
</tr>
<tr>
<td>ML</td>
<td>57.35 (16)</td>
<td>66.01 (78)</td>
<td>58.82 (38)</td>
<td>50.39 (32)</td>
<td>94.41 (38)</td>
<td>52.17 (33)</td>
<td>84.62 (36)</td>
<td>27.27 (11)</td>
</tr>
<tr>
<td>MM</td>
<td>84.62 (13)</td>
<td>83.11 (6)</td>
<td>84.67 (9)</td>
<td>64.08 (15)</td>
<td>75.08 (46)</td>
<td>68.05 (15)</td>
<td>71.76 (37)</td>
<td>68.08 (89)</td>
</tr>
<tr>
<td>MH</td>
<td>66.47 (8)</td>
<td>57.11 (47)</td>
<td>65.29 (52)</td>
<td>67.24 (44)</td>
<td>78.08 (90)</td>
<td>75.06 (29)</td>
<td>71.41 (39)</td>
<td>63.16 (19)</td>
</tr>
<tr>
<td>MLH</td>
<td>72.55 (22)</td>
<td>54.17 (26)</td>
<td>59.09 (22)</td>
<td>64.66 (47)</td>
<td>72.12 (48)</td>
<td>71.40 (7)</td>
<td>81.99 (23)</td>
<td>69.07 (9)</td>
</tr>
<tr>
<td>LL</td>
<td>53.75 (27)</td>
<td>70.00 (20)</td>
<td>63.84 (22)</td>
<td>65.77 (57)</td>
<td>73.22 (61)</td>
<td>70.45 (32)</td>
<td>64.79 (236)</td>
<td>50.06 (38)</td>
</tr>
<tr>
<td>LM</td>
<td>74.10 (51)</td>
<td>78.07 (43)</td>
<td>76.81 (69)</td>
<td>64.71 (68)</td>
<td>78.36 (57)</td>
<td>83.33 (48)</td>
<td>70.89 (56)</td>
<td>83.27 (40)</td>
</tr>
<tr>
<td>LH</td>
<td>68.87 (106)</td>
<td>77.51 (49)</td>
<td>80.77 (52)</td>
<td>60.00 (38)</td>
<td>92.59 (27)</td>
<td>65.12 (43)</td>
<td>73.33 (35)</td>
<td>56.00 (48)</td>
</tr>
<tr>
<td>Overall</td>
<td>64.95</td>
<td>67.39</td>
<td>66.99</td>
<td>62.39</td>
<td>77.15</td>
<td>70.52</td>
<td>79.47</td>
<td>64.63</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>West</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>63.84 (17)</td>
<td>48.00 (35)</td>
<td>65.18 (14)</td>
<td>56.08 (25)</td>
<td>77.08 (30)</td>
<td>92.00 (40)</td>
<td>58.41 (56)</td>
<td>59.36 (59)</td>
</tr>
<tr>
<td>NM</td>
<td>78.15 (19)</td>
<td>68.63 (51)</td>
<td>75.11 (61)</td>
<td>70.03 (48)</td>
<td>78.79 (31)</td>
<td>64.29 (26)</td>
<td>72.73 (22)</td>
<td>76.02 (13)</td>
</tr>
<tr>
<td>ML</td>
<td>64.97 (17)</td>
<td>70.31 (31)</td>
<td>60.00 (58)</td>
<td>70.53 (24)</td>
<td>60.00 (36)</td>
<td>81.25 (16)</td>
<td>86.67 (50)</td>
<td>82.00 (50)</td>
</tr>
<tr>
<td>MLH</td>
<td>87.70 (6)</td>
<td>33.33 (6)</td>
<td>50.08 (8)</td>
<td>60.00 (15)</td>
<td>61.67 (26)</td>
<td>65.96 (47)</td>
<td>80.76 (57)</td>
<td>82.00 (50)</td>
</tr>
<tr>
<td>MH</td>
<td>0.60 (2)</td>
<td>51.42 (26)</td>
<td>71.42 (42)</td>
<td>63.04 (38)</td>
<td>63.77 (69)</td>
<td>63.41 (44)</td>
<td>64.29 (26)</td>
<td>57.14 (14)</td>
</tr>
<tr>
<td>MLH</td>
<td>68.87 (20)</td>
<td>55.37 (20)</td>
<td>67.86 (28)</td>
<td>50.00 (18)</td>
<td>77.27 (22)</td>
<td>45.72 (16)</td>
<td>42.86 (16)</td>
<td>81.82 (11)</td>
</tr>
<tr>
<td>LL</td>
<td>66.87 (21)</td>
<td>63.44 (22)</td>
<td>52.03 (21)</td>
<td>71.43 (21)</td>
<td>64.47 (78)</td>
<td>63.88 (38)</td>
<td>70.01 (279)</td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>76.82 (17)</td>
<td>73.47 (45)</td>
<td>78.76 (56)</td>
<td>60.80 (83)</td>
<td>71.43 (63)</td>
<td>75.00 (48)</td>
<td>68.59 (89)</td>
<td>51.91 (44)</td>
</tr>
<tr>
<td>LH</td>
<td>65.79 (188)</td>
<td>68.42 (56)</td>
<td>72.09 (68)</td>
<td>83.01 (37)</td>
<td>70.87 (15)</td>
<td>75.00 (37)</td>
<td>68.87 (10)</td>
<td>67.90 (40)</td>
</tr>
<tr>
<td>Overall</td>
<td>68.98</td>
<td>67.56</td>
<td>69.41</td>
<td>68.39</td>
<td>66.07</td>
<td>70.20</td>
<td>71.38</td>
<td>72.34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interaction</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>48.80 (16)</td>
<td>52.31 (24)</td>
<td>65.18 (14)</td>
<td>56.08 (25)</td>
<td>77.08 (30)</td>
<td>92.00 (40)</td>
<td>58.41 (56)</td>
<td>59.36 (59)</td>
</tr>
<tr>
<td>NM</td>
<td>77.35 (49)</td>
<td>76.97 (49)</td>
<td>59.13 (86)</td>
<td>60.68 (12)</td>
<td>78.08 (65)</td>
<td>61.66 (44)</td>
<td>73.81 (23)</td>
<td>83.33 (6)</td>
</tr>
<tr>
<td>ML</td>
<td>62.39 (56)</td>
<td>67.57 (111)</td>
<td>61.78 (80)</td>
<td>59.36 (48)</td>
<td>65.87 (57)</td>
<td>82.26 (32)</td>
<td>72.20 (38)</td>
<td>68.07 (3)</td>
</tr>
<tr>
<td>MM</td>
<td>80.10 (10)</td>
<td>89.31 (15)</td>
<td>62.50 (9)</td>
<td>58.08 (31)</td>
<td>64.64 (45)</td>
<td>65.16 (36)</td>
<td>63.97 (98)</td>
<td>58.21 (17)</td>
</tr>
<tr>
<td>MH</td>
<td>71.35 (17)</td>
<td>69.04 (25)</td>
<td>62.20 (57)</td>
<td>63.88 (49)</td>
<td>62.10 (92)</td>
<td>62.75 (90)</td>
<td>57.85 (90)</td>
<td>66.00 (37)</td>
</tr>
<tr>
<td>MLH</td>
<td>66.47 (27)</td>
<td>63.83 (47)</td>
<td>45.16 (33)</td>
<td>51.28 (39)</td>
<td>46.08 (20)</td>
<td>65.32 (28)</td>
<td>70.83 (26)</td>
<td>100.00 (8)</td>
</tr>
<tr>
<td>LL</td>
<td>52.63 (9)</td>
<td>59.25 (32)</td>
<td>65.33 (36)</td>
<td>65.85 (41)</td>
<td>66.67 (61)</td>
<td>78.22 (101)</td>
<td>71.49 (13)</td>
<td>66.25 (98)</td>
</tr>
<tr>
<td>LM</td>
<td>80.72 (27)</td>
<td>82.31 (51)</td>
<td>75.50 (85)</td>
<td>67.78 (86)</td>
<td>78.57 (84)</td>
<td>74.00 (68)</td>
<td>69.33 (48)</td>
<td>66.55 (16)</td>
</tr>
<tr>
<td>LH</td>
<td>71.02 (123)</td>
<td>75.70 (136)</td>
<td>70.24 (86)</td>
<td>60.49 (59)</td>
<td>56.06 (15)</td>
<td>70.21 (47)</td>
<td>76.82 (36)</td>
<td>66.87 (10)</td>
</tr>
<tr>
<td>Overall</td>
<td>65.67</td>
<td>78.19</td>
<td>64.74</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>78.35</td>
<td>68.88</td>
</tr>
</tbody>
</table>

**Sample size presented in parenthesis**
Appendix F-3: NBA Pythagorean Elo – “Outcome Sum” and “Total Sum”

<table>
<thead>
<tr>
<th>Outcome</th>
<th>LL</th>
<th>LM</th>
<th>LH</th>
<th>MH</th>
<th>MM</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>West</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>64.95</td>
<td>67.39</td>
<td>66.69</td>
<td>62.38</td>
<td>77.11</td>
<td>70.52</td>
</tr>
<tr>
<td>Midwest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>62.24</td>
<td>65.42</td>
<td>65.58</td>
<td>64.28</td>
<td>67.31</td>
<td>65.76</td>
</tr>
<tr>
<td>East</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>64.61</td>
<td>67.90</td>
<td>65.95</td>
<td>68.00</td>
<td>69.12</td>
<td>68.12</td>
</tr>
<tr>
<td>Inter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>63.86</td>
<td>65.85</td>
<td>64.51</td>
<td>65.76</td>
<td>65.68</td>
<td>65.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>West</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>64.43</td>
<td>67.39</td>
<td>66.69</td>
<td>62.38</td>
<td>77.11</td>
<td>70.52</td>
<td>78.47</td>
<td>64.43</td>
</tr>
<tr>
<td>Midwest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>64.12</td>
<td>67.90</td>
<td>65.95</td>
<td>68.00</td>
<td>69.12</td>
<td>68.12</td>
<td>67.94</td>
<td>66.42</td>
</tr>
<tr>
<td>East</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>63.41</td>
<td>65.42</td>
<td>65.58</td>
<td>64.28</td>
<td>67.31</td>
<td>65.76</td>
<td>67.80</td>
<td>63.17</td>
</tr>
<tr>
<td>Inter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>63.62</td>
<td>65.61</td>
<td>64.51</td>
<td>65.76</td>
<td>65.68</td>
<td>65.29</td>
<td>65.20</td>
<td>64.49</td>
</tr>
</tbody>
</table>

Sample size presented in parenthesis
### NBA FiveThirtyEight Elo – “Conference Champion” and “Top 4”

#### Conference Champs

<table>
<thead>
<tr>
<th>Team</th>
<th>Overall</th>
<th>East</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>66.47 (337)</td>
<td>60.16 (123)</td>
<td>66.74</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM</td>
<td>66.47 (337)</td>
<td>60.16 (123)</td>
<td>65.78</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HH</td>
<td>65.67</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>64.61 (205)</td>
<td>65.38 (22)</td>
<td>63.41</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>63.41</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH</td>
<td>62.50 (88)</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>61.65 (133)</td>
<td>65.78</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LH</td>
<td>60.61 (123)</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>66.47</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Top 4

<table>
<thead>
<tr>
<th>Team</th>
<th>Overall</th>
<th>East</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>70.19</td>
<td>64.74</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM</td>
<td>70.19</td>
<td>64.74</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HH</td>
<td>69.89</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>68.24</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>67.31</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH</td>
<td>66.26</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>65.17</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LH</td>
<td>64.18</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>70.19</td>
<td>65.53</td>
<td>65.78</td>
<td>69.89</td>
<td>70.35</td>
<td>69.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample size presented in parentheses.
## Appendix F: NBA FiftyThirtyEight Elo – “Positions 5-8” and “Positional Sum”

<table>
<thead>
<tr>
<th>Position 5-8</th>
<th>Positional Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>East</strong></td>
<td><strong>West</strong></td>
</tr>
<tr>
<td>Team</td>
<td>Points (Games)</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td></td>
</tr>
<tr>
<td>1.000 (1)</td>
<td>26.01 (5)</td>
</tr>
<tr>
<td>0.97 (2)</td>
<td>22.02 (5)</td>
</tr>
<tr>
<td>0.95 (3)</td>
<td>18.18 (5)</td>
</tr>
<tr>
<td>0.93 (4)</td>
<td>14.44 (5)</td>
</tr>
<tr>
<td>0.91 (5)</td>
<td>10.76 (5)</td>
</tr>
<tr>
<td><strong>East</strong></td>
<td></td>
</tr>
<tr>
<td>Team</td>
<td>Points (Games)</td>
</tr>
<tr>
<td><strong>Interaction</strong></td>
<td></td>
</tr>
<tr>
<td>1.000 (1)</td>
<td>26.01 (5)</td>
</tr>
<tr>
<td>0.97 (2)</td>
<td>22.02 (5)</td>
</tr>
<tr>
<td>0.95 (3)</td>
<td>18.18 (5)</td>
</tr>
<tr>
<td>0.93 (4)</td>
<td>14.44 (5)</td>
</tr>
<tr>
<td>0.91 (5)</td>
<td>10.76 (5)</td>
</tr>
</tbody>
</table>

Sample rates presented in parentheses.
## Appendix F-6: NBA FiveThirtyEight Elo – “Outcome Sum” and “Total Sum”

<table>
<thead>
<tr>
<th>Outcome Sum</th>
<th>Overall</th>
<th>LH</th>
<th>LL</th>
<th>MM</th>
<th>ML</th>
<th>LH</th>
<th>LM</th>
<th>LL</th>
<th>MM</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>70.09 (107)</td>
<td>64.06 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (15)</td>
<td>32.26 (15)</td>
<td>27.07 (15)</td>
<td>22.18 (15)</td>
<td>18.39 (15)</td>
</tr>
<tr>
<td>F</td>
<td>61.54 (26)</td>
<td>66.07 (26)</td>
<td>50.00 (5)</td>
<td>40.61 (4)</td>
<td>37.19 (4)</td>
<td>30.00 (3)</td>
<td>24.61 (3)</td>
<td>20.12 (3)</td>
<td>18.39 (3)</td>
<td>15.67 (3)</td>
</tr>
<tr>
<td>Outcome</td>
<td>70.19 (107)</td>
<td>64.78 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
<td>27.07 (14)</td>
<td>22.18 (14)</td>
<td>18.39 (14)</td>
</tr>
<tr>
<td>Total Sum</td>
<td>100.00 (12)</td>
<td>70.09 (107)</td>
<td>64.06 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
<td>27.07 (14)</td>
<td>22.18 (14)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>70.09 (107)</td>
<td>64.06 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
<td>27.07 (14)</td>
</tr>
<tr>
<td>Overall</td>
<td>70.09 (107)</td>
<td>64.06 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
<td>27.07 (14)</td>
</tr>
<tr>
<td>F</td>
<td>70.19 (107)</td>
<td>64.78 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
<td>27.07 (14)</td>
</tr>
<tr>
<td>Total Sum</td>
<td>100.00 (12)</td>
<td>70.09 (107)</td>
<td>64.06 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>West</td>
<td>70.09 (107)</td>
<td>64.06 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
<td>27.07 (14)</td>
</tr>
<tr>
<td>Overall</td>
<td>70.09 (107)</td>
<td>64.06 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
<td>27.07 (14)</td>
</tr>
<tr>
<td>F</td>
<td>70.19 (107)</td>
<td>64.78 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
<td>27.07 (14)</td>
</tr>
<tr>
<td>Total Sum</td>
<td>100.00 (12)</td>
<td>70.09 (107)</td>
<td>64.06 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>70.09 (107)</td>
<td>64.06 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
<td>27.07 (14)</td>
</tr>
<tr>
<td>F</td>
<td>70.19 (107)</td>
<td>64.78 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
<td>27.07 (14)</td>
</tr>
<tr>
<td>Total Sum</td>
<td>100.00 (12)</td>
<td>70.09 (107)</td>
<td>64.06 (97)</td>
<td>57.14 (15)</td>
<td>50.00 (5)</td>
<td>44.64 (14)</td>
<td>38.86 (14)</td>
<td>32.26 (14)</td>
</tr>
</tbody>
</table>

Notes: Sample size presented in parenthesis.
Appendix G: Frequency of variable inclusion for regression models

Appendix G-1: Bundesliga division one frequency of variable inclusion
Appendix G-2: Bundesliga division two frequency of variable inclusion
Appendix G-3: NBA frequency of variable inclusion