Dynamic prediction of Australian Rules football using real time performance statistics

Donald Forbes

BSc (Maths), LLB, Grad Dip (App. Sci)

A thesis submitted for
the degree of

Doctor of Philosophy

School of Applied Statistics

Department of Life and Social Sciences

Swinburne University of Technology

March 2006

Acknowledgements

The opportunity to undertake this research was a once in a lifetime opportunity for an avid sports fan such as myself. I am extremely grateful to the Department of Education, Science and Training for having the foresight to provide an Australian Postgraduate Industry Award (APAI) to support me financially during my tenure.

This funding was complemented by Champion Data who have sponsored this project through its entirety and realised the impact that this research could have in the football industry. Special thanks to Ted Hopkins, who has taken a special interest in my work and been a constant source of discussion and debate on football issues. Thanks also to the football department, particularly Glenn Luff, Greg Planner and Darren O'Shaughnessy who have provided me with data and definitions during my research. Without the support of Champion Data this project would not have been possible.

Special thanks to my supervisors Stephen Clarke and Denny Meyer. Stephen has been a constant throughout this research and his knowledge and skills have been invaluable. At times I thought I had confused him with the tack I took, yet he was always quick to read my work and provide me with suggestions on where I could improve my work. Denny was a new addition to the stable, only coming on board for the last 18 months of my tenure. She provided a fresh approach to my work as she had no knowledge of AFL football and was not tainted with pre-conceived ideas. Her technical skills were a huge bonus to me in the crucial stages of my analysis and she provided inspiration and encouragement that was greatly appreciated.

Finally, I am especially grateful to the love and support I received from my family. To my magnificent wife, Elizabeth, who shares little of the passion that I do for sport, thank you for putting up with countless hours of meaningless football matches. To my daughter, Ashley, it was nice to have someone to watch the footy with and generally you went for the right team. And finally to our latest addition, Oliver, your skill at giving me peace to finalise my thesis was unquestionable. You all provided me with inspiration and support during my research and helped me keep things in perspective. Without your patience and love, I doubt that I would have reached this stage.

Candidate's Statement

This document contains no material which has been accepted for the award to the candidate of any other degree or diploma, except where due reference is made in the text of the thesis. To the best of my knowledge, this document contains no material previously published or written by another person except where due reference is made in the text of the thesis. Unless acknowledged, all work found in this thesis has been done by the candidate.

Donald Forbes

Table of Contents

List of Figures

List of Tables

Abstract

This thesis contains detailed analysis of Australian Rules football, played in the Australian Football League (AFL). Data from 1295 matches, dating back to 1998, as collected by the League's official information providers, Champion Data, has been used for the analysis as they were the industry partner for the reserach. The quality and detail associated with the data has enabled analysis to be performed that previously would have been impossible.

Statistical distributions are fit to scoring events, both on attack and defence, for teams in the competition. It is discovered that the Poisson distribution provides a better approximation of the data than the negative binomial distribution for individual teams. Correlations between scoring events are also analysed with a view to developing a pre-match prediction model. Using the results of the exploratory analysis, a static pre-match model that performs better than a model updated at half time, is presented. This model uses negative binomial regression to predict goals and behinds separately for each team and consequently a predicted score. The failure of this model to adapt to dynamic events resulted in models being pursued that could adjust for events as they happened.

An eight state global Markov process model is presented that provides an adequate approximation to AFL football with no regard to location of events on the playing field. Transition probabilities are derived for each state using the transaction files collected by Champion Data for matches in 2003 and 2004. This model is then used for post-match applications, including altering play scenarios and calculating the effect of rule changes, as well as dynamically updating match predictions using live match data. It is expected that these applications will be made available to the wider football community over the next couple of seasons.

The coding of events by Champion Data according to their location on the field enabled a second model to be developed that calculated transition probabilities by zone. This 18

state zone model improved upon the global model due to the inclusion of more information. The zone approach will be more informative to AFL teams as it gives a clearer indication of the functionality of the attacking, midfield and defensive units, rather than looking on these units as a whole. The zone model was used to replicate the applications of the global model and investigate whether different results were produced.

Extra applications were made available with the introduction of the zone model, particularly investigating play strategy in different areas of the ground. Regression models were again developed for predicting match margin at different stages of a match using the transition probabilities up to that stage. The accuracy of these models was good with significant amounts of the variation in final margin explained and this accuracy increased noticeably as a match progressed. The models were used to test the differences in style of play for each team when compared to the competition average. Finally, playing styles of teams were compared for home state games and interstate games to test which transitions differed significantly.

The models presented in this thesis provided accurate approximations of AFL football that has not been seen elsewhere. Some of the applications of these models are already being used by AFL clubs and further commercialisation of the applications will take place over the next season with a view to providing detailed mathematical analysis to the AFL industry in years to come.

Chapter 1: Introduction

1.1 Background to research

The Australian Football League (AFL) is a religion to Australians, particularly in the heartland of Victoria. It is not uncommon for arguments to turn on who is the better player or which club the strongest. And, on any given weekend, thousands of youngsters nationally, can be seen kicking a Sherrin around in the hope that one day they might have what it takes to play at the elite level and grace the hallowed turf of some of the famous sporting arenas Australia has to offer.

The popularity of the AFL has not always been so, particularly for those people from the states of Queensland and New South Wales, of which I can count myself. These areas are traditional heartlands for rugby union and rugby league and the AFL has struggled to break into these markets with any great authority. Recent success for both the Brisbane Lions and Sydney Swans has enabled the locals to embrace the game and appreciate it for the spectacle it is, rather than viewing it as a game of aerial ping pong, not a patch on either rugby code.

Growing up in Sydney, I was lucky enough to be a member of the Sydney Cricket Ground, which played home to the Swans and through my youth I often made the trek to Paddington to watch the Swans do battle against a Victorian side. Around my neck would be a red and white scarf and over the shoulder the same coloured flag, both courtesy of my mother's handy craft skills. The only time the tribal colours weren't on display was when Hawthorn came to town and the red and white flag was replaced by a brown and gold one.

If it had been put to me back then that I would spend a large chunk of my life researching the data associated with Australian Rules football I wouldn't have believed it. Perhaps I could have accepted researching rugby league due to the fact I used to sneak over the back of the SCG to the Sydney Sports Ground, crawl through a hole in the fence and watch either Easts or Souths run around, which I found much more enjoyable. Over time the saturation of AFL football into the northern markets where I lived saw me become an avid follower and keen spectator at matches played in my area.

What has also been noticeable for me in the past few years is the changing face of sports information available during and after matches. Gone are the days where only basic information was relayed during a broadcast such as runs or tries scored. Champion Data (CD) is one of several companies involved in the collection of more detailed sporting data. They are responsible for the collection of match information in the AFL and they have changed the face of data provision to AFL clubs and the media. As the industry partner of the ARC grant that funded this project, CD's data was the only data available for analysis.

1.2 Aims of research

When the opportunity arose to research AFL football with a view to developing a dynamic prediction model I was in no doubt that this was the path I wanted my life to take. Before starting my tenure and in the early stages of it, I always looked upon this research as revolving mainly around gambling and making money from footy. It has only been the latter stages of my research where this driver has been replaced by a desire to better understand the scientific side of football as evidenced by the numbers.

This shift in my ideology was caused by my association with CD. After reading an American book (Lewis, 2003), I was made aware of the reluctance of Major League Baseball managers to accept the power of information that could be derived from performance statistics. In my limited dealings with AFL coaching staff, I was astounded at how embracing they were of the high level analysis I was producing. This inspired me to pursue analytical tools, the likes of which had not been seen in the AFL arena. Ted Hopkins, the managing director of CD was a constant source of ideas and encouragement for my research and his pushing of my research into the public arena helped me maintain my focus and concentrate on achieving the goals I had set for myself.

The main outcomes I hope to see from my research are the development of a number of tools that will be utilized in the football world for differing reasons. Firstly, I hope that my research will help coaches and technical staff make evidence based decisions during matches to enhance their chances of victory. Furthermore, they should be able to identify the strengths and weaknesses of their own charges and the opposition. Thirdly, it is hoped this research will be of great use to media outlets in terms of satisfying their viewers' desire for information in a manner that has never been done before. Fourthly, it is hoped that the decision making of players in game situations may be aided and improved through gaining an understanding of the implications of their decisions from a statistical point of view. Finally, it is hoped that this tool can be used in a live betting environment to accurately price match outcomes at any stage during the game. Unfortunately, Australian law does not allow for internet betting on live events but hopefully in time this rule will be relaxed as there is a licensed bookmaker interested in trialing the product. In the meantime the use of this tool will have to be restricted to the more limited market of phone betting.

1.3 Outline of research

This thesis is made up of 14 chapters. Chapter 1 gives an overview of the thesis as well as some background to the development of the model and how it came about. Chapter 2 provides the background to the literature that is relevant to this work and concentrates on AFL football, prediction models for sporting outcomes, approximating scoring rates and the use of Markov models in sport. This review establishes the framework for where this research sits in terms of previous analyses.

Chapter 3 gives a detailed account of the game of AFL football as well as the history of CD and the information they collect that has made this research possible. Chapter 4 takes the information collected by CD and uses it for some exploratory analysis of relationships within the data. Correlations between events are looked at to provide reasons for decisions made later in the thesis. Analysis of the data are continued in Chapter 5 where statistical distributions are fitted to scoring events in the AFL competition. Again, these results set the foundation for the application of the models seen later in the thesis. Chapter 6 revisits some of the established work on home ground advantage in the AFL competition as well as introducing some different concepts for home advantage relating to match statistics.

The introduction of a pre-match static prediction model is covered in Chapter 7 using a model based on the findings of Chapters 4 and 5. This model is used to highlight the fact that even with a dynamic update at breaks in the game the prediction accuracy is not improved. Therefore a different technique had to be implemented in order to come up with a model that reflected the dynamic nature of an in-game environment. This technique is introduced in Chapter 8 with an eight state Markov process model that uses match statistics collected by CD as its input to calculate transition probabilities. Chapters 9 and 10 present various applications of this model that can be used in both a dynamic and post match environment.

Champion Data's data collection allowed for location to be included in a zone model that is covered in Chapter 11 along with some revisiting of the earlier applications. Chapter 12 considers applications unique to the zone model that revolve around playing strategy and how to maximise a team's chances of victory. The penultimate chapter looks at the characteristics of teams and venues in the competition using the zone model, investigating their transition probabilities. Comparisons are made to the competition average and analysis is performed on interstate sides, comparing their home and away transition matrices. Teams and venues are also compared, to see whether any similarities in style of play exist. Finally, Chapter 14 summarises the findings of this research and possible applications. It also considers limitations of the model along with suggestions as to where this research can be taken in the future.

Chapter 2: Literature Review

This chapter provides an overview of the literature relevant to the research contained in this thesis. Initially, the introduction will set the scene for this research before an overview is given of the general literature relating to quantitative analysis in sport. Finally, research employing the use of Markov techniques in sport will be addressed. Throughout this chapter the literature relating specifically to Australian Rules football will be reviewed where relevant.

2.1 Introduction

Improved information collection of sporting data has provided the backbone for detailed analysis into player performance in a match environment. Gone are the days of recording basic statistics using pen and paper with the move made to more technologically evolved techniques. While suggestions for using computers for the collection of sporting statistics in real time have been around for some time (Patrick, 1985, Patrick, 1992, Croucher, 1992), the AFL has only seen the introduction of electronic statistics collection in recent years. As an evolving process, the level of detail available in match data has increased with time. It is this richness of information that has facilitated the research contained in this thesis, with the ability to construct chains of play and know when and where they occurred on the ground being crucial to the analysis.

A growing area of research involves the investigation of player roles in sporting events. This type of research has been possible due to the improved techniques used to collect match information and the quality of the data that is collected. Work was done investigating how players should be matched up in Australian Rules football games depending on how their opponents line-up (Tomecko, 1999). This work used quantitative and qualitative data to model the expected performance of players in specific positions; however, the model was developed using data derived from a country football league team. It would no doubt benefit from the information that is collected on the AFL competition by CD and used in the research undertaken in this thesis.

Another body of work relating to player performance was conducted in the sport of rugby union and derived a way of valuing the impact of a player's performance relative to their expected involvement in the match (Bracewell, 2002). Although individual player performance is not explicitly addressed in this thesis, the issue of players being hard to rate in a game due to their position has been a concern for CD and their commercialised player rankings. Often, forwards and defenders are under-rated due to the prevalence of the ball in the midfield. Preliminary investigation was done on a better rating of players along the lines of Bracewell's work using cluster analysis; however, this diverged from the focus of this thesis and was not pursued.

It was considered relevant to include these bodies of work as the analysis they contain is dependent on quality and detailed match information, just as this research is. The models presented in this thesis are only possible due to the richness of the AFL data that was available. It is expected that the data available for the AFL competition will enable detailed analysis to be undertaken in the future along the lines of the research of Bracewell (Bracewell, 2002) and Tomecko (Tomecko, 1999). The literature relating to modeling sport will now be looked at as this thesis focuses on approximating AFL football using Markov process models.

2.2 Quantitative analysis of sport

To best analyse sporting data from a prediction viewpoint, it is necessary to approximate match events by fitting a statistical distribution to the observed data. The majority of work in this area has related to scoring in soccer and debate has flourished for decades about which distribution has the most accurate fit. The majority of authors display a preference for the negative binomial distribution; however, as seen later in this review, there is some evidence to suggest that the Poisson distribution could be appropriate, particularly when predicting match outcomes. Definitions and descriptions of these distributions are given in Chapter 5.

Initial analysis in the 1950s and 1960s on English soccer results lent towards the negative binomial distribution as the best descriptor of goals scored by a team in a match. In early work in the area (Moroney, 1956), it was found that the Poisson distribution was not a statistically good fit to goals scored in a soccer match. The author expressed surprise that weather conditions and team-matching did not exert as great an effect as is often supposed. Twelve years later, work followed which also preferred the negative binomial distribution due to it being generated by random or chance mechanisms that underlined the conclusion that soccer is a game dominated by chance. It was suggested that due to this notion, a team who recognised this random element would be able to develop a successful style of play that harnessed the importance of chance on the game (Reep and Benjamin, 1968).

This idea was built on by the same authors (Reep, Pollard and Benjamin, 1971) when they successfully fitted the negative binomial distribution to the number of goals scored in a game of soccer. They displayed a clear preference for the negative binomial distribution over the Poisson distribution and this was reiterated years later (Pollard, 1985). In this work it was argued that the Poisson distribution could not apply to soccer matches, as the goal-scoring rate has to be the same for all games, but in reality, the rate of scoring varies from match to match indicating that the negative binomial distribution is more appropriate.

Later work provided an excellent overview of previous work in this area and built on it through further analysis (Baxter and Stevenson, 1988). A much bigger data set was analysed, with the conclusion that prior to 1970 the negative binomial distribution was preferable; however since then, the Poisson distribution seemed more than adequate. Five possible mechanisms that could be used to approximate soccer scoring were summarised with two being the simple negative binomial and Poisson distributions as presented in earlier work (Pollard, 1985). A third mechanism, previously suggested in earlier work (Cox, 1965) allowed two parameters, depending on the previous number of events in an interval. With appropriate assumptions this mechanism also leads to the negative binomial distribution. The other two mechanisms suggested are perhaps of the most interest as they appear more appropriate from a sporting point of view. Instead of the rate of occurrence being constant across time, it is allowed to vary over time. And secondly, a mixture of negative binomial distributions allows for differences in skill levels and abilities between teams (Baxter and Stevenson, 1988). A Poisson model does not allow for this, as the rate of occurrence is constant.

Complementing this literature is a body of research that suggests that, for the game of soccer at least, goal frequency increases as a match progresses. It was found that for the 1986 soccer World Cup, more goals were scored in the 15 minute period between the 60th and 75th minute than any other period (Jinshan, 1986). Conversely, for the 1990 World Cup, he found that scoring patterns increased over time, with the final 15 minute period containing the most goals (Jinshan, 1993). It was also found that scoring in the Dutch league increased monotonically with time, again using 15 minute intervals (Ridder, Cramer and Hopstaken, 1994). In the Scottish soccer league it was found that a higher than average frequency of goals occurred for the final 10 minutes of play (Reilly, 1996). From an Australian point of view, similar work was conducted on the National Soccer League between 1994 and 1998, which found that there was a significant increase in the number of goals scored in the second half, when compared to the first half (Abt, 2002). They also found that as the match progressed, so too did the frequency of goals scored using 15 minute and 5 minute time intervals.

Another, and a more recent approach to scoring in soccer distinguished itself from previous research by considering soccer scores from ‘a statistical point of view’ (Greenhough, Birch, Chapman and Rowlands, 2002). This work agreed with previous work (Reep, Pollard and Benjamin, 1971, Moroney, 1956) regarding the use of Poisson or negative binomial distributions in English soccer. However, they showed that neither the Poisson nor the negative binomial distribution describes the distribution of worldwide scores in soccer games. They show that extreme value distributions provide a better fit to this data.

The negative binomial distribution was applied to a number of facets of different sports (Pollard, Benjamin and Reep, 1977). Events looked at included passing chains in soccer (and goal scoring), points scored in gridiron, runs in a baseball half-inning, goals scored in ice-hockey, strokes per rally in tennis and runs scored per partnership in cricket. It was found that the negative binomial produced a good fit where there was an occurrence of infrequent events in a team environment e.g. soccer goals. However, when individual performances were looked at, such as in the tennis or cricket examples, there was not a close fit, indicating that individual skill was more significant than chance.

The majority of research into fitting distributions to scoring events relates to soccer. Given its worldwide popularity this is hardly a surprise. All the work in this thesis is concerned with Australian Rules, with markedly different scoring frequencies and systems, and therefore sits outside any previous work done on scoring patterns. Soccer is a game that has minimal frequency of scoring and the score can only advance by one unit. Australian Rules on the other hand, is a high scoring game and the score can advance by six points or one point. For these reasons alone it is considered that the analysis of scoring in the AFL contained in this thesis is worthwhile and unique.

Also, previous analysis in this area has only looked at a team's attacking return, i.e. the number of goals they score. This analysis investigates how teams concede goals, therefore investigating an area that hasn't been looked at previously when fitting distributions to scoring events. The fitting of distributions in this thesis provides the backbone for the implementation of a pre-match prediction model that uses negative binomial regression, which is presented in Chapter 7. Furthermore, the Markov process is reliant on a constant scoring rate, due to the expected number of transitions between scores, which is shown to be the case in the AFL.

Forecasting or predicting the outcome of sporting events is hardly a new area of research. Whether the ultimate aim was to exploit inefficiencies in betting markets or simply better understand the mathematical underpinnings of a sporting contest, numerous techniques have been used. One of the earliest forays in the area was the development of a least squares method for predicting American college football and basketball results (Stefani, 1977). These models used least squares to obtain ratings, and a margin of victory and winner was obtained from these ratings. The good or bad form of teams was accounted for by using a smoothing constant to adjust the ratings against what was predicted.

In 1980, Stefani improved his earlier models by including an adjustment for home advantage and applied his techniques to soccer (Stefani, 1980). The work of Stefani (Stefani, 1980) has underpinned research in predicting AFL football and this will be referred to when the literature relating to AFL football is addressed. Stefani's work paved the way for other academics to document their research into predicting the outcomes of sporting events, particularly in the case of American sports. Around the same time Harville used maximum likelihood estimates to obtain ratings for American pro football results and claimed his method was more accurate than the earlier work of Stefani (Harville, 1980). Home advantage was a necessary component of Harville's model and an autoregressive process that updates the ratings over time. The work of Leake in the 1970s on ranking college football teams (Leake, 1976) gave rise to an approach that accounted for least squares ratings being adversely affected by blowout games. As a result games with large score differences were down weighted so that their effect on the least squares ratings was not as pronounced (Stern, 1993).

Other techniques have been used to rate and subsequently predict sporting outcomes, particularly Poisson regression in soccer. In the early 1980s a simple Poisson regression model was fitted to data from English football (Maher, 1982). This technique was adopted in later years by other researchers. A model was developed to exploit inefficiencies in English soccer betting markets (Dixon and Coles, 1997). This model included a time-dependent effect as an indicator of form throughout the season and the introduction of this factor using a designated betting strategy provided a positive return. Around the same time a model was developed that utilized both offensive and defensive capabilities for teams in the English Premier League, to ascertain whether the team that rates the best statistically is the winner of the competition. Dixon extended his 1997 work to factor in the elapsed time in a match and the current score in order to predict match results as a function of time (Dixon and Robinson, 1998).

In most sporting competition matches there is dependence between the scoring ability of the competing teams. Although this correlation makes modeling a more difficult proposition, it has been shown that by using a bivariate Poisson distribution, model fit and accuracy improve (Karlis, 2003). Similarly, for the Australian rugby league competition, a bivariate negative binomial regression model was used to model scores by taking into account offensive and defensive capabilities (Lee, 1999). This model was then able to be used to determine whether the premiers from the year of analysis were worthy winners.

Prediction models in the AFL are limited. The earliest work is that of Stefani in collaboration with Clarke (Stefani and Clarke, 1992). Their work compared Stefani's least squares approach to Clarke's exponential smoothing model and found very little difference between the two in terms of predictive accuracy. The basis of the pre-match prediction model presented in this thesis is heavily dependent on the techniques used by Clarke in his exponentially smoothed approach to rating AFL teams (Clarke, 1993). More recently, Bailey has sought to expand on the work of Stefani and Clarke by including all historical match data for the AFL competition and using multivariate modeling to derive numerical estimates for travel, ground familiarization, team quality and current form (Bailey and Clarke, 2004).

The issue of home advantage in the AFL is well established and although revisited in this thesis, it is not analysed in great detail. The reason for this is the existing body of work on this topic (Clarke, 2005, Stefani and Clarke, 1992, Bailey and Clarke, 2004). It was felt that little could be gained out of further quantifying home advantage in terms of scoring. However, analysis is presented in Chapter 6 that looks at home advantage from the point of view of match performance statistics in a way that has not been done before for the AFL competition. This is further complemented by the analysis in Chapter 13 that compares the performance of teams at home and away to determine whether they differ significantly in their transition probabilities.

As already stated, the pre-match prediction model in Chapter 7 draws heavily on the work of Clarke by using exponentially smoothed ratings for teams in a match on attack and defence and allowing for home advantage. The derivation of winner and probability of victory is a unique approach that allows for the correlation between goals and behinds as well as either teams ability to score by using a negative binomial regression model. This model is used to predict the number of goals and behinds for each team. This approach has not been used for AFL and compares favourably to the established models in the literature.

The majority of work referred to in this section has been developed for pre-match prediction. The purpose of this research was to develop a dynamic model that utilised the wealth of match information available to update predictions during a match. In light of the discoveries that are presented in the early chapters of this thesis and the static nature of the model in Chapter 7, a dynamic model relying on Markov techniques was pursued and it is pertinent to address the sporting literature in this area.

2.3 Markov techniques used in sport

The use of Markov techniques in sporting situations is strongly grounded in the sport of baseball. One of the earliest pieces of research occurred in the 1960s and was used to obtain the expected number of runs for the remainder of the half-inning (Howard, 1960). This initial model comprised 25 states in the half-inning and was used in later years for further research into baseball events. Work in the 1970s used this model to calculate probabilities for no runs being scored as well as the expected number of runs scored in any state in the half-inning (Trueman, 1977). Around the same time work was done to obtain optimal batting orders using Monte Carlo simulation involving 200,000 baseball games; however a limited number of orders were explored (Freeze, 1974). The 25 state model was expanded to a 2,593 state model to try and better estimate the expected number of runs for a half-inning (Bellman, 1977).

More recently the 25 state model was used to optimize a batting line-up so as to maximize the expected number of runs for the half-inning (Bukiet, Harold and Palacious, 1997). This most recent research proposed a Markov chain model for baseball that found optimal batting orders, run distributions per half inning and per game and the expected number of games a team should win. This involved a 25 state model and therefore a 25 x 25 transition matrix for each player consisting of that player's probabilities of shifting the state of the game to any other during an appearance at the plate. These probabilities are dynamic in the sense that they can be adjusted as the season progresses and form strengthens or wanes.

The most recent work on baseball extended the twenty-five state model described above in a number of ways including a 1,945 state model for expected runs using non-identical players and a 1,434,673 state model to obtain the probability of victory from any state in the game (Hirotsu, 2002). He also addressed strategy issues such as optimal pinch-hitting and substitution for pitchers based on the handedness of the pitcher and player at bat.

In addressing the research associated with the use of Markov models in baseball, it is worthwhile noting that baseball is a discrete event sport and differs from continuous sports such as Australian Rules football. Therefore, analysis of continuous sports is more relevant to this thesis. The major body of work in the area is also by Hirotsu involving a four-state Markov process model (Hirotsu, 2002). He used English Premier League data to derive his transition probabilities via Poisson regression. The model was then used to evaluate the expected number of goals in a match as well as the expected number of league points a team was likely to obtain. It was also useful for investigating strategy issues such as when to substitute or commit a deliberate foul in order to increase the chances of victory.

A model has been developed to analyse strategy decisions in the continuous game of ice-hockey (Thomas, 2006). The author uses a state-space model dependent on possession of the puck and location on the rink to determine expected number of goals scored. Analysis is performed on accepted strategies in ice-hockey to investigate these styles of play and whether they are effective for scoring goals. The data used for this analysis suggested that a continuous time Markov process was not appropriate and therefore, the model used was described as a semi-Markov process.

The only use of stochastic processes to model AFL football is the work done by Clarke and Norman (Clarke and Norman, 1998), who investigated the decision process of when to rush a behind in an AFL game. They looked at when a team's chances of victory could be improved by conceding a point to the opposition. Their model did not utilise actual data, instead the authors chose to assume transition probabilities based on their knowledge of the game. Obviously, any model would be more accurate with the inclusion of transition probabilities derived from observed data. A summary of papers was compiled by Norman (Norman, 1999) in which he looked at 17 papers concerned with ways to utilise stochastic processes for modeling sport. Not all of these models used Markov techniques; however, the paper gives a good background to work done in the area.

Hirotsu's soccer model was the inspiration behind the initial eight state model presented in this thesis (Hirotsu, 2002). It was strongly believed that AFL would be very well suited to a Markov process model and this is been shown to be true in later chapters of this thesis. Although Hirotsu (Hirotsu, 2002) developed a comprehensive model for soccer he did not investigate updating his transition probabilities during a match. This is an integral part of the models in this thesis with the ability for live match statistics to be used to improve the accuracy of predictions as events unfold. This feature is unique in the literature for the use of Markov models. Furthermore, the abundance of applications that these models bring to the game of Australian Rules football is completely unique and revolutionary. The only paper that could be classed as close to the research contained herein is the work of Clarke and Norman (only because it relates to AFL football), however the research presented in this thesis is novel.

Chapter 3: Australian Rules football – the game and the information

3.1 History

The game of Australian Rules dates back to the 1850s where it began when one of its co-founders returned from schooling in England and introduced a hybrid rugby game as a way to keep cricketers fit during the winter off-season. The first recorded game of the new codes was played during 1858 between Scotch College and Melbourne Grammar School. In the same year, the first Australian Rules club was formed, being the Melbourne Football Club, who are still active in the game in its present form.

In 1896 The Victorian Football League (VFL) was established and the following year the League's first games were played among the foundation clubs – Carlton, Collingwood, Essendon, Fitzroy, Geelong, Melbourne, St Kilda and South Melbourne (Sydney). By 1925, the league had welcomed four other clubs, Richmond, Footscray (Western Bulldogs), Hawthorn and North Melbourne (Kangaroos) and continued as a 12 team, Melbourne based, competition until 1987 when it went national by including a team from Perth and a team from Queensland.

The competition evolved into its present state as the AFL by 1997 and is almost a truly national entity with only one state, Tasmania, not enjoying the identity of a local team. It now consists of 16 clubs after Adelaide (in 1991), Fremantle (in 1995), and Port Adelaide (in 1997) joined the AFL and foundation club, Fitzroy, merged with the Brisbane Bears to form the Brisbane Lions after the 1996 season. A chronological history of the game can be found in the official handbook of the AFL, which is published at the start of each season (Lovett, 2004). AFL clubs can be referred to by a number of names and to remove uncertainty, Appendix 1 contains the names that clubs may be referred to in this thesis.

Since becoming a national competition, the game has developed into the major winter football code in the southern states of Australia both for spectators and participants alike.

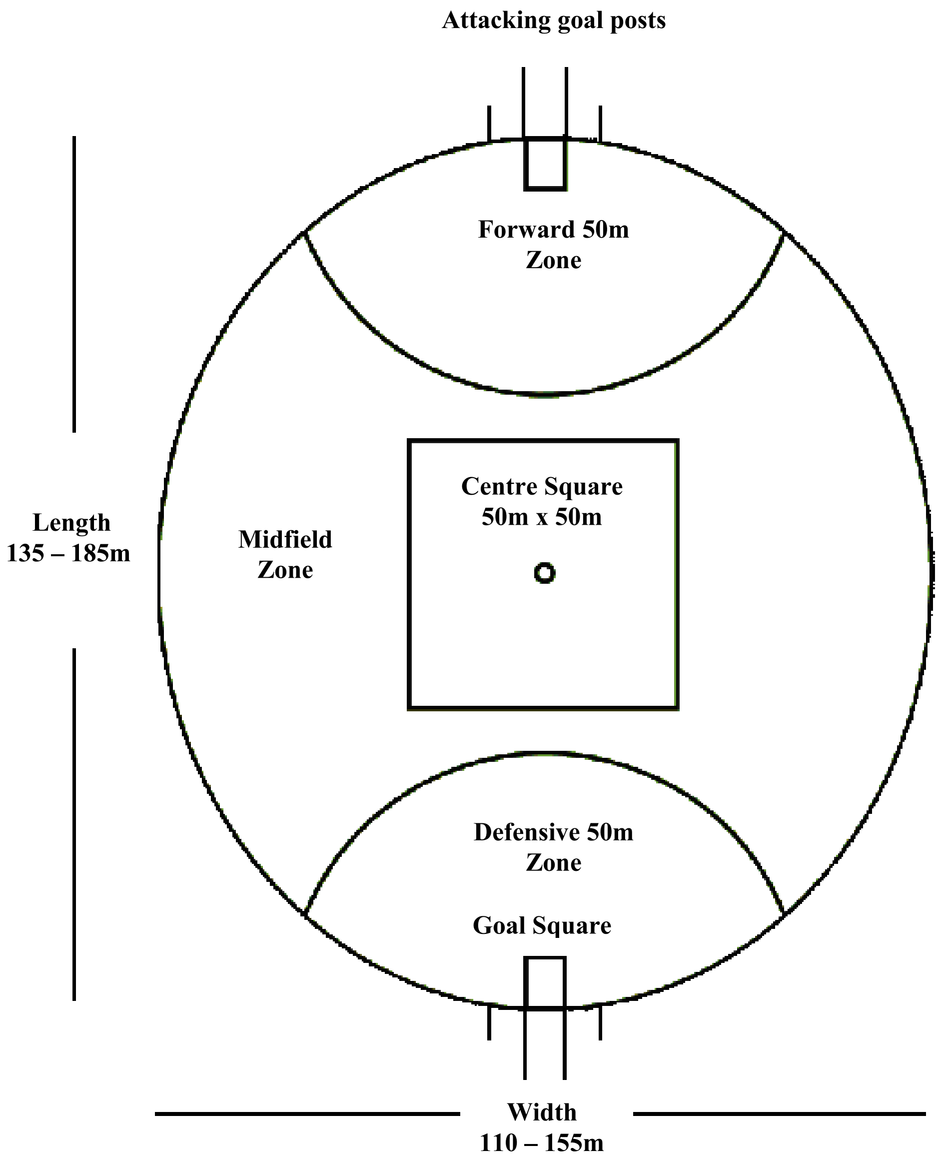
It also enjoys huge popularity in the Northern Territory and Australian Capital Territory. In the states of Queensland and New South Wales, while it runs slightly behind Rugby League and Rugby Union in terms of popularity, it is still widely followed.

The AFL is currently enjoying unprecedented exposure and interest Australia wide with total and average attendances increasing by 50% since the competition became the AFL in 1990 (Lovett, 2004). The average attendance at an AFL match in the 2003 season was 34,333 (Lovett, 2004). The average attendance figures compare very favourably with the major soccer competitions of the world which attract average crowds of 34,000 (England), 25,700 (Spain), 25,200 (Italy) (FAPL, 2002). In 2002, 2.5 million people attended at least one AFL match, making it the highest attended sport in Australia. It is also a sport that is popular with either gender, with 21% of males and 13% of females attending at least one game during the season (Australian Bureau of Statistics, 2003). This is in part due to the national nature of the competition and the relative equality of the participating teams.

3.2 The game

Since first played in the late 1800s, the game of Australian Rules has evolved with a number of rules being changed and introduced, however the overriding tenet of the game has remain unchanged. A complete history of the rules and their changes is contained in the official handbook of the competition including the year they were introduced or abolished (Lovett, 2004). In its present state, 16 clubs compete against each other during the home and away season over 22 weeks before the top eight sides play in the finals series over four weeks to determine the premiership winner. Each game is played over four 20 minute quarters, which in reality last almost 30 minutes due to the clock being stopped for play interruptions such as the ball leaving the playing arena. At the end of each quarter the teams swap direction. Each team consists of a squad of 22 players, however only 18 may be on the field at any one time, with the remaining four players able to be interchanged onto the ground at any time with no restriction on the number of interchanges that can be made. The game is played on grounds that are oval in shape of varying dimensions. At each end of the ground is a semi-circle that signifies 50 metres to the goal posts. Inside the arc of this circle is known as the forward zone for the attacking team and defensive zone for the opposition. The midfield zone lies between the two 50m arcs. Figure 3.1 displays the major features of an Australian Rules football ground.

Figure 3.1: AFL playing field and main features



A match is started by an umpire bouncing the ball in the middle of the centre square where a player from each side jumps for the ball to try and knock it to their team's advantage. Once a team has gained possession of the ball the idea is to advance the ball towards the attacking goal posts and register a score. Territory can be gained by the player running with the ball, provided they bounce or touch it to the ground every 15 metres, or moving it on to a team mate or open space. This is done by what is known as a disposal and constitutes either a hand pass or kick. Players in possession of the ball can be tackled by their opponents and the player in possession must endeavour to dispose of the ball when this happens. Free kicks are awarded when rules are infringed and result in a player being allowed to dispose of the ball without interference from the opposition. If a player catches a kick which has travelled at least 15m, before it touches the ground or another player he is awarded a mark, which has the same effect as a free kick in that he can dispose of the ball without interference. If an umpire deems the ball to be dead, a ball up is called whereby the umpire will bounce the ball and players attack it in a manner similar to a centre bounce. If the ball leaves the playing area it is referred to as 'out of bounds' and is returned into play by an umpire tossing it backwards over his head where players duel for it as for a centre bounce or ball up.

At each end of the ground is a set of four upright posts, which the attacking team uses to score. Scoring can be done by the addition of either six points or one point. A six point score is known as a goal and occurs when a player kicks the ball between the two centre posts. The ball is then returned to the centre of the ground for a centre bounce. If the ball is touched or passes between the two outside posts, a behind worth one point is scored. If the opposition kick the ball or punch it between the two outside posts, the attacking side registers a rushed behind. After the scoring of a behind, the opposition kicks the ball back into play from the goal square via what is known as a kick-in. Using the data obtained from CD, it has been found that the average score in an AFL match is 14 goals, 12 behinds, resulting in 96 points, with draws occurring very rarely. In fact for the 6 years from 1998 to 2003 only nine matches have ended with teams on the same score.

3.3 Development of Champion Data and AFL information collection

As a result of the enormous popularity of the game, many fans desire information about the performance of players and clubs alike. Whether it is via the AFL website or through the print and electronic media, people want statistical information to a level that has not been seen before. This need is now satisfied by CD, who has been professionally collecting AFL match and player statistics for individual clubs since the start of the 1996 season. Since 1998 they have been the collector and provider of the official AFL statistics.

While suggestions for using computers for the collection of sporting statistics in real time have been around since 1985, (Patrick, 1985, Patrick, 1992, Croucher, 1992) such methods were not adopted early in the AFL. Prior to 1996, statistical information on the AFL was collected by APB Sports (many of whose employees now work for CD). Summary statistics were collected and collated at the ground using a pen and paper. However, the level of detail was nowhere near as extensive as fans now expect. For each player statistics consisted of kicks, marks and handballs for each quarter. In addition there were team totals for free kicks for and against, goals and behinds scored, and hit outs and tackles. These statistics were generally published in the print media the Monday following the game. Since they were generally not available until after the match, individual player statistics other than goals scored were rarely referred to in live broadcasts. Statistics were something fans might cast an interested eye over a few days after the match. They were not part of the real time discussion as to how the match was being played, where it was being won or lost, and who were the best players.

The principal of CD is Ted Hopkins and whilst he only played a handful of VFL games, he is well known for the role he played as 19th man for Carlton in the 1970 Grand Final. Sent on at half time with Collingwood a seemingly unassailable 44 points in front, Hopkins kicked four goals from a pocket and was instrumental in an amazing Carlton victory (Devaney, 2002). The match is often cited as a turning point in Australian Rules football. With coach Barassi issuing half time instructions to his players to ‘handball,

handball, handball' it represents the beginning of the more continuous 'play on at all costs' style now common. By 1996 Hopkins was an independent journalist and principal of a multi media publishing company, who had decided to branch out into collecting football statistics.

Hopkins recognized the need for collaboration with established mathematicians and statisticians to increase the potential of the product. As a result, a relationship between Swinburne University and Hopkins began in 1997 when Hopkins began publishing Swinburne Computer predictions on AFL football in The Herald/Sun (Hopkins, 1996) and The Australian Financial Review (Hopkins, 1998b, Wright, 1996). In addition to publishing the computer tips, Hopkins also wrote many articles on other aspects of Swinburne Sports Statistics teaching and research (Hopkins, 1998a). A natural association arose when he developed CD to collect sporting statistics.

Champion Data revolutionised the collection of statistics in AFL football, by the innovative introduction of quality. Right from the beginning Hopkins was not interested in recording what he described as 'Rubbish Statistics', those periods of play where several players may make ineffectual contact with the ball before it is cleared from a pack. The cottage industry that collected kicks, marks and handballs for the following day's newspaper, was transformed into a business providing nearly 100 statistics and match summaries immediately to coaches, the live broadcasters, the public and the media via the internet. While initially there was resistance in some media to 'the boffins from Swinburne', the innovation has changed football broadcasting and reporting forever, with terms such as effective handballs, contested marks, clangers, now part of the language and culture of football. CD has lifted the profile and value of statistics and analysis in the sports media.

Their contribution to the improved collection of AFL statistics in the modern age also lies in their approach to the game and the technology they employ. Originally, a caller and a keyboarder at the ground entered the data directly into a PC via a modified keyboard with one button for each player and one for each statistic. CD has now moved to a computer driven system that involves up to two callers at the ground who describe the occurrences on the field (including interchange). This information is relayed to a keyboarder and back caller at an off site location where it is stored on a computer server. The company has always been in the forefront of modern methods of communication, and before embracing the Internet used ftp to transfer information to clubs. The use of computers and modern methods of data storage and retrieval, has allowed the data base to be analyzed and 'value added' statistics such as player rankings to be developed.

The use of the statistics went much deeper than just measuring player performance. CD was always interested in how the statistics shed light on underlying tactics and strategy in football. What contributed to winning performances? What differentiated premiership or finals contenders from the also-rans? The early years were funded in part by AFL club subscription payments for not only the collected statistics but also technical analysis of games and possible opponents based on past statistics. Hopkins is still the principal writer for CD and often used the statistics as a basis for his articles (Hopkins, 1997, Hopkins, 1998a, Hopkins, 1998b, Hopkins, 1996)

The immediacy and range of statistical data available has led to an increase in the expectations of football followers. Supporters and teams alike have a continual need for match information most of which can and is satisfied by CD's data. The areas where CD's information is now used are numerous and below is a snapshot of the information CD supplies:

- 15 of the 16 AFL clubs receive data from CD relating to past performance, both in the latest game and the season as a whole. They are also given a profile of the relative strengths and weaknesses of their next opponent. The clubs are able to interrogate the databases to extract information as they like or can request certain information from CD's football department.
- CD is party to a contract to be the official information provider for the AFL. Their data are used by the AFL on their website (www.afl.com.au) for the various season statistics they provide (http://afl.com.au/default.asp?pg=stats).

When games are in progress, the live scores and stats on the AFL website are provided by CD. This contract also requires CD to call and collate the same information for the national under 18 and under 16 championships with the data being made available to the relevant recruiting managers from each club. They also do the International Rules matches that are played in Australia.

- CD's information is used by all three of the AFL television broadcasters. Channel 10's football coverage utilizes CD's player rankings. This is a formula that ranks players in a game according to their statistics with players able to both gain and lose points depending on what they do. Channel Nine's football coverage used CD's goal kicking probability model during its match broadcasts. This model was developed thanks largely to the historical recording of angle and distance for each set shot in a match. Fox Footy relies on CD's live information for its match day coverage.
- The Herald Sun carries much of CD's data in its tabloid newspaper. A detailed synopsis of the player statistics is included in Monday's paper. They also publish the rankings points for each round on a Wednesday and include the season totals as well as including a preview of upcoming games with comment on where clubs are doing well and doing poorly according to their season statistics. Various other newspapers around the country, such as The Age in Victoria and The West Australian, carry CD's statistics in less detail than The Herald Sun.
- A day after each round, all match footage and associated statistics are available on digital media for the benefit of clubs. They can use this quickly and easily to compile player footage for post-analysis of completed games or pre-match analysis of upcoming opponents. This is a simpler and easier process than the drawn out procedure of watching VHS tapes and editing and cutting them accordingly.

While the above are the commercial uses of the data, the completeness and accuracy also make it a valuable resource for academic study. For example, the author has used the transaction files to calculate transition probabilities for a Markov Chain model of

Australian rules football (Forbes and Clarke, 2004). It is hoped this model will be used for real time prediction of results and will assist with tactical decisions as described later in this thesis.

The exposure that CD's information receives means it has to be extremely accurate and of a very high standard. The techniques and strict training that CD employs ensure this. The live feed of footage into the off site location is watched by a back caller, allowing for errors and missed statistics to be included in the final product by being edited in at quarter time or fulltime. There is also quality control after the event with the football department from CD regularly auditing games to validate their accuracy and remaining in close contact with the clubs to ensure they are getting accurate records of match day events.

3.4 The data

The level of detail at which CD collects AFL information is extensive. Not only do they collect nearly 100 statistics relating to the events of a game but an impression of the quality and effectiveness of the possessions and disposals at a team and player level can be extracted. For instance a kick is not simply recorded as such but could fall under the category of long, short, ground, ineffective or clanger. A significant advantage with the way the data are collated and stored is the assigning of every statistic to a team and indeed a player. This means that summary totals for teams and players can be extracted for any time frame, be it a quarter, a match, a season or a career. For instance, one may be interested in calculating the number of free kicks a player has given away in his career and this is very easy to achieve as every player has a unique identifier in the database that they carry for their playing career.

The match data are recorded in Microsoft Access and Oracle databases that contain every game from a season. Within the database are various tables that can be linked to each other in order to extract meaningful information. The advantage of the transaction file is that it enables complete sequences of play to be extracted and analysed - for instance the chain of events leading up to a goal. This is a huge progression forward from previous data stores that recorded only the summary statistics for a match. Previously, there was no way to tell when or where a player gained his kicks. The data recorded and stored by CD defines when and where the statistics occur, as well as by whom, and what occurs before and after a particular event.

Champion Data takes a rigorous approach to recording the context of a game by collecting information such as the attendance and venue of the match. Other information is logged according to recognized 'business rules' such as the home team, whether there has been interstate travel, who won the toss and what end they chose to kick to. Over 20 different variables help to set the scene for a match that could prove important in trying to scientifically analyze the game. These also include weather conditions for every match ranging from the nature of the surface, to wind strength and direction and air temperature.

Clearly, the level of detail collected by CD about each match is very comprehensive. The level of detail in the match performance statistics is just as comprehensive, with nearly 100 different statistics recorded during a match. The match statistics recorded range from an umpire's report of a player through to a player's disposal or possession gather. When it comes to the possession and disposal statistics for players within a game, these are rated according to set definitions developed by CD.

As well as the statistics recorded by the caller, CD has set up their information systems to generate a number of statistics after the event. These are referred to as derived statistics. The system has various triggers that will recognize when a derived statistic is to be included. For instance, when a player kicks a ball to a teammate who subsequently shoots at goal and scores, the initial player would be credited after the event with a goal assist. The advantage of this is the human element is taken out of the process, as these types of statistics are system generated.

In addition to the information on what happens during the game, there are also a number of indicators for some special events. For instance, for every free kick, the reason it was paid is given; as well as the source of a shot on goal; type of shot on goal; type of miss for a shot on goal; and the direction that the defending team kicks in after a behind. For every set shot on goal the angle on goal and distance from goal is recorded. This information is recorded by a caller at the ground according to the co-ordinates of the shot in relation to the goalposts. This kind of information has never been collected or available before and is important to being able to thoroughly understand and comment knowledgeably on the game. For example, the angle and distance of set shots has been used to develop the goal kicking probability model that is used by Channel Nine during their match coverage. The model predicts the chances of a goal being scored from the set shot, and depends on the kicker's past performance as well as the difficulty of the shot.

3.5 Examples

To gain a better understanding of CD's systems and processes, two examples are included. The first example is a snapshot of a match from the 2003 season. The extract from the transaction file shows how the information can be grouped together to turn it into something that is meaningful from an analysis point of view. Included before the transaction file is a transcript of the call from the ground relayed back to the off site location, which gives an idea of how much is extracted from what appears to be very little.

The match is the round 10 game between Port Adelaide and Collingwood, played at Football Park on the 30th May 2003 in front of 43,321 spectators. The weather conditions were cold and fine with a light wind and the surface was hard. The toss was won by Collingwood who chose to kick to the Northern end.

Here is the actual call that is relayed back to the off site location and transformed into the data presented below:

- Umpire James Bounces;
- Wanganeen Hard ball, Handball;
- Stevens Receive, Kick Long;
- Cockatoo-Collins Hard ball, Handball;
- Tredrea Receive, Disspossessed by Buckley;
- Shaw Loose ball, Handball;
- Burgoyne Free against Holding the man Umpire James;
- Advantage Licuria, Handball;
- Johnson Receive, Handball;
- Lokan Receive, Handball;
- Johnson Receive, Handball;
- Shane Wakelin Receive, Kick Long;
- Rocca Dropped Mark;
- Darryl Wakelin Loose Ball get;
- Brogan Block;
- Darryl Wakelin Kick Long, Inside, Out of Bounds.

It is clearly evident from the inclusion of the verbal call that CD's systems and processes add a lot of value to the simple match call. As the collection and provision is all done in real time, it is important that the call is as abbreviated and succinct as possible. The extract above shows that this is the case, however the amount of information that the system produces is crucial to the final product. This extract accounts for only the first 45 seconds of the game but already there are 42 different occurrences in the match that have been recorded beginning with the players who started on the bench and finishing with the ball out of bounds in Port Adelaide's forward 50m. 15 players have been involved in the match within the first 45 seconds and there has been 21 different statistics recorded, highlighting the level of detail that CD record and collect from a game of AFL.

Table 3.1: Extract from transaction file for Port Adelaide v Collingwood

Quarter	Time (secs)	Transaction Type	Zone	Club	Player
1	0	Interchange Off	Midfield	Collingwood	Alan Didak
1	0	Interchange Off	Midfield	Collingwood	Brodie Holland
1	0	Interchange Off	Midfield	Collingwood	Steven McKee
1	0	Interchange Off	Midfield	Collingwood	Richard Cole
1	0	Interchange Off	Midfield	Port Adelaide	Jarrad Schofield
1	0	Interchange Off	Midfield	Port Adelaide	Stuart Cochrane
1	0	Interchange Off	Midfield	Port Adelaide	Jared Poulton
1	0	Interchange Off	Midfield	Port Adelaide	Brent Guerra
1	0	Start Quarter	Midfield	UMPIRE	Umpire James
1	0	Centre Bounce	Midfield	UMPIRE	Umpire James
1	8	Hard ball get - in play	Midfield	Port Adelaide	Gavin Wanganeen
1	8	CB First Possession	Midfield	Port Adelaide	Gavin Wanganeen
1	9	Effective Handball	Midfield	Port Adelaide	Gavin Wanganeen
1	9	Centre Bounce Clearance	Midfield	Port Adelaide	Gavin Wanganeen
1	11	Handball Received	Midfield	Port Adelaide	Nick Stevens
1	11	Long Kick	Midfield	Port Adelaide	Nick Stevens
1	15	Hard ball get - in play	Midfield	Port Adelaide	Che Cockatoo-Collins
1	15	Effective Handball	Midfield	Port Adelaide	Che Cockatoo-Collins
1	17	Handball Received	Midfield	Port Adelaide	Warren Tredrea
1	18	Dispossessed	Midfield	Port Adelaide	Warren Tredrea
1	18	Dispossesses	Midfield	Collingwood	Nathan Buckley
1	18	Loose Ball Get	Midfield	Collingwood	Rhyce Shaw
1	21	Ineffective Handball	Midfield	Collingwood	Rhyce Shaw
1	21	Free Kick Against	Midfield	Port Adelaide	Shaun Burgoyne
1	21	Free Kick For	Midfield	Collingwood	James Clement
1	24	Free kick - advantage	Midfield	Collingwood	Paul Licuria
1	25	Effective Handball	Midfield	Collingwood	Paul Licuria
1	26	Handball Received	Midfield	Collingwood	Ben Johnson
1	26	Effective Handball	Midfield	Collingwood	Ben Johnson
1	28	Handball Received	Midfield	Collingwood	Matthew Lokan
1	29	Effective Handball	Midfield	Collingwood	Matthew Lokan
1	30	Handball Received	Midfield	Collingwood	Ben Johnson
1	30	Effective Handball	Midfield	Collingwood	Ben Johnson
1	32	Handball Received	Midfield	Collingwood	Shane Wakelin
1	32	Long Kick	Midfield	Collingwood	Shane Wakelin
1	32	Long Kick to advantage	Midfield	Collingwood	Shane Wakelin
1	37	Mark - Dropped	Midfield	Collingwood	Anthony Rocca
1	39	Loose Ball Get	Midfield	Port Adelaide	Darryl Wakelin
1	41	Block	Midfield	Port Adelaide	Dean Brogan
1	42	Long Kick	Midfield	Port Adelaide	Darryl Wakelin
1	42	Inside 50m	Midfield	Port Adelaide	Darryl Wakelin
1	44	Out of Bounds	Attacking	Port Adelaide	UMPIRE

The following is the extract of match statistics from the AFL website for the 2003 AFL Grand Final, played between the Brisbane Lions and Collingwood Magpies at the Melbourne Cricket Ground on Saturday 24th September. This information is archived for two years on the AFL website.

(http://afl.com.au/default.asp?pg=2003round&spg=results&m_tournamentmatch_id=1143)

BRISBANE LIONS: q1: 5.5, q2: 11.7, q3: 14.12, q4: 20.14 (134)
GOALS: Akermanis 5, Lynch 4, Caracella, Brown 2, McRae, Pike, Hadley, Black, Leppitsch, Hart, Bradshaw 1.

COLLINGWOOD: q1: 3.3, q2: 4.7, q3: 9.7, q4: 12.12 (84)
GOALS: Didak 3, Burns 2, Davis, Woewodin, Buckley, Licuria, Tarrant, Fraser, Scotland.

INJURIES: None.

CHANGES: None.

REPORTS: None.

UMPIRES: McBurney, Kennedy, Allen

CROWD: 79,451 at the MCG

Tables 3.2 and 3.3 contain the individual player statistics for the match. The letters stand for the following match occurrences:

K – Kick; H – Handball; P – Possession = Kick + Handball; M – Mark; HO – Hit Out; T – Tackle; FF – Free Kick For; FA – Free Against; G – Goal; B – Behind

Table 3.2: Brisbane Lions player statistics, 2003 Grand Final

Player	K	H	P	M	HO	T	FF	FA	G	B
Jason Akermanis	18	2	20	3	0	3	1	0	5	2
Marcus Ashcroft	1	3	4	2	0	0	0	0	0	0
Simon Black	16	23	39	2	0	9	2	0	1	0
Daniel Bradshaw	11	4	15	7	1	2	0	0	1	0
Jonathan Brown	9	6	15	8	0	2	0	2	2	0
Blake Caracella	9	7	16	3	0	1	2	1	2	0
Jamie Charman	2	2	4	2	24	1	0	2	0	0
Robert Copeland	7	4	11	4	0	1	0	2	0	0
Richard Hadley	5	3	8	2	0	3	0	0	1	1
Shaun Hart	16	5	21	4	0	5	1	1	1	0
Chris Johnson	12	3	15	4	0	3	0	1	0	0
Clark Keating	3	5	8	1	27	1	1	3	0	1
Nigel Lappin	13	6	19	7	0	0	1	0	0	0
Justin Leppitsch	6	5	11	3	0	3	1	0	1	0
Alastair Lynch	10	2	12	8	0	0	0	1	4	2
Ashley McGrath	5	3	8	3	0	2	1	1	0	1
Craig McRae	9	5	14	1	0	2	1	1	1	1
Malcolm Michael	4	3	7	1	1	0	0	0	0	0
Martin Pike	10	6	16	4	0	0	0	2	1	1
Luke Power	8	8	16	6	0	4	0	1	0	1
Michael Voss	9	9	18	5	0	2	1	0	0	2
Darryl White	8	6	14	5	2	3	1	0	0	0

Table 3.3: Collingwood Magpies player statistics, 2003 Grand Final

Player	K	H	P	M	HO	T	FF	FA	G	B
Nathan Buckley	17	7	24	1	0	3	2	1	1	1
Scott Burns	8	14	22	4	0	5	1	0	2	0
James Clement	6	2	8	4	0	1	0	1	0	0
Jason Cloke	1	3	4	1	3	0	1	1	0	0
Richard Cole	4	5	9	2	0	4	0	1	0	0
Leon Davis	6	4	10	2	0	1	0	0	1	1
Alan Didak	9	4	13	5	0	1	2	0	3	1
Josh Fraser	15	8	23	7	21	0	5	2	1	0
Brodie Holland	4	4	8	0	0	2	0	1	0	1
Ben Johnson	12	7	19	2	0	5	1	0	0	0
Ben Kinnear	1	3	4	2	1	4	0	0	0	0
Paul Licuria	14	7	21	3	0	7	1	3	1	1
Matthew Lokan	2	0	2	1	0	0	0	2	0	0
Ryan Lonie	2	2	4	0	0	4	0	0	0	0
Shane O'Bree	9	4	13	0	0	2	0	0	0	0
Simon Prestigiacomo	6	3	9	4	0	2	0	0	0	0
Heath Scotland	8	6	14	2	0	2	0	0	1	0
Rhyce Shaw	5	3	8	3	0	0	2	1	0	0
Chris Tarrant	9	3	12	6	1	0	0	0	1	2
Shane Wakelin	7	1	8	2	0	1	1	0	0	0
Tristen Walker	1	0	1	0	5	1	0	0	0	0
Shane Woewodin	9	4	13	1	0	3	2	0	1	1

This snapshot of a match after it has been completed demonstrates the wealth of information that can be extracted from the call of an AFL match by CD. The information is also provided in real time to the website as well as to the coaches and media outlets. To give an idea of the detail recorded by CD as well as the systems and processes needed to collate the amount of information into meaningful output, some figures are presented for this match. In all, there was just over two hours of game time, with all of the four quarters running for over thirty minutes each. During this time CD recorded 2301 individual occurrences within the match at an average of 575 per quarter. This translates to a recordable event every 3.3 seconds.

After each round during the season the file for each game is added to the yearly database, forming a comprehensive record of the season elapsed. By the end of the season the information contained in the transaction files is extremely large and comprehensive. With 185 matches in an entire season and an average of nearly 2200 recordable occurrences per match, the transaction file for a season contains over 400,000 records. In addition, every statistic is attributed to one of the players or umpires involved in the competition. In one week of football alone, there are 22 individual players making up each of the 16 different teams as well as 3 different umpires for each of the 8 games in the round.

3.6 Summary

A summary is given of the major features of CD's information that sets it aside from previous AFL information collection:

- Range of statistics – information available previously was extremely limited with a minimal range of statistics recorded.
- Quality coded – never before has the quality of the possessions or disposals been recorded until CD's entry into the market.
- Zone coded – the location statistics occur on the ground is recorded according to whether it is attacking, defensive or in the midfield. This information can give an

idea of a player's versatility within a team and a rough idea of the type of position the player is assuming.

- Time coded – the time of the occurrence within the quarter is generated automatically when the statistic is keyed in. This information opens up the opportunity to investigate time dependent analysis of what happens on the field.
- Recorded in transaction files – data are stored chronologically according to when it takes place. This allows for entire sequences of play to be extracted instead of purely summary statistics.
- Derived statistics – statistics such as clearances and assists are automatically generated depending on subsequent statistics, removing the possibility of these extra statistics being omitted from the call.
- Match information at a macro level – information is recorded relating to the environment of the game such as the venue, weather, crowd, toss winner etc.
- Stats collected and available in real time – CD's process allows for the information to be available as it happens making it accessible for television and Internet broadcasts as well as directly to the coaches during the game.

The analysis of such a comprehensive set of statistics, along with the ability to investigate sequences of plays, allows insights into the game. For example CD recently produced the statistics of the 1970 Grand final from a video replay, and found some of the myths about that game did not stand up to scrutiny. In the second half of the match, Carlton was attributed with four handballs in their defensive zone and every one of these resulted in a turnover to the opposition. As the coach had instructed them to 'handball, handball, handball', the count of only four handballs in defence does not indicate that the players adopted the approach.

CD's collection of information is an ever-evolving system with the process and provision of information being reviewed at the end of every season to see how it can be enhanced and improved. Through CD's close association with the AFL and the clubs, this process allows for different statistics that were previously not measured, to be included in the match day analysis.

In today's football environment where teams, supporters and the media alike require comprehensive information relating to the game of AFL, it is easy to understand why the provision of detailed information has risen. It wasn't long ago that the only stats recorded were basic kicks, marks and handballs but one only has to look at the standard CD has set for the collection and provision of information to see this is a thing of the past.

Chapter 4: Exploratory analysis of the data

4.1 Introduction

The last chapter illustrated that the data recorded by CD for AFL matches are very comprehensive and detailed. Six seasons of data made up of 185 games per season were available for analysis. Such a large data set needs some preliminary analysis. This analysis paved the way for the next chapter on AFL scoring events and set the framework for the models that have been developed. The first issue addressed is whether clubs score at different rates from one another. This would require analysis at a club level instead of at the competition level. Secondly, correlations will be investigated for team scoring events within the data set to establish whether there is independence between events or otherwise. This will be followed by analyses on scoring as a function of time.

4.2 A global competition or individual team approach?

Intuitively it is to be expected that in any regulated sporting competition, no matter how even it might seem, the teams competing will score and concede at different rates to each other. As this section is investigating correlations and relationships within the competition, it is believed that the analysis has more power with team effect removed. A chi-square test was performed on the observed number of goals each of the 16 teams kicked for the six-year period, 1998 to 2003, to ascertain whether there is a difference in score returns. This test provided extremely strong evidence to suggest that teams in the AFL score at different rates (chi-square stat = 269.2, 15 d.f., p-val < 0.00001). Although this was the expected result, the magnitude of the evidence suggests that there is no need to investigate relationships for the competition and instead relationships for the individual clubs should be the focus. Further evidence in section 4.4 on scoring and time dependency will indicate that this is a valid approach and that little can be taken from an analysis undertaken for overall competition rates.

4.3 AFL club scoring events and their correlation

The next chapter will show that the Poisson distribution is suitable for approximating team's scoring and conceding of goals in the AFL as well as for behinds and the combination of both, known as scoring shots. Chapter 7 will present a pre-match Poisson prediction model that derives margin and probability of victory using approximated scoring rates for goals and behinds for each of the competing teams. An expected score is derived for each team and a subsequent winner and predicted margin of victory.

From a prediction and modeling viewpoint it is important to ascertain the level of correlation between events. If the correlation is minimal the modeling process is much simpler as independent Poisson distributions can be used to model each team's goals both on attack and defence. Previous research involving soccer has found the presence of relatively low correlations between the number of goals scored by the two opponents (Lee, 1997) (Karlis and Ntzoufras (2000)), however it was ignored when modeling due to the complications it presents. Recent work by Karlis and Ntzoufras (2001) proposed a bivariate Poisson model. This allowed for the correlation between team's scoring to be included in the model and resulted in an improved fit.

Analysis has been performed on the 1110 matches for the six seasons in the data set broken down by club. Each club has played a different number of matches due to the inclusion of nine finals per season; however, they have all played a minimum of 132 matches. Table 4.1 contains the number of matches each club has played over the six years of analysis broken down by home and away and finals. Pearson correlation calculations are based on the assumption that both X and Y values are sampled from an approximately normal population. Prior to calculating the correlation coefficients, the data have been transformed using square roots to stabilise the variance and bring the distribution closer to normality. However, nonparametric Spearman rank order correlation coefficients have been used for comparison purposes. These are based on ranking the two variables, and so make no assumption about the distribution of the values.

Table 4.1: Games played by AFL clubs between 1998 and 2003

Club	Home & Away	Finals	Matches
Adelaide	132	10	142
Brisbane	132	15	147
Carlton	132	9	141
Collingwood	132	6	138
Essendon	132	13	145
Fremantle	132	1	133
Geelong	132	1	133
Hawthorn	132	5	137
Melbourne	132	8	140
Kangaroos	132	10	142
Port Adelaide	132	9	141
Richmond	132	3	135
St. Kilda	132	2	134
Western Bulldogs	132	5	137
West Coast	132	5	137
Sydney	132	6	138

Relationships for individual clubs were investigated to ascertain whether there is dependency between goals and behinds for a team, both on attack and defence,. The relationship between goals and behinds was investigated for attack and defence and the results are presented in Table 4.2. In order to avoid a lot of spurious positives, the alpha value needs to be lowered to account for the number of comparisons being performed. The Bonferroni correction is a multiple-comparison correction used when several dependent or independent statistical tests are being performed simultaneously. In this instance, the significance level has been set at 0.002 (i.e. 0.05/32) for each test, the Bonferroni correction for 32 tests. The Bonferroni correction is applied throughout this thesis.

Using Pearson's coefficients, Table 4.2 shows that at the revised significance level of 0.002, only the West Coast has a significant correlation between goals and behinds on both attack and defence. It is also noted that the correlations are generally positive indicating that the more or less goals a team kicks, the more or less behinds they will score respectively. Scatter plots of West Coast's attack and defence relationship between the transformed variables of goals and behinds are presented in Figures 4.1 and 4.2 below.

Table 4.2: Pearson and Spearman correlations between goals and behinds for AFL clubs on attack and defence

Club	Attack				Defence			
	Pearson	P-val	Spearman	P-val	Pearson	P-val	Spearman	P-val
Adelaide	0.12	0.145	0.10	0.232	0.24	0.004	0.29	0.005
Brisbane	0.16	0.052	0.14	0.085	0.24	0.003	0.21	0.010
Carlton	0.25	0.003	0.23	0.006	0.19	0.021	0.18	0.030
Collingwood	0.22	0.011	0.22	0.009	0.16	0.066	0.13	0.139
Essendon	0.25	0.002	0.21	0.012	-0.04	0.674	-0.01	0.858
Fremantle	0.14	0.107	0.15	0.088	0.21	0.016	0.20	0.019
Geelong	0.02	0.862	0.01	0.880	0.10	0.230	0.09	0.314
Hawthorn	0.18	0.033	0.17	0.042	0.18	0.040	0.15	0.076
Melbourne	0.19	0.022	0.19	0.022	-0.11	0.210	-0.09	0.267
Kangaroos	0.01	0.878	0.01	0.827	0.14	0.101	0.10	0.214
P. Adelaide	0.10	0.223	0.07	0.393	0.18	0.029	0.22	0.009
Richmond	0.08	0.385	0.11	0.195	0.20	0.020	0.17	0.054
St. Kilda	0.16	0.070	0.16	0.065	0.25	0.003	0.27	0.001
W. Bulldogs	0.07	0.397	0.09	0.313	0.16	0.058	0.18	0.031
West Coast	0.29	0.001	0.32	0.000	0.30	0.001	0.26	0.002
Sydney	0.09	0.285	0.08	0.359	0.14	0.090	0.21	0.015

Figure 4.1: Plot of West Coast sqrt(behinds) vs sqrt(goals) on attack

```
      5 ˆ
        ,
        ,
        ,                        *     **         *
        ,                       **   **    *
      4 ˆ                    *  *    ***      *
        ,                    *  *   * **  *     *
sqrt(bh),                 *  ** *    * *** *      *
        ,         *     *    ** ***  ***** **
        ,                 ** ** *   ***  *    *   *
      3 ˆ           *     ** *  **   *   *
        ,                 *  *   ** * **
        ,                 *   *     **
        ,                 *    **
        ,                     *  *  **
      2 ˆ                       *
        ,                            *
        ,
        ,
        ,
      1 ˆ
        Šƒˆƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒˆƒ
          1        2        3        4        5        6

                             sqrt(gl)
```

Figure 4.2: Plot of West Coast sqrt(behinds) vs sqrt(goals) on defence

```
      5 ˆ
        ,                           *      *
        ,
        ,                         *
        ,                           *  *    * *
        ,                  *    *    **** * * * *
      4 ˆ                  *     **  *  * *    *
```

```
        ,                 *  * *     *      *
        ,            * * *  ** *** ** ****   * *
        ,            *    *  *  ** ** *   *
        ,                   ** * *  **    *     *
sqrt(bh),                *  *   ** *** ** ** *
      3 ˆ                   *  *      **
        ,                 *    *     * * *
        ,                   ** **  *  *  *
        ,                      *      *
        ,
        ,        * *
      2 ˆ                          *
        ,
        šƒˆƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒˆƒ
          1        2        3        4        5        6

                             sqrt(gl)
```

Whilst not as evident in the attack figure as the defence figure, there are outlying values that may be unfairly influencing the correlation between the variables, making them significant. It is acknowledged that although 30 out of the 32 correlations are positive and the percentage of significant correlations is larger than what we would expect by chance, the dependence between goals and behinds is minimal. The relative weakness of the correlation coefficients for each team on attack and defence suggests that the occurrence of goals and behinds can be treated as independent events. To investigate this further some calculations have been done assuming both independent and dependent distributions for goals and behinds.

A real example was looked at to compare the dependent and independent probabilities for teams kicking greater than 18 behinds given they had kicked greater than 18 goals. Assuming dependence, it was a simple case of counting the data and calculating the probability. There were 357 teams that kicked more than 18 goals in a match, and of these teams, 31 also scored more than 18 behinds. This gives a conditional probability for a team kicking more than 18 behinds given that it has kicked more than 18 goals of 0.087. If there is independence, then the conditional probability for more than 18 behinds given more than 18 goals is the same as the unconditional probability for more than 18 behinds. Out of the 1110 matches, the number of matches that had more than 18 behinds was 100 giving an unconditional probability of 0.090. The closeness of these probabilities indicates that an approach assuming independence or dependence could be justified as valid.

Of more importance from a modeling point of view is how the goals (and scores) between opposing teams in a match are related. It is easier to model independent events by using independent distributions rather than having to allow for the correlation between events. The correlation between goals scored by the two teams has been calculated for each team when it is the home side and when it is the away side. These results are presented in table 4.3. The correlations for the total score are presented in table 4.4. The Bonferroni correction is applied for each table again producing a significance level of 0.002.

Table 4.3: Correlations between goals for AFL clubs named as home or away team

Club	Home				Away			
	Pearson	P-val	Spearman	P-val	Pearson	P-val	Spearman	P-val
Adelaide	0.08	0.512	0.08	0.525	-0.24	0.042	-0.23	0.050
Brisbane	-0.11	0.338	0.11	0.359	-0.12	0.318	-0.12	0.305
Carlton	-0.18	0.149	-0.17	0.168	-0.27	0.023	-0.23	0.048
Collingwood	0.01	0.964	0.04	0.773	-0.18	0.153	-0.18	0.150
Essendon	0.05	0.656	0.04	0.733	-0.03	0.793	-0.05	0.686
Fremantle	-0.11	0.382	-0.16	0.194	-0.26	0.035	-0.26	0.038
Geelong	0.07	0.551	0.07	0.572	0.02	0.900	0.03	0.819
Hawthorn	-0.20	0.096	-0.25	0.042	0.01	0.902	0.03	0.805
Melbourne	-0.35	0.003	-0.34	0.004	0.12	0.305	0.11	0.357
Kangaroos	-0.18	0.132	-0.12	0.303	0.14	0.248	0.17	0.154
P. Adelaide	0.23	0.057	0.12	0.307	-0.22	0.063	-0.26	0.029
Richmond	-0.01	0.955	-0.05	0.695	0.10	0.416	-0.01	0.954
St. Kilda	0.03	0.823	0.01	0.915	-0.34	0.005	-0.25	0.038
W. Bulldogs	-0.14	0.239	-0.16	0.193	0.01	0.925	-0.10	0.410
West Coast	-0.27	0.026	-0.29	0.019	0.11	0.356	0.12	0.317
Sydney	-0.33	0.005	-0.23	0.052	-0.10	0.399	-0.09	0.439

Table 4.4: Correlations between score for AFL clubs named as home or away team

Club	Home				Away			
	Pearson	P-val	Spearman	P-val	Pearson	P-val	Spearman	P-val
Adelaide	0.04	0.719	0.04	0.738	-0.23	0.048	-0.24	0.041
Brisbane	-0.15	0.208	-0.13	0.264	-0.16	0.180	-0.14	0.248
Carlton	-0.21	0.084	-0.18	0.146	-0.29	0.013	-0.24	0.040
Collingwood	-0.01	0.916	0.01	0.955	-0.22	0.073	-0.21	0.080
Essendon	-0.02	0.835	-0.06	0.596	-0.10	0.422	-0.13	0.275
Fremantle	-0.18	0.148	-0.22	0.079	-0.32	0.008	-0.30	0.014
Geelong	0.07	0.586	0.06	0.648	-0.02	0.874	0.01	0.948
Hawthorn	-0.27	0.025	-0.29	0.016	-0.09	0.455	-0.06	0.628
Melbourne	-0.41	0.000	-0.39	0.001	0.09	0.477	0.09	0.440

Kangaroos	-0.20	0.089	-0.12	0.306	0.10	0.423	0.10	0.406
P. Adelaide	0.19	0.113	0.11	0.358	-0.29	0.015	-0.29	0.013
Richmond	-0.07	0.597	-0.08	0.500	0.08	0.538	-0.02	0.844
St. Kilda	-0.04	0.726	-0.06	0.617	-0.38	0.002	-0.31	0.009
W. Bulldogs	-0.20	0.104	-0.20	0.103	-0.05	0.713	-0.15	0.217
West Coast	-0.34	0.005	-0.39	0.001	0.07	0.576	0.09	0.445
Sydney	-0.36	0.003	-0.26	0.033	-0.12	0.312	-0.11	0.387

It is interesting to note the spread of positive and negative correlations as we would expect to see negative correlations in competitive team sport i.e. the more a team scores the less their opponent scores. There are only seven positive correlations in Table 4.4. To ascertain whether the proportion of negative correlations between home and away scores is significantly greater than 50%, a sign test was performed on the data, indicating that overall, Team A's score is negatively correlated with Team B's score in the AFL competition , ($P(X \leq 7 | X \sim Bin(32; 0.5)) = 0.001$). Those teams that display a positive correlation indicate that when their attack plays better than average so too does their defence. The teams with negative correlations indicate that if the attack plays above average, the defence plays below average or vice versa. Possible reasons for positive correlations could be the move of an influential player from attack or defence, whilst negative correlations could be put down to variations in form of the team. Other reasons for correlation could be the effect of weather conditions or playing at a venue that is more likely to produce high or low scoring. There may also be other reasons for the existence of positive and negative correlations.

Looking at the correlations between goals from Table 4.3, there are no significant correlations for any team at either home or away. From Table 4.4, Melbourne at home and St. Kilda away have significant Pearson correlations between their score and their opponent's score. Melbourne and West Coast have significant Spearman correlations at home. These relationships have been analysed more closely with scatter plots of the two variables presented in Figure 4.3, 4.4 and 4.5 respectively. Towards the extremities of each plot there are some outlying data points that may be contributing to the significance of the correlation for these teams.

Figure 4.3: Plot of Melbourne score vs opposition score at home

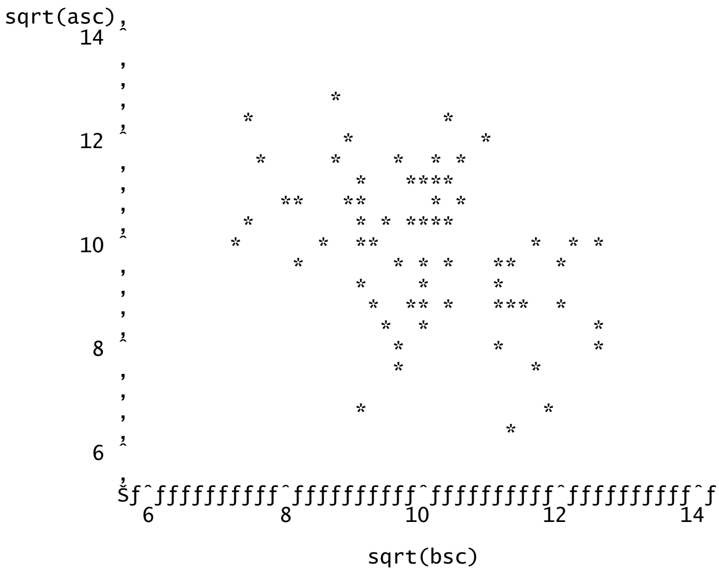


Figure 4.4: Plot of St. Kilda score vs opposition score away

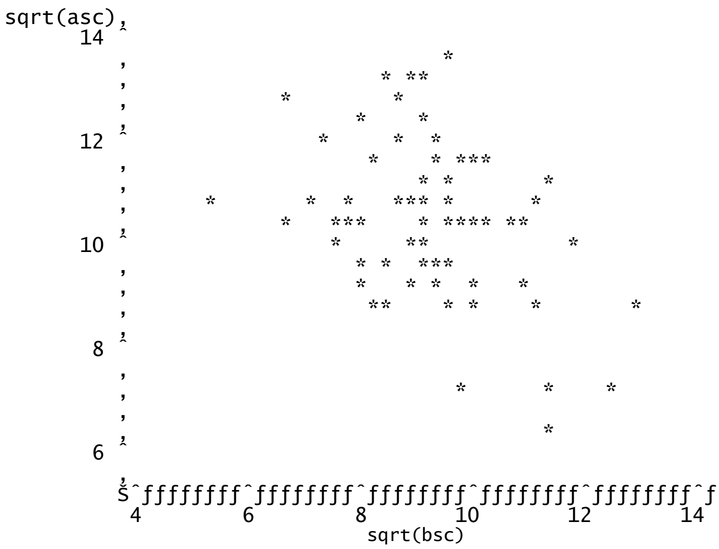
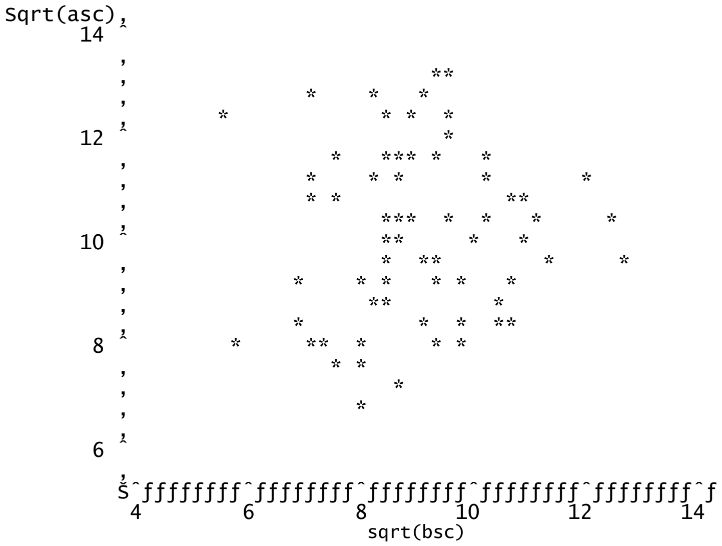


Figure 4.5: Plot of West Coast score vs opposition score at home



With only three of the 32 correlations significant, we have ascertained that the relationship between scores in the AFL competition is negligible and independent distributions could be used in a prediction model.

4.4 Scoring and time dependency in the AFL

Another issue to be explored is whether time dependency affects scoring. Do different parts of the game (e.g. near the end) contribute significantly more goals than other parts? Researchers have shown that scoring in soccer is dependent on time with more goals being scored towards the end of a match (Abt, 2002, Jinshan, 1986, Jinshan, 1993). This phenomenon has been explained by a number of factors such as player fatigue and lapses in concentration (Reilly, 1996). It has also been suggested that play could become more frantic towards the end of the match as teams chase a result. Again the distinction between soccer and AFL must be drawn here due to the scoring systems in place. The game of AFL rarely sees a drawn result as identified in Chapter Three, with only nine matches drawn between 1998 and 2003. Therefore, one is not likely to see radical changes in team line-ups to force a result at the end of a game producing the expectation of more goals in the second half. To investigate this further, differences in means have been investigated between the total points scored in the first and second half of matches for each of the 16 clubs. Table 4.5 below displays the summary statistics for these comparisons.

From Table 4.5 there is no evidence to suggest that halves of football can not be modeled independently. In the table, none of the clubs show a significant difference in their scoring returns between the first and second half at the revised level of significance of 0.003 (0.05/16). This indicates that there is not enough evidence to accept dependency between halves of football for the AFL competition.

Table 4.5: Summary statistics comparing team scoring in the first and second half of matches, 1998-2003

Club	Games	1st Half		2nd Half		Difference		t stat	d.f.	p-value
		Mean	S.D.	Mean	S.D.	Mean	S.D.			
AFC	142	48.8	16.2	48.0	19.2	0.7	23.5	0.37	141	0.71
BFC	147	52.6	17.4	55.1	18.1	-2.5	22.5	-1.33	146	0.18
CAFC	141	45.4	18.0	49.8	18.7	-4.3	20.1	-2.55	140	0.01
COFC	138	46.0	15.8	49.1	15.4	-3.1	19.2	-1.91	137	0.06
EFC	145	54.6	18.6	52.1	20.2	2.5	21.0	1.45	144	0.15
FFC	133	42.3	14.4	44.2	16.4	-1.9	19.9	-1.08	132	0.28
GFC	133	45.9	16.8	45.1	17.0	0.8	22.4	0.44	132	0.66
HFC	137	47.0	16.8	45.0	18.8	2.0	22.8	1.03	136	0.30
MFC	140	47.2	17.3	49.7	17.7	-2.6	21.7	-1.40	139	0.16
NMFC	142	53.1	16.1	52.3	19.5	0.8	24.6	0.36	141	0.72
PAFC	141	45.7	16.8	49.6	18.2	-3.9	20.5	-2.25	140	0.03
RFC	135	44.7	16.0	44.3	16.7	0.3	19.5	0.19	134	0.85
SKFC	134	45.7	16.9	43.3	16.2	2.4	22.3	1.25	133	0.22
WBFC	137	49.0	16.4	54.0	16.4	-5.0	20.6	-2.82	136	0.01
WCFC	137	46.2	17.1	47.1	18.5	-0.9	18.8	-0.57	136	0.57
SFC	138	48.1	15.9	50.0	17.8	-1.9	22.3	-1.00	137	0.32

It is pertinent to look also at the relationship between quarters for each club and whilst this analysis has not been presented here, the results will be discussed. The level of significance has been revised to 0.001, being 0.05 divided by 96 as there are six comparisons for 16 clubs. The correlation between what happens between quarters can be broken down as follows: - no team showed significant correlation in scoring between quarter one and two; three teams had significant correlation between quarter one and three; two teams had significant correlation between quarter one and four; two teams had significant correlation between quarter two and three; two teams had significant correlation between quarter two and four; and one team had a significant correlation between quarter three and four. Overall, there are ten correlations between quarters that are significant. Perhaps the changing of ends at quarters reduces the correlation as things such as end familiarity and wind conditions come into play. However, in light of the results above to do with scoring across halves, it has been assumed that a constant scoring rate is acceptable for modeling AFL matches. To further test whether this assumption is valid, analysis has been done to establish if scoring rates differ significantly across a match in the AFL competition.

This was done in SAS using a generalised linear model with a nested design. The variables quarter, club, season and match id were used as independent variables with the square root of score used as the dependent variable. Season and match are random effects with match nested within season, while club and quarter are regarded as fixed effects. The tests carried out tested whether: scores vary across seasons; scores vary across clubs; scores vary across quarters (on average) and whether clubs score differently across quarters (on average).

The hypotheses tested from the model indicate that scores vary significantly across seasons (F-statistic = 7.42 on 5 d.f., p-value <0.0001) and clubs (F-statistic = 11.06 on 15 d.f., p-value <0.0001), however, there is no significant difference between quarters of a match (F-statistic = 1.65 on 3 d.f., p-value = 0.1749) nor do different clubs score significantly differently across quarters (F-statistic = 1.03 on 45 d.f., p-value = 0.4122). The club result confirms the analysis done in section 4.1 about analysing the data for each club separately. Furthermore, the club by quarter result indicates that there is no reason to use a different scoring rate for different periods of a match and that one constant rate for each club would be acceptable.

4.5 Summary

The analysis in this chapter provides the framework for the models presented later in this thesis. It was shown that teams in the AFL score at different rates. Goals and behinds on offence were shown to have minimal correlation with goals and behinds conceded. This supports the use of separate attack and defence ratings for each team rather than a "whole of competition" rating.

Scoring goals and behinds showed minimal correlation indicating that these events could be modeled independently. A real example was considered assuming both dependence and independence and the difference in probabilities was negligible. The scores for competing teams in a match are negatively correlated for the competition overall. Some clubs showed a positive correlation with what their opponent's score, however a sign test indicated that the number of negative correlations was significantly higher. Finally, there was no evidence to suggest that a different rate should be used for different periods of the game. This meant that one rate could be used for the match (distinct for each team), simplifying any prediction model.

Chapter 5 Goals scored and conceded in the AFL

5.1 Introduction

A suitable starting point for any sporting prediction model is an understanding of the scoring outcomes that are inherent in the event. It is important that a mathematical distribution can be fitted to scoring events to make the prediction of such events more robust. Researchers have been doing this since the 1950s (Moroney, 1956) with advances in the literature more recently (Greenhough, Birch, Chapman and Rowlands, 2002). It is generally accepted that the frequency of scoring events in sport will fall under one of two distributions - either the negative binomial distribution (NBD) or Poisson distribution. However, the majority of work in this area has been performed on soccer data sets and there are a number of reasons why the analysis performed here is important.

Firstly, the scoring system in Australian football is distinctly different to soccer. Secondly, the frequency of scoring events is much higher than soccer. In soccer it is not uncommon for a match to involve no scoring events at all with a match ending in a nil all draw. It would be very rare to see more than six goals scored in a match by both sides. Chapter 3 highlighted that in the AFL, the average score per side is close to 100 points with around 50 scoring events per match. Because of the nature of AFL football and the scoring system, a draw has only one chance in one hundred of happening. Thirdly, the AFL competition is unique in that there are no national or international competitions or concurrent competitions taking place during a season. This means the best players, if healthy, are on the field each week and therefore the strength of sides remain fairly constant throughout a season. Finally, distributions have not been fitted to team's defensive capabilities in the academic literature. There has been plenty of work done on attacking scores for individual teams but the distribution of concession rates has not been looked at. This analysis looks at the defensive capabilities of sides as well as their attacking capabilities.

This chapter examines the seven seasons of AFL data from 1998 to 2003. Due to the results from Chapter 4 the analysis will be performed on a team by team basis, with scoring events examined in an attempt to fit a suitable distribution to these events both for offence and defence. In Chapter 6 on home advantage, analysis will be done to investigate whether there is any difference in scoring returns according to venue.

5.2 Overview of statistical distribution applicable to scoring events

It is accepted that the occurrence of scoring events in sport can be approximated by one of two distributions. These distributions are the Poisson distribution and the NBD and both are used to model the number of events occurring in given time intervals e.g. a sporting match. A description of both of these distributions and how they apply in a sporting sense will be given before the analysis.

5.2.1 The Poisson distribution

The Poisson distribution describes the number of events occurring in equal intervals of time if these events occur independently, randomly, singly and uniformly with the same rate in each interval (Baxter and Stevenson, 1988). In this case, the events of interest are the number of scoring events an AFL side registers or concedes in a match. To specify the Poisson distribution, only the mean scoring rate (or concession rate) is required with the probability of observing x events given by:

$$\Pr(X = x) = {e^{-\lambda}\lambda^{x}}\big/{x!} \qquad x = 0, 1, 2, \ldots,$$

where λ is the mean scoring rate and the variance and differ from team to team.

The assumptions that must be met for the Poisson distribution to apply are that the probability of an event within a certain interval does not change over different intervals and that the probability of an event in one interval is independent of the probability of an event in any other non-overlapping interval. In an AFL match it might be argued that teams can get on a roll and 'bang' on a couple of goals in quick succession, however it is assumed that the second assumption is satisfied in an AFL game. One would expect that different venues, weather conditions or the strength of the opposition might reduce or increase the probability of scoring and therefore cause the first assumption to fail. If this were the case, the NDB would be the appropriate distribution for AFL scoring events.

5.2.2 The negative binomial distribution

The NBD is similar to the Poisson distribution in that events must occur independently, randomly, singly and uniformly with the same rate in each interval. However, the rate of occurrence of these events is a random variable following the gamma distribution (Johnson N.L. & Kotz, 1969). This goes against the first assumption of the Poisson distribution and suggests that scoring rates differ from match to match but during any particular game the events will still occur randomly. Unlike the Poisson distribution, the NBD requires two parameters, namely the mean and variance of the scoring rate.
The probability of observing x events in an interval is given by:

$$P(X=x) = \binom{k+x-1}{k-1} p^k (1-p)^x \qquad x = 0, 1, 2, \ldots\ldots\ldots,\infty$$

where $k > 0$ and $0 < p < 1$.

The parameters can be estimated in a number of ways and this analysis uses the following formulae for the population parameters:

$$\mu = {k(1-p)}\big/{p}$$

$$\sigma^2 = {k(1-p)}\big/{p^2}$$

For the distribution to hold $\sigma^2 > \mu$
The limit of this distribution, as $x \rightarrow \infty$, approaches the Poisson distribution.

Intuitively, the NBD would seem to be the most appropriate distribution as teams are unlikely to score at the same rate week in week out, for the reasons given above. The analysis contained in this chapter will show that this is not necessarily the case, with the Poisson distribution providing a better fit to AFL team scoring data than the NBD.

5.3 The AFL competition

Initially, the work of others (Reep and Benjamin, 1968, Baxter and Stevenson, 1988, Pollard, Benjamin and Reep, 1977) was followed to investigate the distribution of scoring events by team in the AFL competition. The distinction between the work contained herein and others is the unusual scoring system in AFL. Instead of investigating goals, behinds had to be looked at also. It was also decided to combine the two events and look at the distribution of scoring shots. Detailed analysis is given on goals due to the fact they are worth six times a behind, with brief analysis on behinds and scoring shots later in the chapter.

Although analysis will only be performed for individual teams, it is worthwhile to present an overview of the results relating to the entire competition. Firstly, analysis was done on the total scoring events in a match and as expected, it was found that for goals the NBD was suitable (p-val = 0.51) whilst the Poisson distribution was not (p-val < 0.001). For behinds, the NBD was not applicable as the variance is less than the mean whilst the Poisson distribution provided an appropriate, albeit weak fit (p-val = 0.11). Matches were then analysed according to team goals on attack and defence and it was found that neither goals nor behinds could be approximated by the Poisson distribution or NBD with highly significant p-values in each instance. These results are hardly surprising as we would expect scoring returns in a match to differ according to the strength of the opposition, weather conditions, importance of match and other similar factors. The strength of environmental effects (strength/weakness of the opposition, weather conditions, importance of match etc.) and the variation in goals from match to match contribute to this lack of fit for either distribution.

Finally, matches were broken down by quarter and it was found that the Poisson distribution provided an appropriate fit for goals (p-val = 0.32) and behinds (p-val = 0.77) and the NBD was also appropriate for goals (p-val = 0.44) and behinds (p-val = 0.70). Because of this, the analysis for individual teams was performed using a rate that was constant across the match. It has already been shown that one rate can be used for each quarter, on attack and defence, although the rate differs from team to team.

For individual team scoring rates in the AFL competition, the Poisson distribution is a better approximation than the NBD distribution. As this information forms the backbone of the Markov process model that will be introduced later in this thesis, the following sections concentrate on demonstrating that the Poisson distribution is the best approximation. It will be shown that goals occur randomly but at a constant rate throughout a match and from match to match.

5.4 Individual team goals on offence

Table 5.1 presents summary statistics for each club's offence and p-values indicate the appropriateness of the Poisson distribution for describing the results on a quarter basis. Brisbane has played the most games with 147, whilst Fremantle and Geelong have played the least with 133. Of most importance is the fact that the number of scores per quarter for each of the 16 AFL clubs is suitably approximated by the Poisson distribution. Collingwood has the lowest p-value of 0.19 and this may be explained by the instability of their results over the period of analysis where they have had varying degrees of success, however there is no reason to discount the Poisson distribution for any club.

Table 5.1: Summary statistics for goals for by quarter and Poisson fit, 16 AFL clubs, 1998 to 2003

Team	Matches	Goals	Mean	Variance	p-value
Adelaide	142	1989	3.50	3.58	0.64
Brisbane	147	2299	3.91	3.89	0.20
Carlton	141	1957	3.47	3.98	0.68
Collingwood	138	1910	3.46	3.05	0.19
Essendon	145	2274	3.92	4.06	0.53
Fremantle	133	1660	3.12	3.01	0.95
Geelong	133	1761	3.31	3.31	0.98
Hawthorn	137	1852	3.38	3.74	0.77
Melbourne	140	1986	3.55	3.69	0.24
Kangaroos	142	2196	3.87	3.91	0.26
Port Adelaide	141	1945	3.45	3.53	0.98
Richmond	135	1746	3.23	3.29	0.28
St Kilda	134	1750	3.26	3.45	0.57
Western Bulldogs	137	2072	3.78	3.82	0.57
West Coast	137	1869	3.41	3.61	0.90
Sydney	138	1993	3.61	3.55	0.90
Competition	2220	31259	3.52	3.65	0.32

5.4.2 Individual team goals on defence

Having looked at offence in the previous section it is logical to examine how clubs concede goals and whether this can also be approximated by the Poisson distribution. Table 5.2 presents the summary statistics for each team in the competition. Defence, like attack, is also well approximated by the Poisson distribution. Only one rate of the 16 shows evidence to suggest the Poisson distribution is not appropriate (Port Adelaide). On the other hand, a number of clubs are very well approximated by the Poisson distribution (Brisbane, Geelong, Kangaroos) and there is no reason to suspect the Poisson distribution cannot be used to approximate club scoring concession in the AFL competition.

Table 5.2: Summary statistics goals against by quarter, 16 AFL clubs, 1998 to 2003

Team	Matches	Goals	Mean	Variance	p-value
Adelaide	142	1915	3.37	3.55	0.78
Brisbane	147	1919	3.26	3.52	0.92
Carlton	141	2007	3.56	3.77	0.83
Collingwood	138	1916	3.47	3.71	0.53
Essendon	145	1794	3.09	2.63	0.07
Fremantle	133	2083	3.92	4.08	0.64
Geelong	133	1900	3.57	3.47	1.00
Hawthorn	137	1860	3.39	3.28	0.86
Melbourne	140	2066	3.69	3.55	0.71
Kangaroos	142	2130	3.75	3.98	0.90
Port Adelaide	141	1845	3.27	3.45	0.00
Richmond	135	1867	3.46	3.60	0.07
St Kilda	134	2059	3.84	4.45	0.10
Western Bulldogs	137	2079	3.79	3.70	0.32
West Coast	137	1959	3.57	3.81	0.36
Sydney	138	1860	3.37	3.25	0.61
Competition	2220	31259	3.52	3.65	0.32

5.4.2.1 Port Adelaide defence analysis

Analysis has been done similar to the above on a season by season basis for Port Adelaide's defence with the results contained in Table 5.3.

Table 5.3: Summary statistics goals against by quarter by season, PAFC, 1998 – 2003

Season	Position	Matches	Goals	Mean	Variance	p-val
1998	10th	22	294	3.34	3.77	0.17
1999	7th	23	316	3.43	3.96	0.14
2000	14th	22	339	3.85	4.01	0.28
2001	5th	24	301	3.14	3.00	0.88
2002	3rd	25	304	3.04	3.13	0.67
2003	3rd	25	291	2.91	2.61	0.08
Total		141	1845	3.27	3.45	0.00

Although Port Adelaide's defence is not approximated by the Poisson distribution for the six seasons combined, there is not one season alone that can not be approximated by the distribution. The implication is that once allowance is made for differing performance over seasons, the Poisson distribution approximates concession in the AFL. Again it must be noted that better fits would result from fitting the data by season and there is no reason to assume that the Poisson distribution does not provide an adequate approximation of the data.

5.5 Behinds

Having seen that goals on attack and defence can be well approximated by the Poisson distribution for AFL teams, this section looks at behinds. It is evident from Table 5.4 that the Poisson distribution is also a suitable approximation for attacking and defensive behinds. Only Geelong's offensive rate of scoring behinds has a p-value less than 0.05 and this could be expected by chance. Therefore, there is no reason to reject the Poisson distribution as a suitable approximation for the distribution of behinds in the AFL.

Table 5.4: Summary statistics behinds by quarter

Club	Games	Attack			Defence		
		Mean	Variance	P-value	Mean	Variance	P-value
Adelaide	142	3.19	3.24	0.72	2.83	3.05	0.40
Brisbane	147	3.46	3.34	0.60	2.93	2.67	0.13
Carlton	141	2.98	3.47	0.47	2.99	3.02	0.77
Collingwood	138	3.02	2.96	0.50	3.17	2.94	0.78
Essendon	145	3.14	2.88	0.30	2.88	3.08	0.74
Fremantle	133	2.89	2.83	0.72	3.05	3.17	0.25
Geelong	133	2.88	3.02	0.04	2.85	3.03	0.53
Hawthorn	137	2.72	2.69	0.81	3.06	3.05	0.84
Melbourne	140	2.95	2.70	0.32	3.11	3.16	0.90
Kangaroos	142	3.16	3.33	0.42	2.96	2.64	0.73
Port Adelaide	141	3.15	2.88	0.43	2.78	2.64	0.91
Richmond	135	2.85	2.82	0.98	3.04	3.01	0.90
St. Kilda	134	2.65	2.85	0.85	3.17	3.24	0.76
Bulldogs	137	3.07	2.86	0.81	3.08	3.22	0.87
West Coast	137	2.87	2.74	0.41	3.12	3.07	0.60
Sydney	138	2.88	2.90	0.37	2.94	3.01	0.92

5.6 Scoring shots

In the AFL competition, the combination of goals and behinds make up scoring shots, even though a team may have had more 'actual' scoring shots than the score line reflects. As it has already been shown that goals and behinds can be well approximated by the Poisson distribution, it is expected that the combination of both would be able to be approximated by the Poisson distribution as well. Table 5.5 indicates that this is the case with only Fremantle's attacking shots and St. Kilda's defensive shots having a p-value less than 0.05. There is no reason to suggest this is not due to chance. As a result it can be assumed that shots for and against for teams in the AFL competition can be well approximated by the Poisson distribution.

Table 5.5: Summary statistics scoring shots by quarter

Club	Games	Attack			Defence		
		Mean	Variance	P-value	Mean	Variance	P-value
Adelaide	142	6.69	6.93	0.27	6.20	7.18	0.09
Brisbane	147	7.37	6.94	0.39	6.19	6.30	0.66
Carlton	141	6.45	7.64	0.65	6.55	7.15	0.60
Collingwood	138	6.48	6.09	0.48	6.64	6.39	0.73
Essendon	145	7.06	7.14	0.23	5.97	5.23	0.81
Fremantle	133	6.02	6.05	0.04	6.96	6.99	0.58
Geelong	133	6.19	6.26	0.30	6.42	6.33	0.51
Hawthorn	137	6.09	6.65	0.34	6.45	6.45	0.54
Melbourne	140	6.49	6.65	0.80	6.80	6.15	0.33
Kangaroos	142	7.02	7.12	0.93	6.71	6.51	0.48
Port Adelaide	141	6.60	6.04	0.48	6.05	6.40	0.83
Richmond	135	6.08	5.85	0.91	6.50	6.15	0.15
St. Kilda	134	5.92	6.07	0.54	7.01	8.43	0.01
Bulldogs	137	6.85	6.48	0.60	6.87	7.04	0.54
West Coast	137	6.28	6.88	0.41	6.70	7.16	0.72
Sydney	138	6.49	6.13	0.64	6.31	6.43	0.54

5.7 Summary

This chapter has shown that competition scoring rate is best approximated by the negative binomial distribution. Perhaps this is to be expected as venues, quality of competitors and weather conditions can affect the number of goals scored. Individual club's attacking and defensive rates are best approximated by the Poisson distribution with the NBD inapplicable in a number of cases due to under dispersion.

Of the 32 goal rates analysed, only Port Adelaide's defence shows significant evidence to suggest the Poisson distribution will not approximate it. However, when Port Adelaide was analysed on a season by season basis there is no evidence to suggest the Poisson is not a good approximation. The fit of the distribution to the majority of clubs indicates there is solid evidence to suggest the Poisson distribution is a very good approximation to both attack and defence in the AFL competition. Behinds and scoring shots were also analysed and the fit of the Poisson distribution to these rates was adequate. Therefore, it can be assumed that team scoring and concession in the AFL competition can be approximated by fitting the Poisson distribution.

The analysis in this chapter supports the pre-match prediction model using historical data that is presented in Chapter 7. There is no need to use a different rate for each quarter of a match. As a result, a model has been implemented with only one rate used for the attack and defence of each side. Negative binomial regression is used to predict the goal and behinds of each team in a match. The Poisson nature of AFL team scoring suggesting that goals occur randomly but at a constant rate is also strong support to use a Markov process to approximate AFL football and this will be introduced in Chapter 9.

Chapter 6: Home/venue advantage in the AFL competition

Home advantage is a widely accepted phenomenon in sport and the AFL competition is no exception. Clarke's work on the topic is extensive and will be extended upon by the analysis in this chapter for the period 1998 to 2003 (Clarke, 2005, Stefani and Clarke, 1992). Elements of it will be reproduced in this chapter allowing for comparison where necessary.

6.1 Home and away level HAs – AFL nominated home team

In the AFL competition the team named in the draw as the home side is labelled team 'A' and the visiting side is named team 'B'. Using traditional measures of HA, various indicators have been calculated to investigate the presence of HA in the competition. Table 6.1 shows the percentage of results experienced across the six seasons, 1998 – 2003. Over this period, the home side won 656 matches, lost 445 matches and drew a further nine matches. By including a draw as half a win, the nominal home side won 59.5% of matches indicating that teams do perform better at home than away. In a manner similar to Clarke, the value of HA has also been calculated according to the average number of points scored in the match compared to the average margin of victory by the home side (Stefani and Clarke, 1992, Clarke, 2005). This ratio translates to the number of points scored for every point attributed to HA. Of particular interest is the six year totals for home advantage of 8.2 points. This value is the same as Clarke found for the period 1980 to 1995 and indicates that whilst HA varies from season to season its underlying value remains fairly constant.

Table 6.1: Match results and HA in points ratio for the nominal home team, 1998-2003

Season	Matches	Home Win %	Away Win %	Draw %	HA in Points	Total Points	Points:HA
1998	185	63.2	36.2	0.5	8.3	186.7	22.4
1999	185	58.4	41.1	0.5	8.3	191.1	22.9
2000	185	57.8	41.1	1.1	10.9	207.1	18.9
2001	185	50.8	49.2	0.0	3.1	194.6	61.9
2002	185	65.4	33.5	1.1	11.0	189.5	17.2
2003	185	58.9	39.5	1.6	7.6	188.5	24.7
98 - 03	1110	59.1	40.1	0.8	8.2	192.9	23.4

6.2 Difference in performance advantage according to venue

Interstate sides enjoy a healthy home advantage when they are playing their matches in their home state against a visiting opposition. As has been mentioned elsewhere, this has been researched extensively elsewhere and will not be revisited here (Stefani and Clarke, 1992, Bailey and Clarke, 2004, Clarke, 2005). The pre-match prediction model, introduced in Chapter 7, factors in home advantage by using a rating for each team according to where the game is being played. Melbourne clubs have seen the erosion of their home advantage with the use of only two grounds in Melbourne from 2006 onwards. In light of this, it is important to determine if Melbourne teams perform differently at these two venues. If, as expected, this is the case, a different scoring and concession rate will be used for each venue. Alternatively, Melbourne clubs would use only one rate for games played in Melbourne if there is no significant difference between the two venues. Table 6.2 contains the average team score at both venues for all of the Melbourne clubs as well as Geelong, who play home games in Melbourne from time to time. To investigate if any significant difference exists, a z-test was conducted on the mean points scored per team in a match at each venue using the summary statistics from Table 6.2.

Table 6.2: Summary statistics for matches played at MCG and Docklands, 2000 – 2003

Venue	Matches	Mean	Variance
Docklands	181	102.21	711.21
MCG	188	95.52	764.76

From the data, a z-statistic of 3.35 is derived indicating that there is sufficient evidence to suggest that the average score by a side differs between Docklands and the MCG (p-val <0.001) by nearly seven point. This result suggests that Melbourne clubs should be rated differently according to whether they are playing at the MCG or Docklands and the different ratings for sides will be explained further in Chapter 7.

As a national competition, the AFL regularly involves a match where one side has had to travel interstate. With the rationalisation of grounds in Melbourne it can be said that the only team who enjoys a home advantage is Geelong when they play at Kardinia Park. It was decided to reduce the data set to matches that involved at least one side having travelled interstate or Geelong games at Kardinia Park. Games where both teams were interstate, such as games at Manuka Oval or York Park were omitted from the analysis. Further, some clubs experimented 'selling' their home games to an interstate side and in these instances the home side was actually the interstate side. Adjustments have been made to reflect the true scenario of the match. This has also been done for some finals in Melbourne where an interstate side was playing a Melbourne side but was named as the home team. By applying these changes to the data set, the number of matches where there was an actual home advantage reduced to 673 games. This data set has been used for the rest of the analysis in this chapter.

6.3 Home and away level HAs – actual home advantage involved

By removing games where no home advantage is perceived to exist, it is hoped that a more accurate reflection of home advantage within the AFL will be achieved. The analysis from 6.1 has been replicated on this data set to illustrate the effect of 'actual' home advantage within the AFL competition. Table 6.3 contains the data broken down by season for the 673 matches.

Table 6.3: Match results and HA in points ratio for perceived home teams, 1998-2003

Season	Matches	Home Win %	Away Win %	Draw %	HA in Points	Total Points	Points:HA
1998	114	60.5	38.6	0.9	8.2	184.6	22.5
1999	114	59.6	40.4	0.0	12.4	188.5	15.2
2000	108	58.3	39.8	1.9	13.5	206.8	15.3
2001	112	55.4	44.6	0.0	11.1	191.7	17.3
2002	112	70.5	28.6	0.9	16.1	190.3	11.8
2003	113	63.7	33.6	2.7	12.6	186.4	14.8
98 - 03	673	61.4	37.6	1.0	12.3	191.3	15.6

It is evident that when neutral games are removed, the effect of home advantage is more pronounced. Although the win percentage for the home side does not increase by much (59.1% cf. 61.4%), the points value for home advantage has increased by over four points. Furthermore, the ratio of points to home advantage has dropped from 23.4 to 15.6 points. Whilst most of the six seasons are comparable between all games and the perceived home advantage games, the 2001 season stands out. During this season the home advantage between both data sets differs greatly (3.1 cf. 11.1) and this must be due to the strong performance of the away named side in games that were at neutral venues. From this analysis, it can be seen that playing at home produces a strong advantage, particularly when one team has travelled interstate. With this in mind, any predictive modeling done for AFL data will be improved by taking the venue into consideration.

6.4 Impact of home advantage on match performance statistics

Although home advantage is generally couched in terms of points and usually looked at from a scoring point of view, it was decided to look at its effect on other facets of the game. This is particularly important for the research in this thesis as the Markov process models presented later rely heavily on transition probabilities between "states of play". It is widely accepted that home advantage can be caused by psychological factors and generally these are extremely difficult, if not impossible, to quantify. To try and gain a better understanding of this effect in AFL descriptive analysis has been performed using various performance statistics for the 673 matches where there was a perceived home advantage. The analysis contained in this section mirrors the work done earlier on home advantage as a function of points.

The first area of interest was the total disposals per side within a match. Disposals consist of all instances where a player disposes of the ball by hand or foot in a positive or negative manner. The effect that home advantage has on the number of disposals within a match is contained in Table 6.4.

Table 6.4: HA as a function of disposals within a match for the home team, 1998 – 2003

Season	Matches	Home Win %	Away Win %	Draw %	HA in Disposals	Total Disposals	Disposals:HA
1998	114	63.2	36.8	0.0	10.9	566.2	51.9
1999	114	62.3	36.8	0.9	13.7	591.6	43.3
2000	108	55.6	44.4	0.0	12.6	613.6	48.9
2001	112	62.5	36.6	0.9	13.0	589.9	45.4
2002	112	73.2	25.9	0.9	19.0	586.5	30.9
2003	113	64.6	35.4	0.0	13.0	586.4	45.1
98 - 03	673	63.6	36.0	0.4	13.7	588.8	43.0

It is evident from the table that the home team enjoys an advantage over their opponents in having the ball. In fact, for the six years of analysis, the home side has enjoyed 13.7 extra disposals on average per match. Upon looking at the match as a whole, for every 43 possessions within the match, one disposal can be attributed to home advantage.

Of more interest may be the relationship between home advantage and soft and hard possessions. What is meant by hard possessions is the gaining of the ball in a contest where it is seemingly “up for grabs”. Hard possessions have been categorised as including the following: hard ball get, contested mark, earned mark and ruck get. On the other hand, a soft possession is where there is no contest for the ball and includes the following: loose ball get, uncontested mark and gather. The following tables display the effect of home advantage for these types of possession.

Table 6.5: HA as a function of hard possessions within a match for the home team, 1998-2003

Season	Matches	Home Win %	Away Win %	Draw %	HA in Hard Poss	Total Hard Poss	Hard Poss:HA
1998	114	55.3	42.1	2.6	3.0	85.2	28.8
1999	114	50.0	48.2	1.8	1.2	98.1	79.3
2000	108	53.7	40.7	5.6	1.4	98.7	72.0
2001	112	59.8	37.5	2.7	2.3	104.7	44.6
2002	112	58.0	38.4	3.6	2.2	96.3	43.7
2003	113	61.9	35.4	2.7	2.4	95.0	38.9
98 - 03	673	56.5	40.4	3.1	2.1	96.3	45.9

Table 6.6: HA as a function of soft possessions within a match for the home team, 1998-2003

Season	Matches	Home Win %	Away Win %	Draw %	HA in Soft Poss	Total Soft Poss	Soft Poss:HA
1998	114	57.9	39.5	2.6	3.2	294.7	92.8
1999	114	57.0	42.1	0.9	7.8	308.4	39.4
2000	108	55.6	42.6	1.9	4.0	314.9	78.4
2001	112	56.3	41.1	2.7	3.5	285.8	80.8
2002	112	66.1	33.0	0.9	9.5	277.2	29.1
2003	113	55.8	44.2	0.0	3.8	279.5	73.8
98 – 03	673	58.1	40.4	1.5	5.3	293.3	55.2

The tables show that whilst there is a slight advantage to the home side for both hard and soft possessions, the overall effect is minimal. For every 46 hard possessions, in a match only one is due to home advantage and for soft possessions, only one can be attributed to home advantage for every 55 that occur. We would expect a slight advantage to the home side because they enjoy more disposals and score more points and this is what is seen here. The issue of psychology cannot really be raised here as the advantage is minimal, however further investigation is required for free kicks, errors and one percent acts.

An analysis of free kicks may provide an insight into whether umpires are likely to be influenced by the home crowd. Crowd effect is an accepted component of home advantage (Agnew, 1994)and perhaps the crowd is able to influence the umpire's decision making in some way. It would be hoped that the home teams aren't treated favourably by the umpires otherwise one would have to question whether teams are competing on a level playing field. Table 6.7 contains the summary information for free kicks.

Table 6.7: HA as a function of free kicks within a match for the home team, 1998-2003

Season	Matches	Home Win %	Away Win %	Draw %	HA in FRFO	Total FRFO	FRFO:HA
1998	114	56.1	37.7	6.1	1.1	29.7	26.5
1999	114	50.9	42.1	7.0	1.0	31.2	29.9
2000	108	61.1	26.9	12.0	2.2	30.3	13.9
2001	112	56.3	35.7	8.0	1.5	28.3	19.5
2002	112	54.5	37.5	8.0	1.2	30.6	25.8
2003	113	62.8	32.7	4.4	2.3	30.3	13.2
98 - 03	673	56.9	35.5	7.6	1.5	30.1	19.5

Table 6.7 shows that the home side receives an extra 1.5 free kicks in a match over their opposition, which equates to a free kick once in every 20 given as a result of playing at home. Crowd effect would surely play a part in this but so too could the fact that the home team has an extra 14 disposals per match, on average, over their opponents, giving the opposition more opportunity to infringe the rules. With this in mind, the extra 1.5 free kicks to the home side do not indicate anything abnormal.

One-percent acts were looked at next to see if home sides outperformed their opponents in the less obvious areas of a game such as tackling, shepherding and spoiling. For this area, the data set does not include all of the statistics for the whole period covered. As a result, it will be noticed that the number of acts per match will increase with time. The summary for these acts is contained in Table 6.8.

Table 6.8: HA as a function of 1% acts within a match for the home team, 1998-2003

Season	Matches	Home Win %	Away Win %	Draw %	HA in 1%s	Total 1%s	1%s:HA
1998	114	55.3	43.0	1.8	0.9	35.3	39.9
1999	114	64.9	33.3	1.8	3.9	69.5	17.7
2000	108	57.4	39.8	2.8	2.9	88.0	30.7
2001	112	58.0	38.4	3.6	4.8	102.1	21.4
2002	112	59.8	36.6	3.6	5.8	144.5	25.1
2003	113	58.4	40.7	0.9	2.5	147.9	59.0
98 - 03	673	59.0	38.6	2.4	3.4	97.7	28.4

Seasons 2002 and 2003 saw the inclusion of spoils as a recorded statistic for a match and this is why there is a significant increase in the average number of 1% acts per match for these seasons. It is evident that the home team outperforms their opponents in the small things in the range of about three per game.

Finally, the number of errors committed by each side was analysed. The error count consists of the number of direct turnovers by hand or foot to the opposition, commonly known as the number of "clangers". The results of the analysis are contained in Table 6.9.

Table 6.9: HA as a function of errors within a match for the home team, 1998-2003

Season	Matches	Home Win %	Away Win %	Draw %	HA in Errors	Total Errors	Errors:HA
1998	114	55.3	36.8	7.9	1.7	36.5	22.0
1999	114	54.4	40.4	5.3	1.0	38.5	37.2
2000	108	51.9	43.5	4.6	0.9	45.3	49.5
2001	112	49.1	42.0	8.9	-0.2	30.3	-135.8
2002	112	50.9	43.8	5.4	0.5	28.6	60.4
2003	113	48.7	47.8	3.5	-0.2	28.5	-179.2
98 - 03	673	51.7	42.3	5.9	0.6	34.6	55.9

Intuitively, one would expect the home side to commit fewer errors than the visitors and this is the case for two of the six seasons. Overall, the number of extra errors the home side commits over their opposition is less than one and perhaps this can be attributed to the extra disposals of the home side, increasing the opportunity for error.

6.5 Summary

The analysis in this Chapter evaluates the advantage that the home teams enjoy in the AFL competition. The competition requires different analysis to other sporting competitions around the world due to the rationalization of grounds in Melbourne. This phenomenon has seen the nature of the competition draw change so that there are often matches where both teams are playing at their home venue. This is also the case when interstate teams from Perth and Adelaide play against each other. It was shown that if analysis is done using the home named team and the away team, with no regard to who is competing, the home team enjoys an advantage of 8.2 points and this is consistent with earlier work in this area.

The performance of Victorian clubs at Docklands and the MCG was analysed to ascertain whether the scoring return differed at these venues. It was found that there is a significant difference between these venues with Docklands a higher scoring venue than the MCG by almost seven points. This result justifies the approach to pre-match prediction in the next chapter of using different ratings at these venues for the Victorian clubs who play a substantial number of games at these venues. Ignoring all matches in which there was no clear home team and away team, an even stronger home advantage of 12.3 points was calculated.

The final part of the chapter investigated the advantage that was gained by playing at home for different types of match statistics. This was done with a view to the later chapters of this thesis where models are presented that use transition probabilities based on these match statistics. As expected, home teams enjoy an advantage over their opposition in the number of disposals they accumulate and there are also small advantages for hard and soft possessions. The home side commits more errors than the away side and whilst intuitively this is not expected, the extra ball the home team has would affect this result. The penultimate chapter of this thesis looks at the performance of teams locally and interstate and determines which transition frequencies differ significantly between the two scenarios.

Chapter 7: Pre-match prediction

7.1 Introduction

The prediction of sporting outcomes prior to the commencement of a match has long been of interest to those in academic research. Many different techniques and a wide range of sports have been the subject of analysis. With the advent and subsequent growth in sports betting both worldwide and in particular in Australia in recent years, the general public has joined the quest to accurately predict match outcomes.

Whilst Australian Rules football is a sport predominantly enjoyed only by Australians, it has been popular in the literature from a prediction point of view. Clarke has developed a prediction model that uses only the competing teams and venue to derive a winner, margin of victory and chance of victory, via an exponential smoothing process that produces team ratings and venue advantages (Clarke, 1993). The exponential smoothing process assigns exponentially decreasing weights to scores as the matches get older. This means recent matches are given relatively more weight for forecasting than older matches. Although this is a simple approach to prediction, the model has proved that it is as good, if not better, than media experts in picking winners and margins of victory. In more recent times it has been made available to subscribers who have used the estimated probabilities of victory and 'fair' lines to exploit market inefficiencies with great success[1]. More recently, Bailey (Bailey and Clarke, 2004) has applied a multiple linear regression model to 100 seasons of AFL/VFL data dating back to 1897 to develop a highly successful prediction model that shows substantial profits when applied against historical bookmaker data. He also compared his work to that of Clarke and found his results to be superior.

This chapter intends to show that a static pre-match model reaches a certain ceiling level, in terms of accuracy and appropriateness, which is very hard to exceed. The randomness associated with sporting events is the major contributor to this ceiling level as some

[1] http://www.smartgambler.com.au/testimonials/afl04.html

matches don't unfold as the ratings suggest they will. It is hoped that the models presented later in this thesis which are of a more dynamic nature will be better able to identify and adjust to match events as they happen, and to quickly identify matches that are not tracking according to form. For this reason, the model contained in this chapter is fairly basic but it serves to show that different techniques can be used to produce very similar results. The last part of this chapter will investigate whether a pre-match prediction model can be improved upon if the margin at half time is used to update the prediction. This will be done with reference to the model presented in this chapter and benchmarked against the model of Clarke (Clarke, 1993).

7.2 Overview of this approach

As Clarke(Clarke, 1993) and Bailey (Bailey and Clarke, 2004) have shown their models to be the best in the area as far as predicted winner's percentage and absolute margin of error are concerned, there is no point in trying to replicate or outperform their techniques using a similar approach and explanatory variables. For these reasons, the pre-match prediction model demonstrated in this thesis uses a slightly different method by not only including attacking capabilities but also defensive capabilities. It was believed that the interaction between one team's attack and their opponent's defence may provide a reasonable approximation to the attacking score.

This model follows on from the earlier chapter relating to statistical distributions and scoring in the AFL by making the distinction between goals and behinds. For the purposes of this model attacking ratings are broken into goals and behinds and so too are defensive ratings. It was thought that prediction accuracy may be improved if expected score was derived by predicting score as a function of goals and behinds. As a result there are four parameters of interest, namely attacking goals and attacking behinds as well as defensive goals and defensive behinds.

For obvious reasons that have been explored in Chapter 6, these measures of attack and defence need to be differentiated according to whether a team is playing at home or interstate. For the sides from Melbourne that play regularly at the MCG or Docklands, both as the home and away named side, the distinction is made according to venue. To illustrate the difficulties of a location approach to each team, Table 7.1 contains the possible locations of matches for each club in the competition. Data have been used dating back to 1998. Although more data were available and could have been used as done by Bailey, it was decided that this was a sufficiently large data set with all teams having played a minimum of 154 games if they had never made the finals. Furthermore, the present form of the AFL competition as a 16 team national competition began only in 1997, with the introduction of Port Adelaide and dismissal of Fitzroy.

Another reason for only going back to 1998 was the change of the AFL's own ground from Waverley Park to Docklands in 2000. It was felt that, with the inclusion of seven year's worth of matches, a model which uses attack and defence ratings according to location would have enough data points in order to accurately reflect team performance levels. Earlier games were considered irrelevant for present prediction purposes.

Table 7.1: Venues that are used for attack and defence ratings for each club in AFL

Team	Home	Away	MCG	Docklands	Other
Adelaide	Football Park	All Other	-	-	-
Brisbane	Gabba	All Other	-	-	-
Carlton	-	All Other	MCG	Docklands	-
Collingwood	-	All Other	MCG	Docklands	-
Essendon	-	All Other	MCG	Docklands	-
Fremantle	Subiaco	All Other	-	-	-
Geelong	Kardinia Park	All Other	MCG	Docklands	-
Hawthorn	-	All Other	MCG	Docklands	York Park
Melbourne	-	All Other	MCG	Docklands	-
Kangaroos	-	All Other	MCG	Docklands	Manuka Oval
Port Adelaide	Football Park	All Other	-	-	-
Richmond	-	All Other	MCG	Docklands	-
St. Kilda	-	All Other	MCG	Docklands	-
Western Bulldogs	-	All Other	MCG	Docklands	-
West Coast	Subiaco	All Other	-	-	-
Sydney	SCG, Olympic Stadium	All Other	-	-	-

The attack ratings are derived by exponentially smoothing each team's score for home and away games as described by Table 7.1. For instance, Geelong games at Kardinia Park are treated as home games, MCG games and Docklands games are rated separately (due to the result from Chapter 6) and all other games are treated as away games. Exponential smoothing makes allowances for form thereby producing the best team rating for match prediction, as shown by Bailey (Bailey and Clarke, 2004). Note that for each team, attack and defence ratings are calculated independently for home and away matches.

7.3 Development of the model

As demonstrated in Chapter 5, the Poisson distribution provides an adequate approximation for AFL scoring events. For this reason, it was decided to investigate a model that made use of this distribution to try and predict match outcomes. One such technique was based loosely on the American college ice hockey model known as CHODR (Lock, 2000) and whilst the results were adequate, the technique assumed independence between one team's attacking rating and their opposition's defensive rating, and this may not necessarily be the case as investigated in Chapter 4. Furthermore, the predicted scores using this model could often be quite extreme with teams often given a probability of victory close to one. It was hoped that another model could be developed that was a little more conservative. As a result of the possible dependence between each team's ratings for attack and defence and desired conservativeness in predicted victory probabilities, it was decided to investigate a regression model that may better fit the data, allowing for interaction between variables not only between teams but also between goals and behinds. Due to the nature of AFL scoring events, the first model investigated used Poisson regression techniques. This type of approach has been shown to be highly successful for soccer prediction (Dixon and Robinson, 1998, Dixon and Coles, 1997), (Karlis, 2003).

The intentions of the analyses was to model goals and behinds separately for the attacking and defensive team, and combine the predicted values to obtain an expected score for the attacking team. In trying to predict goals, the dependent variable is the number of goals that a team kicked in a match. The independent variables are the team's attacking and defensive rating for goals as well as their opponent's attacking and defensive rating for goals. These values are derived using exponential smoothing and implicitly allow for venue/home advantage as the rating differs according to where a team is playing. Similarly, for behinds the same variables are used except that they pertain to behinds rather than goals. The SAS procedure, REG, was used on matches from 1998 to 2003 and the output for the goals model produced a deviance value of 3260, on 2513 degrees of freedom. The resultant p-value from the Chi-squared distribution is less than 0.0001 indicating that the goal model is not a good fit to the data. Similarly, for the behinds model, a deviance value of 2873, on 2513 degrees of freedom resulted with the p-value again less than 0.0001. Tables 7.1 and 7.2 contain the model fit statistics for the goals and behinds model using Poisson regression. Definitions of the parameters are:

scr_gl_att = smoothed rating for goals attacking team scores
scr_gl_def = smoothed rating for goals attacking team conceded
scr_gl_att = smoothed rating for goals opposition team scores
scr_gl_att = smoothed rating for goals opposition team concedes

These definitions hold for behinds too.

Table 7.2: Model fit statistics for Poisson regression model of goals

Parameter	DF	Estimate	Std Err.	Chi-Sq	P-value
Intercept	1	1.6445	0.109	229.8	<.0001
scr_gl_att	1	0.0427	0.004	138.4	<.0001
scr_gl_def	1	0.0409	0.004	125.3	<.0001
opp_gl_att	1	-0.0077	0.004	4.4	0.037
opp_gl_def	1	-0.0056	0.004	2.3	0.130

Table 7.3: Model fit statistics for Poisson regression model of behinds

Parameter	DF	Estimate	Std Err.	Chi-Sq	P-value
Intercept	1	1.5172	0.151	100.4	<.0001
scr_bh_att	1	0.0556	0.006	94.9	<.0001
scr_bh_def	1	0.0316	0.006	31.4	<.0001
opp_bh_def	1	0.0006	0.006	0.0	0.919
opp_bh_att	1	-0.0080	0.006	1.9	0.167

Unfortunately, a Poisson regression model was not an adequate fit to the data, which seems a little odd given that AFL scoring is well approximated by the Poisson distribution. One reason for this may be the proliferation of scoring in an AFL match. Soccer, where Poisson regression has been used with success, has a relatively low frequency of goals scored. AFL on the other hand has an average score of 14 goals and 12 behinds for each team and as a result, around 26 scoring events for a team, on average, per match.

A negative binomial regression model was then investigated. Again, the model was set up in the same manner and the same variables were used for the analysis. This time the fit was much better with the goals model having a deviance value of 2552, on 2513 degrees of freedom and a p-value of 0.29, whilst the behinds model has a deviance value of 2562, on 2513 degrees of freedom and a p-value of 0.24. These results would seem to indicate that a negative binomial model for predicting AFL matches is more suitable than a Poisson model; however there is very little difference between the parameter estimates of both models. In fact, the only noticeable difference is the difference in magnitude of the intercepts for both models. The Poisson model has a larger intercept for both the goals and behinds model, whereas the parameter estimates for the four ratings used are close to equal. The following tables display the estimated coefficients for goals and behinds models respectively.

Table 7.4: Model fit statistics for negative binomial regression model of goals

Parameter	DF	Estimate	Std Err.	Chi-Sq	P-value
Intercept	1	1.6431	0.123	178.54	<.0001
scr_gl_att	1	0.0427	0.004	106.68	<.0001
scr_gl_def	1	0.0410	0.004	97.52	<.0001
opp_gl_att	1	-0.0077	0.004	3.36	0.067
opp_gl_def	1	-0.0057	0.004	1.81	0.179

Table 7.5: Model fit statistics for negative binomial regression model of behinds

Parameter	DF	Estimate	Std Err.	Chi-Sq	P-value
Intercept	1	1.5165	0.161	89.11	<.0001
scr_bh_att	1	0.0556	0.006	84.37	<.0001
scr_bh_def	1	0.0316	0.006	27.85	<.0001
opp_bh_def	1	0.0005	0.006	0.01	0.929
opp_bh_att	1	-0.0079	0.006	1.68	0.195

In both cases, the opposition's ratings on defence for goals and behinds, are not significant in the model and these parameters should be excluded from the model. These models have included no interactions between ratings for goals and behinds. It has been shown that there is asignificant correlation between AFL scoring events in Chapter 4 and any model that is to be used should include interactions of some kind. Various interactions were tested. The best fitting model for goals was found to be one that included an interaction between the scorings side attacking and defensive rating and a separate interaction between the opposition's attacking and defensive rating. This is a nice result and agrees with the correlations from Chapter 4. The inclusion of these interactions rendered the defence and attack variables by themselves as insignificant. The following table contains the summary statistics for the best-fit goal model, which had a deviance of 2552 on 2515 degrees of freedom. The p-value associated with this model was 0.30, indicating adequate goodness of fit.

Table 7.6: Summary statistics for best-fit negative binomial regression model of goals

Parameter	DF	Estimate	Std Err.	Chi-Sq	P-value
Intercept	1	2.148	0.061	1226.3	<.0001
scr_gl_at*scr_gl_def	1	0.003	0.000	250.31	<.0001
opp_gl_at*opp_gl_def	1	-0.001	0.000	6.34	0.01

It is evident from Table 7.6 that both interactions used in the model are highly significant for predicting goals scored, although the scorer's interaction is of more importance than the defender's interaction. It is necessary to complement this model with one that predicts the number of behinds a team is expected to score. Using the same approach as the goals model and including the same interactions between variables, a model was developed with the summary statistics contained in Table 7.7.

Table 7.7: Summary statistics for best-fit negative binomial regression model of behinds

Parameter	DF	Estimate	Std Err.	Chi-Sq	P-value
Intercept	1	1.904	0.0511	1389.8	<.0001
scr_gl_at*scr_gl_def	1	0.001	0.0002	9.7	0.002
scr_bh_at*scr_bh_def	1	0.003	0.0004	59.8	<.0001

The behinds model had a deviance value of 2193 on 2145 degrees of freedom and an associated p-value of 0.23, indicating that the goodness of fit is adequate using the negative binomial distribution. Interestingly, the interaction between the scorer's ratings for goals is a significant predictor in the behinds model. It has been shown here that by including an interaction for team's goals when modeling their behinds, the prediction is improved.

In conclusion, two models have been developed using negative binomial regression to predict goals and behinds and the goodness of fit of both models is satisfactory. Both models are then combined to obtain a predicted score for the match for each team. These scores are then compared to one another to ascertain a winner and margin of victory. Below is an example of how this works.

Sydney v Hawthorn at S.C.G, round 9, 2004

Sydney: predicted goals = 14.3, predicted behinds = 12.5;

Predicted score = (6*14.3) + 12.5 = 98.3

Hawthorn: predicted goals = 11.7, predicted behinds = 10.1;

Predicted score = (6*11.7) + 10.1 = 80.3

Sydney is the predicted winner by a margin of 18 points

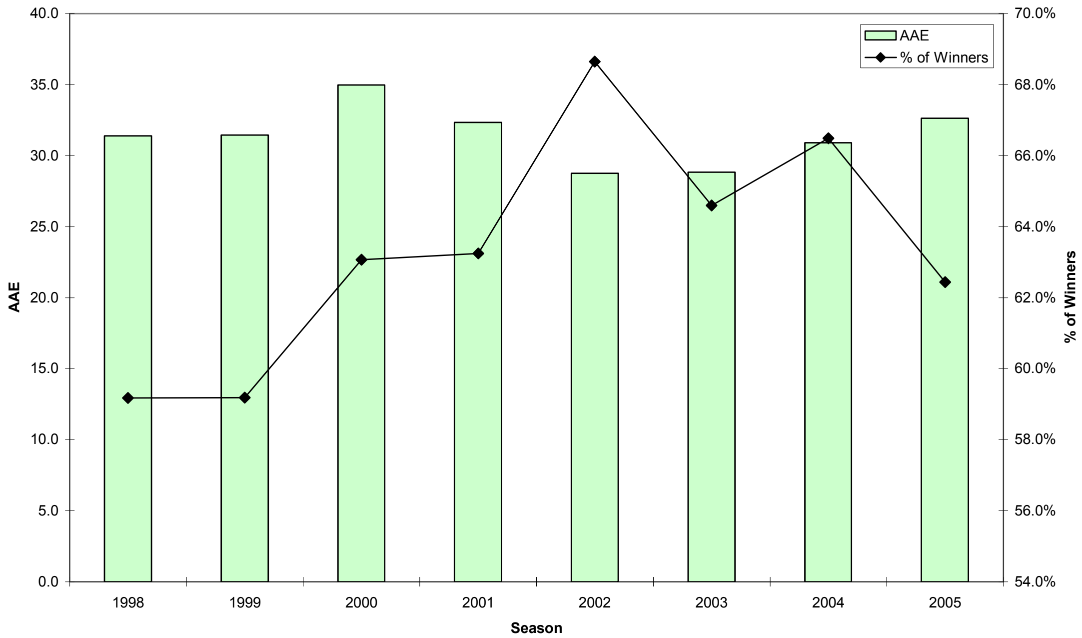
Sydney won the match by one point with a score line of 11 goals, 14 behinds, 80 points to 12 goals, 7 behinds, 79 points.

7.4 Results

The model derived from the data set of games between 1998 and 2003 was applied to a holdout sample of games from 2004 and 2005 to provide a more valid test of predictive capability. The results of the model for these seasons have been included in Figure 7.1 to ascertain its fit. Two indicators are used to measure the success of the model in predicting the results of the AFL competition. Firstly, one is concerned with the % of correct winners predicted. Secondly, the average absolute error (AAE) is also a useful indicator for ascertaining the accuracy of the margin prediction. It is obtained by subtracting the predicted margin of victory from the observed margin and taking the absolute value before averaging.

In the Sydney v Hawthorn example given above, the absolute error for that prediction would be $|1 - 18| = 17$ as the model predicted a margin to Sydney of 18 points and they won by one point. This measure can better reflect the accuracy of the model over the straight percentage of winners. For example, in the match from above, if the model had picked Hawthorn to win by two points and they lose by one, the pick is incorrect, however, the absolute error is only three points. Whereas a tipster who picks Sydney to win by 40 points has got the result right but is 39 points away from the actual margin. Therefore, the AAE can be used as a measure of accuracy of the model. The following figure gives an indication of how the negative binomial model has performed over the years in question.

Figure 7.1: AAE and % of winners by season for pre-match negative binomial regression prediction model



It is expected that the model would improve with time because an exponential smoothing technique is being used to obtain ratings for each team according to where they are playing. This is evidenced in Figure 7.1 with the worst percentage of winners picked in 1998 and 1999. It is also worthwhile noting that 2000, which corresponds with the highest AAE in a season, was the first year that Docklands was used as a venue and the smoothing needed time to adjust the ratings. The last four seasons not including the holdout sample, have shown improvement in both percentage of winners picked and AAE with a high 68.6% of winners picked in 2002 and a low AAE of 28.8 in 2002 and 2003.

The model applied to the holdout sample of 2004 and 2005 has performed well, picking 66.5% and 62.4% of winners respectively. The AAE for these seasons also compares favourably with the seasons used for fitting the model. Further improvement would be expected by including the data from the holdout sample when fitting the negative binomial regression models. It should be pointed out that this model has been utilised by the author for the past two AFL seasons to derive a substantial return on investment in the AFL head to head betting market. In 2004, 103 bets were made for a profit of 30% on turnover, whilst in 2005, 97 bets were placed for a profit of 24% on turnover.

Bailey (Bailey and Clarke, 2004) provided comparison of Bailey's model against Clarke's model and this information will be used to investigate the appropriateness of this approach. His results were obtained for the period 1997-2003 whilst this model has been applied to 1998-2004 data. For the period in question, Bailey's model had an AAE of 30.2 ± 0.6 and a percentage of winners picked of 65.8%. The benchmark model he referred to (Clarke, 1993) had an AAE of 30.5 ± 0.6 (95% C.I.) and a percentage of winners picked of 64.6%. The negative binomial regression model had an AAE of 31.3 ± 1.4 (95% C.I) and a percentage of winners picked of 63.5%. Although this model performs slightly worse than Bailey's and Clarke's, an approach that uses both attack and defence ratings based only on which team is playing and where the match is being played has definite merit. It must also be noted that Bailey's model used a number of predictor variables including ground familiarity, playing personnel and travel distance, and so would be expected to perform better than the minimalistic approach employed in this study. Further, Clarke's modeling takes a more advanced approach to venue rating than is used for this model. In light of these factors the model presented in this chapter appears to provide an adequate tool for pre-match prediction of AFL match results. The conclusion that can be drawn from this comparison is that pre-match modeling reaches a limit in terms of the accuracy and correctness that can be achieved. All three models perform to a similar level of prediction accuracy. The next section will look at whether updating predictions throughout the course of a match has the desired result of improving accuracy in the long run.

7.5 Updating predictions during a match

The purpose of this research was to develop a dynamic prediction model for AFL football that updates throughout the course of a match. This section will investigate whether the pre-match prediction model that has been presented in this Chapter can be improved upon if updates are made to the predicted margin of victory at half time of an AFL match. Intuitively, one would expect that if you had a predicted margin before the match, the knowledge and application of what had taken place in the first half would improve prediction accuracy. The model of the author and Clarke's model were used to investigate whether half time knowledge could improve predictions.

The data used for the analysis included every AFL match between 2001 and 2004, which numbered 740 matches. Previous seasons were excluded due to the fact that the introduction of new venues meant that a prediction was impossible under the author's model for a number of matches, there being no history on which to base the exponential smoothing. A predicted margin for each match was obtained from both models and a prediction for the 2nd half margin by dividing the prediction by two. An alternative prediction of the 2nd half margin is that actual margin divided by two. Correlations between the first half margins and the observed second half margins were then obtained as well as differences between the first half margins (observed and predicted) and the second half observed margin. A priori, we expected the observed 1st half margin to be the better predictor as it took into account player personnel, weather, ground size etc. The results are presented in the following table.

Table 7.8: Correlations and errors of predicted and observed ½ margins

Prediction for 2nd half Margin	Correlation with actual 2nd ½ Margin	P-Val	1st minus 2nd Half Margin		
			Mean	Std Dev	Mean Square Error
Observed	0.32	<0.0001	-0.22	33.08	1094.3
Predicted Forbes ½ Margin	0.35	<0.0001	0.32	26.28	690.7
Predicted Clarke ½ Margin	0.42	<0.0001	0.24	25.73	662.1

The column titled 'Correlation with 2nd ½ Margin' in Table 7.8 shows that the pre-match prediction models outperform the observed margin at half time as a predictor. Both prediction models have a stronger correlation between 1st and 2nd half margins than the observed value. This is an interesting result as it suggests that little improvement could be made to the static model used prior to a match by updating it at half time. The fact that Clarke's model outperforms the model presented above is reflected by the higher correlation and lower mean square error for that model when compared to the author's model.

7.6 Summary

Drawing on the results from Chapters 4 and 5 regarding scoring events in the AFL and their underlying statistical distributions, a model has been developed for pre-match prediction. Whilst a Poisson regression model was not appropriate, it was found that a negative binomial approach using only location and the interaction between attack and defence was suitable for modeling goals and behinds separately to obtain predicted margins for a match. This model, whilst not as accurate as other benchmark models used in the area, performed well and has been used to make substantial profits betting against the bookmakers. A particularly interesting result was the fact that the model could not be improved upon greatly by updating it at half time with what had taken place in the first half. This suggests that in order to dynamically approximate AFL matches a different approach has to be explored. The next Chapter will introduce a Markov process model that can be used to approximate AFL matches in a more dynamic manner.

Chapter 8: An eight state Markov process to globally approximate Australian Rules football

8.1 Introduction

This chapter introduces the first of the Markov process models developed to approximate AFL football in a dynamic environment. AFL football is a continuous sport and the models presented in this thesis have been developed with this in mind. In a continuous time Markov chain the process makes a transition from one state to another, after an interval of time has been spent in the preceding state. This interval is defined as the state holding time. For a discrete time Markov chain the holding time is 1, while in continuous time Markov chains it is exponentially distributed. Although the models in this thesis are not time structured, the model does not contain an absorbing state and only ends when the relevant number of transitions (match, half, quarter or segment of a quarter) have expired.

In a model developed to investigate ice-hockey (Thomas, 2006) the holding times for each state were not exponentially distributed, meaning that a continuous time Markov process had to be replaced in favour of a semi-Markov process. In the research presented in this thesis, the time spent in each state has not been included and therefore, it is assumed that the times spent in each state are exponentially distributed; however, further research may indicate that a semi-Markov process similar to the ice-hockey model is more appropriate.

The Markov model used for association football (Hirotsu, 2002) relied on a first-order Markov process which assumed that the Markov process was valid. Hirotsu noted that further research could investigate the validity of the assumption and the time-dependency of parameters. His work provided an excellent approximation to soccer using a very simplistic approach and would only be improved upon by including time as an element of the model. Fortunately for Hirotsu, he was able to use summary statistics from the

English Premier League yearbook to derive the transition probabilities for his model. The ability to do this gives weight to his use of a first order Markov process model as more complicated relationships in the data did not need to be accounted for.

The assumption of a first-order Markov process model is that the system is memory less, and futures states can be determined by knowing only the current state. This assumption has been made for AFL football when, in reality, this is most likely not the case. The transition probabilities used in the AFL models had to be calculated with reference to the match transaction files and the relationship between match statistics. With this in mind and the nature of AFL football it would be expected that higher order models that take into account chains of play would give a better approximation to AFL football than a first-order model.

There is strong justification for accepting the first order models contained in this thesis for approximating Australian Rules football. As will be shown in the following chapters, the fit of the model is unquestionable with a good approximation provided in both instances. Furthermore, the first order model is preferable from a simplicity point of view. With the introduction of a second order model comes the need for many more states to be included in the model. It is believed that the small gain in accuracy and fit, which may be achieved with higher order models, cannot be justified by the increased analysis and computation that is required. Also, the introduction of higher order models would see a drastic reduction in observed cell counts for an AFL match, reducing the power of transition probabilities. When quarters or segments of quarters are investigated, the available data would be minimal and result in analysis and simulation that may be inaccurate.

8.2 The 8-state model

The development of an eight state model stemmed from an initial seven state model which did not encapsulate the events of a match as accurately as desired. The initial model contained the following states: team A possession, team B possession, ball in dispute, team A goal, team B goal, team A behind and team B behind (Forbes and Clarke, 2004). Although the results for this setup were promising, it was believed that by not including the three different types of stoppages in the game, any practical applications of the model may be limited. With this in mind, a new model was developed that included as separate states, the three stoppage types. With this amendment, it was no longer necessary to include separate states for team A and B goals as this was implicit in the model by the inclusion of centre bounces. In order to model the game of Australian Rules football, the following eight states need to be defined:

State 1: Centre Bounce – this state is entered at the beginning of each quarter and after either team kicks a goal.

State 2: Ball Up Bounce – similar to a centre bounce, however it can take place anywhere on the field during general play.

State 3: Throw In – when the ball leaves the playing arena it is returned into play by the boundary umpire via a throw in.

State 4: Dispute – when neither team has the ball and either is a theoretical 50/50 chance of gaining possession.

State 5: Team A has possession of the ball.

State 6: Team B has possession of the ball.

State 7: Team A behind – Team A scores a behind and ball returned into play by Team B via a kick in.

State 8: Team B behind – Team B scores a behind and ball returned into play by Team A via a kick in.

Table 8.1 contains the definition for each transition from one state to another that is possible within this model.

Table 8.1: Definition of transition probabilities in an AFL game

Transition (states)	Probability	Definition
1 to 2	a_{12}	Centre bounce to a secondary ball up bounce
1 to 4	a_{14}	Centre bounce to dispute
1 to 5	a_{15}	Centre bounce to Team A possession
1 to 6	a_{16}	Centre bounce to Team B possession
2 to 2	a_{22}	Ball up to secondary ball up
2 to 3	a_{23}	Ball up to throw in
2 to 4	a_{24}	Ball up to dispute
2 to 5	a_{25}	Ball up to Team A possession
2 to 6	a_{26}	Ball up to Team B possession
2 to 7	a_{27}	Ball up to Team A behind
2 to 8	a_{28}	Ball up to Team B behind
3 to 2	a_{32}	Throw in to secondary ball up
3 to 3	a_{33}	Throw in to throw in
3 to 4	a_{34}	Throw in to dispute
3 to 5	a_{35}	Throw in to Team A possession
3 to 6	a_{36}	Throw in to Team B possession
3 to 7	a_{37}	Throw in to Team A behind
3 to 8	a_{38}	Throw in to Team B behind
4 to 2	a_{42}	Dispute to ball up
4 to 3	a_{43}	Dispute to throw in
4 to 4	a_{44}	Dispute to dispute
4 to 5	a_{45}	Dispute to Team A possession
4 to 6	a_{46}	Dispute to Team B possession
4 to 7	a_{47}	Dispute to Team A behind
4 to 8	a_{48}	Dispute to Team B behind
5 to 1	a_{51}	Team A kicks a goal
5 to 2	a_{52}	Team A to ball up
5 to 3	a_{53}	Team A to throw in
5 to 4	a_{54}	Team A to dispute
5 to 5	a_{55}	Team A to Team A possession
5 to 6	a_{56}	Team A to Team B possession
5 to 7	a_{57}	Team A kicks a behind

Table 8.1: Definition of transition probabilities in an AFL game (cont.)

Transition (states)	Probability	Definition
6 to 1	a_{61}	Team B kicks a goal
6 to 2	a_{62}	Team B to ball up
6 to 3	a_{63}	Team B to throw in
6 to 4	a_{64}	Team B to dispute
6 to 5	a_{65}	Team B to Team A possession
6 to 6	a_{66}	Team B to Team B possession
6 to 8	a_{68}	Team B kicks a behind
7 to 2	a_{72}	Team B kick in to ball up
7 to 4	a_{74}	Team B kick in to dispute
7 to 5	a_{75}	Team B kick in to Team A possession
7 to 6	a_{76}	Team B kick in to Team B possession
8 to 2	a_{82}	Team A kick in to ball up
8 to 4	a_{84}	Team A kick in to dispute
8 to 5	a_{85}	Team A kick in to Team A possession
8 to 6	a_{86}	Team A kick in to Team B possession

Hirotsu derived the data used in his model from the Carling Opta Football Yearbook (Hirotsu, 2002). The statistics recorded in this book are very well defined for the purposes of his model and required little interpretation as to which transition they may constitute. This is not the case when it comes to AFL match statistics as recorded by CD. Over 80 different match occurrences can be recorded by CD for any one game of AFL football. The model uses only 30 of the 84 statistics to assign transition probabilities between each state. The statistics had to be coded according to what transition they constituted. This was done using CD's and the AFL's accepted event definitions. For example, a short kick is defined as a kick of less than 40m that finds a team-mate and as such guarantees the ball stays with the team in possession. The watching of matches off tape also assisted in best approximating the events of play with the proposed model. Table 8.2 contains the 30 events used in the model as recorded by CD.

Table 8.2: Statistics used from an AFL match to derive Markov transition probabilities

Stat Code	Description	Stat Code	Description
BEHI	Behind	KIIN	Ineffective kick in
BUBO	Ball up bounce	KILO	Long kick in
EQTR	End of quarter	KISE	Kick in self
FRFO	Free kick for	KISH	Short kick in
GATH	Gather	KKCL	Clanger kick
GEHA	Hard ball get	KKGK	Ground Kick
GELO	Loose ball get	KKIN	Ineffective kick
GERU	Ruck hard get	KKLO	Long kick
GOAL	Goal	KKSH	Short kick
HBCL	Clanger handball	MACO	Contested mark
HBEF	Effective handball	MAER	Earned mark
HBIN	Ineffective handball	MAUN	Uncontested mark
HBRE	Handball received	RUSH	Rushed behind
KIBU	Kick in ball up	SQTR	Start of quarter
KICL	Clanger kick in	THIN	Throw in

There are several reasons why the remaining statistics were not used in the model. Firstly, only count data was used. This ensured categorical variables were omitted e.g. inside 50 and rebound 50, interchange on or off. These variables added no numerical value to the model. Secondly, some variables offer no evidence of what has taken place, as far as possession and scoring goes, within the game. Examples of these are free kicks where advantage is played, bounces and tackles. Finally, not all of the statistics are mutually exclusive and may be recorded twice. The statistical package, SAS 8.01, was used to transform the raw data into a form that constituted individual transition matrices for each game of the 2003 and 2004 seasons. In order to do this redundant statistics had to be removed from the analysis. In certain instances some statistics will be included in other codes as well as their own. This happens with derived statistics such as a long kick to advantage, which will also be included as a long kick. For instance, in a match if Team A had 30 long kicks and 10 long kicks to advantage, the system would record them as having had 40 long kicks. The same issue arose with goals and the kick that resulted in the goal. The data will record each goal scoring kick within the kick code as well as recording the goal. These doubled up occurrences had to be eliminated so that what happened in the game was reflected as accurately as possible by the numbers used for transition probabilities. Furthermore, a transition that is defined by the characteristic of the event taking place (KKSH guarantees possession) did not need to have the associated possession gather included as well. Watching games off tape, accompanied by the transaction files allowed for decisions to be made on what to include in the analysis and what to leave out.

Of the 49 transition probabilities, 23 have zero probability, as the associated transition cannot occur. An example would be team A having the ball and team B kicking a goal. The remaining 26 transition probabilities need to be calculated using counts of the data extracted from each match and this will be demonstrated later in this chapter. The 26 transitions and the relevant match statistics that comprise them are given below. Following these summaries for each state is a table that contains the code for each statistic and a description of what it constitutes.

- *CEBO → BUBO: There is no possession for either side after a CEBO and a BUBO results immediately.*

- *CEBO → Disputed Possession: Occurs after a CEBO when either team kicks the ball off the ground (KKGK) without physically taking possession.*

- *CEBO → Team Possession: Team A (or B) gains the next possession (GATH, GEHA, GELO, GERU, FRFO) following the CEBO.*

In an initial model the scoring of a goal meant a reversion to state 3 with probability 1. However, projections are more accurate when the resulting possession after the goal is attributed to the relevant team, as evidenced by the first possession after a centre bounce.

- *BUBO → BUBO: There is no possession for either team after a BUBO with another BUBO following immediately.*

- *BUBO → THIN: There is no possession for either team after a BUBO and the ball goes out of bounds resulting in a THIN.*

- *BUBO → Disputed Possession: A ground kick (KKGK) is the first statistic that occurs after a BUBO.*

- *BUBO → Team Possession: Either team takes possession of the ball straight after a BUBO similar to first possession from a CEBO.*

- *BUBO → Behind: The ball is forced through for a behind directly from a BUBO.*

- *THIN → BUBO: A BUBO results directly from the ball being returned from out of bounds without a statistic in between.*

- *THIN → THIN: The ball is forced out of play directly from a THIN without a statistic in between.*

- *THIN → Disputed Possession: A ground kick (KKGK) is the first statistic after the ball is returned from out of bounds.*

- *THIN → Possession: Similar to CEBO/BUBO → Possession, either team gains the first possession after a THIN.*

- *THIN → Behind: The ball is 'rushed' through for a behind directly from a THIN.*

- *Disputed Possession → BUBO: The ball has become disputed via a disposal that does not guarantee the team retains possession (HBIN, KKIN, KKLO, KKGK) and a BUBO is the next transition.*

- *Disputed Possession → THIN: Similar to the previous transition, however a THIN is the next transition.*

- *Disputed Possession → Disputed Possession: Only occurs when the ball is in dispute and either team advances it via a KKGK..*

- *Disputed Possession → Possession: The ball is in dispute and either team gains possession of it out of dispute (GEHA, GELO, GATH, MACO, MAER, FRFO).*

- *Disputed Possession → Behind: The ball is rushed through for a behind when neither team has possession of it.*

- *Possession → CEBO: A goal is kicked by the team in possession and the ball returns to the centre for a CEBO.*

- *Possession → BUBO: A BUBO stems directly from either team having possession of the ball. Usually when a tackle is made and there is no chance to dispose of the ball.*

- *Possession → THIN: Similar to above however a THIN results.*

- *Possession → Disputed Possession: When either team disposes of the ball in a manner that does not guarantee they retain possession (HBIN, KKIN, KKLO, KKGK) and theoretically makes the ball available to either team.*

- *Possession → Team Possession: Team A (or B) has possession and the disposal ensures they retain possession (KKSH, HBEF, KKLA).*

The definition of these disposal types guarantees that the team that has the ball retains it, either via the foot or hand, for the next play. The redundant statistic that needs removal is the possession gather after the disposal i.e. the mark or ball get as it has already been accounted for due to the definition associated with the disposal type.

- *Possession → Opposition Possession: Clanger disposals (HBCL, KKCL) guarantee the opposition has possession of the ball.*

- *Possession → Team Behind: The team in possession of the ball kicks a behind.*

- *Possession → Opposition Behind: The team in possession of the ball 'rushes' a behind for their opposition.*

- *Behind → BUBO: The player kicking in steps over the goal square resulting in a KIBU and a ball up.*

- *Behind → Disputed Possession: The player kicking in puts the ball into dispute from his kick in (KIIN, KILO).*

- *Behind → Team Possession: Team A (or B) scores a point and Team B (or A) gains the next possession (KILA, KISH, KISE).*

  Similar to when a goal is scored, the model was more accurate when probabilities were assigned to transitions based on the kick-in statistics for the match rather than assuming the non-scoring side gained possession with probability 1. The kick-in codes above result in the team that kicks the ball back into play retaining the ball.

- *Behind → Opposition Possession: A clanger kick in (KICL) by the player kicking in ensures possession goes directly to the opposition from the kick in.*

Table 8.3 comprises the statistic codes, a description of the code and the transition that they comprise. This gives a better understanding of the mechanics of the model.

Table 8.3: Statistic codes, descriptions and transition for match occurrences contained in model

Stat. Code	Description	Transition
BEHI	Behind (1 point)	POSS → BEHI
GATH	Gather of Loose Ball	DISP → POSS
GEHA	Hard ball get	DISP → POSS
GELO	Loose ball get	DISP → POSS
GERU	Gather from a ruck	DISP → POSS
GOAL	Goal (6 points)	POSS → CEBO
HBCL	Clanger Handball	POSS → OPPOSITION
HBEF	Effective Handball	POSS → POSS
HBIN	Ineffective Handball	POSS → DISP
KIBU	Kick in resulting in a ball-up	BEHI → BUBO
KICL	Clanger Kick-in	BEHI → OPPOSITION
KIIN	Ineffective Kick-in	BEHI → DISP
KILA	Long Kick-in to advantage	BEHI → POSS
KILO	Long Kick-in	BEHI → DISP
KISE	Kick-in to self	BEHI → POSS
KISH	Short Kick-in	BEHI → POSS
KKCL	Clanger Kick	POSS → OPPOSITION
KKGK	Ground Kick	DISP → DISP
KKIN	Ineffective Kick	POSS → DISP
KKLA	Long Kick to advantage	POSS → POSS
KKLO	Long Kick	POSS → DISP
KKSH	Short Kick	POSS → POSS
MACO	Contested Mark	DISP → POSS
MAUN	Uncontested Mark	DISP → POSS
RUSH	Rushed Behind (1 point)	DISP → BEHI

8.3 Calculation of match transition probabilities

Having elicited the count data for a match and allocating it to the relevant transition, the calculation of transition probabilities is very simple and can be well explained with reference to a transition matrix for the model contained in Table 8.4. The cells that contain a zero are impossible transitions as referred to above.

Table 8.4: AFL Markov process model transition matrix

State/Description		1 CEBO	2 BUBO	3 THIN	4 DISP	5 APOS	6 BPOS	7 ABEH	8 BBEH	Row Total
1	**CEBO**	0	sum_{12}	0	sum_{14}	sum_{15}	sum_{16}	0	0	**$CEBO_{TOT}$**
2	**BUBO**	0	sum_{22}	sum_{23}	sum_{24}	sum_{25}	sum_{26}	sum_{27}	sum_{28}	**$BUBO_{TOT}$**
3	**THIN**	0	sum_{32}	sum_{33}	sum_{34}	sum_{35}	sum_{36}	sum_{37}	sum_{38}	**$THIN_{TOT}$**
4	**DISP**	0	sum_{42}	sum_{43}	sum_{44}	sum_{45}	sum_{46}	sum_{47}	sum_{48}	**$DISP_{TOT}$**
5	**APOS**	sum_{51}	sum_{52}	sum_{53}	sum_{54}	sum_{55}	sum_{56}	sum_{57}	0	**$APOS_{TOT}$**
6	**BPOS**	sum_{61}	sum_{62}	sum_{63}	sum_{64}	sum_{65}	sum_{66}	0	sum_{68}	**$BPOS_{TOT}$**
7	**ABEH**	0	sum_{72}	0	sum_{74}	sum_{75}	sum_{76}	0	0	**$ABEH_{TOT}$**
8	**BBEH**	0	sum_{82}	0	sum_{84}	sum_{85}	sum_{86}	0	0	**$BBEH_{TOT}$**

Each cell's transition probability is calculated simply by dividing the cell count by the row total and it is evident then that the row probabilities sum to one. So for instance, the probability, a_{55}, that team A has the ball and retains it directly is computed by:

$$a_{55} = Sum_{55}/APOS_{TOT}$$

If, for example, team A has the ball 300 times and on 120 occasions they directly retain the ball, the probability a_{55} = 120/300 = 0.4. The row probabilities provide a profile of how the ball has been distributed from any state during an AFL match and can be accumulated as shown in the next Chapter to gain a better understanding of how particular teams play the game.

8.4 Ascertaining fit of the model

Once the transition matrices for each game from the 2003 and 2004 season (370 games) have been calculated, the long-term behaviour of each matrix is of interest. As mentioned in the introduction for this chapter, a first order Markov process has been assumed, in which, the transition probabilities are independent of the previous sequence of events. Whilst this may not necessarily be true, future research can be used to validate this assumption. The first-order Markov process from each match has an associated limiting probability distribution:

$\pi = (\pi_0, \pi_1, ...,\pi_8)'$, where $\pi j > 0$ for $j = 0, 1, ...,8$ and $\sum j \pi j = 1$

This convergence means that in the long run, the probability of finding the Markov chain in state j is approximately π_j no matter in which state the chain began at time 0. Relying on this knowledge, the limiting distributions for each match were derived by raising the transition matrix until convergence resulted in SAS. We can also think of π_j as the proportion of time in each match that the chain spends in state *j* if each transition takes the same time, which is not the case in this model. Alternatively, they are the probabilities of being in state *j* after any transition. The limiting distribution can then be used to calculate the proportion of match transitions that the model spends in each state. These expected counts for each state can then be compared to the observed data as calculated previously to obtain errors for each state.

Before the observed and expected counts can be compared, adjustments need to be made to the observed count data to allow for the last transition in each quarter as well as the four CEBO's that start each quarter. Each statistic that takes place immediately before the end of quarter siren has to be removed from the observed count as the ball enters the relevant state but has no chance to leave it and therefore could affect the errors when observed and expected data are compared. Similarly, four CEBO's need to be removed from the observed data matrix so that only the CEBO's that occur as a result of a goal to either side are left for comparison. These adjustments more accurately reflect the observed data and the results can be seen in Table 8.5, which contains the mean error for each state and a 95% c.i. for the errors, as well as the mean square error.

Table 8.5: Mean error for estimated counts for each state and 95% c.i. for mean error assuming a normal distribution

State	Mean Error	Std Dev	Lower Bound	Upper Bound	Mean Square Error
CEBO	-0.12	0.08	-0.13	-0.11	0.02
BUBO	0.49	0.35	0.45	0.53	0.36
THIN	-0.08	0.10	-0.09	-0.07	0.02
DISP	-0.35	0.49	-0.40	-0.30	0.37
APOS	0.27	4.05	-0.15	0.68	16.48
BPOS	-0.11	4.05	-0.53	0.30	16.38
ABEH	-0.04	0.17	-0.06	-0.02	0.03
BBEH	-0.05	0.16	-0.07	-0.04	0.03

It can be seen from Table 8.5 that for the 370 matches the mean error for each of the eight states is relatively small with none exceeding an absolute value of one. It must be noted that for only two of the states (APOS, BPOS) does the 95% c.i. for the mean error contain zero. However, during the transition probability computation process, a small number of adjustments and edits had to be made to the transaction files recorded by CD. Potential errors in the data that led to relationships between statistics that were not allowed for in the probability derivation program may affect the overall errors.

The distribution of errors may not necessarily be normal so a chi-squared goodness of fit test with 2959 degrees of freedom has been performed on the observed and expected counts for the 370 games in the data set. The results are very impressive (X^2 = 49.38, d.f. = 2959, p-val = 1.00). In addition, each match was looked at individually using a chi-squared test with seven degrees of freedom to test the significance of the chi-squared residuals. Again, the results of these tests were impressive with a lowest p-value for the 370 matches of 0.97. Chi-squared tests are typically used to test goodness of fit. With large sample sizes, these tests tend to be over sensitive, favouring lack of fit hypothesis. It was therefore surprising that the above results were so supportive of the model, however, it should be noted that data is being used from completed matches to reproduce information about those matches. The results of the Chi-squared goodness of fit tests provide strong evidence to suggest that the Markov process model that has been implemented for approximating AFL football does an excellent job. These results coupled with the mean errors contained in Table 8.5 suggest that the model is satisfactory for approximating Australian Rules football.

8.5 Summary

An eight state Markov process model is a more than adequate model for approximating Australian Rules football. Even though such a model is clearly more complicated than Hirotsu's four-state model for soccer, it remains a fairly simple approximation in view of the complexity of the game. The eight states included in the model encapsulate the different states that a match can be in and provide the added advantage of capturing a team's performance characteristics in the main areas of a match.

The key behind the success of the model has been the richness in detail of the information that has been collected by CD. Without their data, the implementation of such a model would have been virtually impossible as it has been the ability to link match statistics as they take place that has allowed for relatively simple and accurate calculations of transition probabilities. There is no reason why the techniques used here to model AFL football could not be applied to other continuous sports such as rugby union or league with similar success. The key ingredient would be a level of information that is similar to CD's AFL databases.

This initial model, described as a global approach, pays no regard to location on the ground. More accurate approximations could be made if location was included in the model and better reflections of team performance across different areas of the ground would result. This step has been taken and the model that results from this adjustment will be presented in a later chapter. Regardless of the inclusion of location, this model has some very interesting applications to the game of AFL and a number of these applications will be discussed in detail in the following chapters.

Chapter 9: Analysing matches after their completion using a global Markov process model

9.1 Introduction

The eight-state Markov process model for approximating Australian Rules football presented in Chapter 8 is an innovative tool that can be used for analysis of AFL matches after they have been completed. The application of such a model would be useful for not only AFL teams but also for media outlets and the public in general. This chapter will be devoted to outlining some of the applications where such a model could be used, using examples to demonstrate these applications in practice. Hirotsu used his soccer model to derive optimum times for strategic decisions such as when to make substitutions in a match or advantageous times for fouling an opponent. These types of strategic applications are similar to Clarke and Norman's analysis on when to rush a behind in an AFL match in order to further one's own chances of winning the game. These applications are implemented post match using data derived or deduced from a completed game.

Similarly, this chapter is concerned with applications of the model after matches have been completed. For instance, it is common practice in AFL circles to hear phrases such as, 'we were robbed' or 'that cost us the game'. Previously such statements have not been able to be quantified numerically as no one has had the tools to validate exactly how much one simple mistake or good piece of play means in the overall scale of victory or defeat. The global model is a way of doing this using the actual match transition probabilities and simulation. Furthermore, a transition matrix that has been derived from a completed match could be altered in selected ways to reflect a different game plan or improved ball use with the chances of victory and expected score calculated under the changed conditions. These scenarios are tested using a simulation tool that will be described in the next section.

Simulation and math modeling are widely used tools in the research of sporting outcomes and have been used as far back as the 1970s to answer strategic issues. Such issues include the optimum batting line-up a team should take into a baseball match in order to maximise the runs they are expected to score (Bukiet, Harold and Palacious, 1997, Freeze, 1974). Transition matrices were used to compute run distributions via math modeling with the final results similar to those from simulations. Other researchers use simulation to derive projected ladders for sporting competition based on calculated ratings. The AFL ladder at the end of the home and away season is a perfect example of this and the projected outcomes are considered good enough to be sold as a benchmark for betting purposes (Clarke, 1993). The literature is full of examples of simulation being used in a sporting context with games such as Cricket (Dyte, 1998, Croucher, 2000), Soccer (Koning, Koolhaas, Renes and Ridder, 2003, Dobson and Goddard, 2003, Dyte and Clarke, 2000), Rugby League (Lee, 1999) and Beach Volleyball (Glasson, Jeremiejczyk and Clarke, 2001).

9.2 Simulation process

Simulating using Markov models is not a new phenomenon and in fact off the shelf packages are available for doing exactly this. The development of the global model described in Chapter 8 is a ready made tool for AFL analysis using simulation. Different scenarios can be investigated using the actual probabilities from matches, seasons, and clubs etc to investigate playing strategies and relationships, as well as making adjustments to the probabilities to investigate the effect. In order to do this a complex program was developed using SAS.

The underlying idea of the program is to generate random numbers that determine which state the model will move into next. i.e. if the ball (model) is at a centre bounce, the only states it can move into are a ball-up bounce, dispute, Team A possession or Team B possession. The probabilities for each of these happening may be 0.2, 0.1, 0.4 and 0.3 respectively. The system generated random number, r, dictates the next state as follows:

If $r < 0.2$, state_{i+1} = ball-up bounce;
If $0.2 \le r < 0.3$, state_{i+1} = dispute;
If $0.3 \le r < 0.7$, state_{i+1} = Team A possession;
If $0.7 \le r$, state_{i+1} = Team B possession;

The process has been broken down into four parts to reflect more accurately an AFL match, which is divided into four quarters. The starting state for each quarter is state one (centre bounce). In order to get an accurate reflection of a match, the average number of transitions for a quarter is used. This is done by taking the total number of transitions from a match and dividing this by four. The simulation is run, in general,10,000 times but there is no restriction, save CPU processing time, to the number of times it can be run. To give an example of speed, on a notebook with 256MB of RAM, a season comprising 185 matches, can have each match simulated 10,000 times in roughly 20 minutes.

Upon running a simulation of a match, or set of matches, there is a variety of information that can be gleaned from the process. Perhaps the most important and useful data are the projected score of each team for each simulated match. Goals are calculated for each team by counting the number of times either Team A or B had possession directly before a centre bounce. These occurrences are then multiplied by six and added to the number of times the model entered the behind state for each team to come up with a match score, which is then used to ascertain the match outcome. The probability of victory for each team and the likelihood of a draw is then a simple calculation according to the expected outcome. Other information that can be extracted from a simulation is the proportion of time the 'match' is likely to spend in any one state and the number of occurrences of each state within a simulated match. This will be highlighted in the first application that will be discussed regarding the effect of rule changes.

The simulation process was validated by the results given under different scenarios. These scenarios will be referred to later in this thesis where they arise. However, they will be addressed here too as they indicate that the simulation process is accurate and valid. Home advantage in the AFL competition between 1998 and 2003 was found in Chapter 6 to amount to 12.3 points. When the 2004 season was simulated using a transition matrix for each match, the average margin in favour of the home side amounted to 11.8 points. A matrix was derived for the 2005 season based on the 185 matches and used to simulate an ‘average’ match. The score line that resulted from the simulation was 99-90 in Team A’s favour. Using the actual data from the 2005 season, the average score was 102-87, which again is a favourable comparison to the simulated results.

9.3 Examining the effect of rule changes within the AFL

Traditionally, the AFL/VFL has not seen the necessity to tinker with their rules and as such the game is quite static in its appearance. This is quite a different approach to other organised sporting competitions in Australia, such as the National Rugby League competition, which makes rule changes regularly, presumably, without any knowledge of the effect the rules changes may have on the game. This can be a very dangerous approach as rules that seem meritorious in the boardroom may not actually be so on the playing field. The AFL has had the luxury in recent years of being able to trial rule changes[2] in their pre-season competition and gauge their effects and so determine whether they would be suitable in the home and away competition. The simulation model could be used by the authorities to provide mathematical evidence as to the effect of rule changes on the game. Two changes will be looked at as examples, one taken from other sports such as soccer and basketball and the other a left field idea from one of the competition’s leading coaches.

9.3.1 A different approach to secondary ball-up bounces

Much has been made in recent seasons regarding the proliferation of secondary stoppages in AFL matches. Officials, fans and the media alike have all described the lack of clearance at stoppages, particularly at ball-up bounces, as a blight on the game. The issue

[2] e.g. 9 points for a goal kicked from outside 50m, 3 points for a rushed behind, no delay to kick-in

further came to a head when the chief executive of the AFL described one of the clubs in the competition as playing ugly football (Lonergan, 2005) due to the fact they are happy to bottle the ball up and turn the game into a dour scrap. Furthermore, the modern day phenomenon of 'flooding' (putting large numbers behind the ball in an attempt to slow attacking movements forward) has exacerbated the problem.

Rumblings have emanated from AFL headquarters about ways in which to make the game more attractive and reduce the number of secondary stoppages. The debate about this issue reached a crescendo prior to the start of the season and it was at this time that legendary AFL coach, Kevin Sheedy, went public with his plan to reduce the abundance of secondary stoppages. In a newspaper article (Ryan, 2005), Sheedy "conceived a radical plan to eliminate the unsightly blemish of consecutive stoppages from the game: allow the controlling umpire to kick or throw the ball into the corridor to open up play". He acknowledged in the article he would have to "put a little more thought into it". However the AFL would be loath to implement such a radical rule change without having some kind of evidence as to its effect within a match and this is where a Markov process model could be utilised.

The 185 matches from the 2004 season were used to ascertain whether Sheedy's plan would result in a 'better'[3] game. Firstly, the transition matrices as derived from each match were used in the simulation program. These matches were simulated 10,000 times and the number of ball-up bounces as well as total match points was recorded. The following table contains the summary statistics for each of the eight states in the model derived from the 1.85 million simulated matches.

---

[3] The assumptions made as to what constitutes a 'better' game will be explained in the analysis

Table 9.1: Summary statistics of eight states from a simulated 2004 average match

Variable	Mean	Std Dev
BUBO	28.7	12.1
THIN	29.7	10.1
DISP	156.6	20.0
APOS	294.2	32.8
BPOS	284.6	32.8
ABEH	11.7	5.0
BBEH	11.3	5.2
GOALA	14.6	6.0
GOALB	12.7	5.6

The state CEBO has been broken into GOALA and GOALB to indicate which team kicked the goal leading to the centre bounce. A combination of these variables would give the relevant statistics for CEBO, not including the four centre bounces needed to start each quarter. Having looked at the matches as they actually happened, the next step was to make adjustments to each match's transition matrix to reflect Sheedy's idea of allowing the "umpire to kick or throw the ball into the corridor to open up play".

He was mainly referring to secondary ball ups, so this analysis has only concentrated on them. As a result, adjustments were only made to the three stoppage states of CEBO, BUBO and THIN. To reflect Sheedy's idea the assumption was made that instead of the ball being bounced by the umpire it would return to either team with probability 0.5 as this is the theoretical probability of either side gaining possession from the ball-up bounce. For instance, if four throw ins resulted in secondary ball-up bounces, Team A and B would each be attributed with two extra possessions from a throw in.

To illustrate this in practice, a match from 2004 is presented below. Table 9.2 contains the actual count data with Table 9.3 displaying the transition matrix. Table 9.4 contains the adjusted count data with Table 9.5 displaying the amended transition matrix.

Table 9.2: Observed count data

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH	Total
CEBO	0	9	0	1	5	7	0	0	22
BUBO	0	18	5	0	17	23	0	0	63
THIN	0	9	1	0	9	10	0	0	29
DISP	0	24	22	3	49	32	0	0	130
APOS	11	0	0	62	140	26	11	0	250
BPOS	9	3	1	63	18	136	0	8	238
ABEH	0	0	0	1	2	7	0	0	10
BBEH	0	0	0	0	8	0	0	0	8

Table 9.3: Observed transition matrix

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH
CEBO	0.0000	0.4091	0.0000	0.0455	0.2273	0.3182	0.0000	0.0000
BUBO	0.0000	0.2857	0.0794	0.0000	0.2698	0.3651	0.0000	0.0000
THIN	0.0000	0.3103	0.0345	0.0000	0.3103	0.3448	0.0000	0.0000
DISP	0.0000	0.1846	0.1692	0.0231	0.3769	0.2462	0.0000	0.0000
APOS	0.0440	0.0000	0.0000	0.2480	0.5600	0.1040	0.0440	0.0000
BPOS	0.0378	0.0126	0.0042	0.2647	0.0756	0.5714	0.0000	0.0336
ABEH	0.0000	0.0000	0.0000	0.1000	0.2000	0.7000	0.0000	0.0000
BBEH	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000

Table 9.4: Adjusted count data

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH	Total
CEBO	0	0	0	1	10	12	0	0	22
BUBO	0	0	5	0	26	32	0	0	63
THIN	0	0	1	0	14	15	0	0	29
DISP	0	24	22	3	49	32	0	0	130
APOS	11	0	0	62	140	26	11	0	250
BPOS	9	3	1	63	18	136	0	8	238
ABEH	0	0	0	1	2	7	0	0	10
BBEH	0	0	0	0	8	0	0	0	8

Table 9.5: Adjusted transition matrix

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH
CEBO	0.0000	0.0000	0.0000	0.0455	0.4318	0.5227	0.0000	0.0000
BUBO	0.0000	0.0000	0.0794	0.0000	0.4127	0.5079	0.0000	0.0000
THIN	0.0000	0.0000	0.0345	0.0000	0.4655	0.5000	0.0000	0.0000
DISP	0.0000	0.1846	0.1692	0.0231	0.3769	0.2462	0.0000	0.0000
APOS	0.0440	0.0000	0.0000	0.2480	0.5600	0.1040	0.0440	0.0000
BPOS	0.0378	0.0126	0.0042	0.2647	0.0756	0.5714	0.0000	0.0336
ABEH	0.0000	0.0000	0.0000	0.1000	0.2000	0.7000	0.0000	0.0000
BBEH	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000

This match contained a season high 63 ball-up bounces, 36 of which were secondary bounces. From Table 9.2 it can be seen that nine resulted from centre bounces, 18 from ball-up bounces and nine from throw ins. Table 9.3 shows that a ball-up was likely to result 41% of the time from a centre bounce, 29% of the time from a ball-up and 31% of the time from a throw in. With the adjustments made in Table 9.4, it is seen that Team A and B are credited with an extra 4.5 possessions from a centre bounce and throw in and nine possessions from a ball-up. Having made these adjustments to each of the 185 games for 2004, the simulation was run again using the adjusted probabilities. Table 9.6 contains the summary statistics for each state from the adjusted simulation.

Table 9.6: Summary statistics of eight states from a simulated 2004 average match with secondary ball-up bounces removed

Variable	Mean	Std Dev
BUBO	16.0	6.7
THIN	29.7	10.0
DISP	158.8	21.3
APOS	298.8	32.5
BPOS	289.6	32.6
ABEH	11.9	5.1
BBEH	11.5	5.2
GOALA	14.8	6.1
GOALB	12.9	5.7

The issue has to be addressed now of what constitutes a better game. Obviously, in a time where secondary stoppages are a concern, any rule change would want to see the number of stoppages, particularly ball-up bounces reduced. Furthermore, a more attractive game for spectators may be where more scoring happens. It is logical that paying spectators would prefer to see a high scoring, free flowing affair rather than a dour struggle with not much scoreboard action. Therefore, the assumptions have been made that a 'better' game, which would justify rule changes, involves a reduction in stoppages, particularly ball-ups, as well as an increase in points scored. In comparing Tables 9.6 and 9.1, it can be seen that there is a reduction in ball-up bounces, but we would expect this as we are removing any secondary ball-ups through the umpire putting the ball into play if one occurs. Aside, from that, there is a slight increase in the number of times the ball enters DISP and possessions for Team A and B but more importantly there is a slight increase in scores for both teams.

To test whether these differences are significant, paired t-tests were carried out on the observed difference between the values for the states BUBO, THIN, DISP, APOS and BPOS as well as the total number of points scored in the match from the two scenarios. The following table contains the summary results for each comparison using the mean results from the 185 matches.

Table 9.7: Paired comparison t-test results between observed simulation and no secondary ball-up bounce simulation for the 2004 season

Difference	DF	Mean	Std Dev	t Value	Pr > |t|
BUBO	184	-12.7	6.3	-27.59	<.0001
THIN	184	-0.0	0.5	-0.94	0.347
DISP	184	2.3	1.4	22.74	<.0001
APOS	184	4.6	3.0	20.69	<.0001
BPOS	184	5.1	3.0	23.36	<.0001
SCORE	184	3.0	1.6	24.83	<.0001

Whilst all of the comparisons except for throw-ins (THIN) are statistically significant, the enormous size of the data set implies that a slight change would indicate significance. There is a decrease in ball-up bounces and throw-ins with an increase in disputed ball, possessions for both sides and total match score. In practical terms, the increase in score amounts to an extra three points, on average, per match and probably wouldn't be enough to justify such a radical change to the rules. Furthermore, the assumption that either team gets the ball from the umpire's return into play with probability half, may not accurately reflect reality. The putting of the ball into open space may encourage secondary bounces by allowing both teams to reach the ball simultaneously and scrimmage for possession. In light of this analysis, it would seem unlikely that such a rule change would have the desired effect of improving the game by reducing stoppages and increasing scoring.

9.3.2 Adopting a possession return over the throw-in when the ball goes out of play

The rules of AFL football are different to many other sports when it comes to the return of the ball into play after it has gone out. The boundary umpire is required to throw the ball in over his head whilst facing out from the arena in an attempt to make the ball available to either team with an element of randomness. In other sports the team who took the ball out of play are usually penalised by giving possession (or the best chance at possession in Rugby Union and League) to the opposition. Analysis has been done to investigate the effect on AFL of totally removing the throw in from play and replacing it with the ball being returned by the team that did not touch the ball last.

This analysis follows on from the first part of 9.3.1 with a different adjustment relating to throw-ins. In most cases, throw-ins result from the ball being in dispute. For this scenario, along with throw-ins stemming from ball-up bounces, an assumption has been made that either team would return the ball into play half the time. For the times when a throw-in results directly from a team possession, the opposition is given the possession for returning the ball into play. An occurrence that occurs often is the transition for a secondary throw in (THIN → THIN) and to reflect this in the adjusted probabilities, it has been treated as though the team returning the ball into play has turned possession over directly to their opponents. It is assumed that when this occurs, either side does it half of the time.

Table 9.8: Summary statistics of states from a simulated 2004 season with throw-ins removed

Variable	Mean	Std Dev
BUBO	24.9	10.8
THIN	0.0	0.0
DISP	162.3	22.6
APOS	307.4	32.1
BPOS	297.1	32.1
ABEH	12.2	5.2
BBEH	11.7	5.3
AGOAL	15.2	6.1
BGOAL	13.2	5.8

It can be seen from Table 9.8 that apart from the expected number of ball-up bounces, the mean entries into the other seven states is very similar to what was obtained from Table 9.1 for the simulation done using actual probabilities. Perhaps this is a legacy of the fact that there are fewer transitions with throw-ins removed and therefore the chance of extra entries into any one state is slightly diminished. Paired t-tests were again carried out using the data from the actual simulation and the simulation with throw-ins removed and these are displayed in Table 9.9.

Table 9.9: Paired comparison t-test results between observed simulation and no throw-in simulation

Difference	DF	Mean	Std Dev	t Value	Pr > |t|
BUBO	184	-3.7	2.7	-18.6	<.0001
THIN	184	-29.7	8.6	-46.9	<.0001
DISP	184	5.7	2.4	31.9	<.0001
APOS	184	13.3	6.3	28.6	<.0001
BPOS	184	12.5	5.9	29.0	<.0001
SCORE	184	7.4	2.1	47.2	<.0001

Again, it is seen that all of the differences are significant but this is due to the size of the data set. In reality, the only difference that would seem practically significant would be the reduction in ball-up bounces of just under four per game. An extra goal would be scored, on average, per match; however, the change to the nature of the game would probably not justify these results as significant enough to change the rules.

9.3.3 Summary

It has been shown that the Markov model presented in Chapter 8 can be utilised to investigate the effect on match statistics and scenarios within games with great success. For the purposes of this thesis, two rule amendments were looked at and in both cases it was found that their implementation would have a minimal effect in terms of improving the quality of the game. The indicators used for quantifying an improvement were a reduction in stoppages (i.e. ball-up bounces) and an increase in scoring events. It is believed that such a tool could be useful for the AFL in determining whether the benefit of changes outweighs the effect of altering the set up of the game. Further rule changes can be investigated provided the adjustment to the rules is able to be reflected via adjustments to transition count data and hence the probability matrices.

9.4 Using simulation to investigate matches after their completion

The investigation of rule changes and their effect presented above would be a useful application of the model for the administrators of the AFL. This section will look at applications of the model that would be of most interest to the media and football public in general. There would also be some benefit to the clubs in terms of addressing strategical issues or match tactics, which could improve the chances of victory. Once again, the analysis here is concerned with matches after they have been completed with two distinct applications looked at. Firstly, the relative chances of victory of each side based on the transition probabilities stemming from the match will be investigated. Following that, analysis will be done on where teams could make changes to how they played in a particular game to improve their chances of victory.

9.4.1 Relative probability of match outcomes based on observed transitions

Sometimes in a game of AFL football, the scoreboard may not be indicative of who deserved to win the match. For instance, the scoreboard may flatter one team over the other and indicate a close game when, in reality, that was not necessarily the case. Furthermore, a match where scores ended up level at the completion may be an injustice to one of the teams competing. The simulation program described earlier in this chapter can be utilised to calculate the expected proportion of games either team in a match would win or draw based on the probabilities derived from the match. For the purposes of this analysis, the number of transitions that were observed in the match is used for the length of the simulation. This application would be of interest to the print and electronic media as a way of quantifying the ‘real’ chances of victory of each team and could be used in post match analysis to credit or discredit a team's performance. Several games from the 2004 and 2003 seasons that finished close on the scoreboard have been chosen to illustrate this application.

9.4.2 Examples using close games from 2003 and 2004

To illustrate this application of the model, seven close games from each season were selected and analysed. One final in 2003 was chosen where the margin was 12 points, and one in 2004 where the margin was nine points. These games aside, the size of the margin for the remaining 12 games was six or less and one game was a draw. Each match was extracted from the data set and simulated independently with the number of transitions used for the simulation reflecting the number of transitions that the match actually contained. The score for each team was then calculated and a result derived based on these scores. Table 9.10 contains the teams for each match, the winner and margin and the probability of either team winning or a draw from the 10,000 simulations.

Table 9.10: Likelihood of victory for either team in close matches from 2003 and 2004

Season	Round	Team A	Team B	Expected Margin	Pr(A win) %	Pr(B Win) %	Pr(Draw) %
2003	2	HA	WC	2	50.5	48.2	1.4
2003	3	NM	BL	0	47.3	51.5	1.2
2003	11	NM	RI	3	50.8	47.8	1.4
2003	14	GE	PA	1	55.9	42.6	1.5
2003	14	FR	BL	3	50.8	47.7	1.5
2003	16	FR	AD	1	43.4	55.3	1.3
2003	QF	PA	SY	12	36.5	62.3	1.2
2004	3	ES	WC	6	54.7	44.3	1.0
2004	6	RI	HA	1	48.4	50.1	1.5
2004	11	AD	CA	-4	44.3	54.4	1.3
2004	22	CA	CO	1	49.4	49.6	1.0
2004	EF	ME	ES	-5	41.6	57.2	1.2
2004	PF	BL	GE	9	62.9	35.7	1.4
2004	PF	PA	ST	6	63.3	35.6	1.2

For the 14 matches analysed, on three occasions the team that lost the game was actually a better chance of winning than their opposition. The biggest discrepancy came in the round 16 match from 2003 between Fremantle and Adelaide, which Fremantle won by one point. In this instance, Fremantle was a 43.4% chance whilst Adelaide was a 55.3% chance. These types of matches can make or break a season for a team and highlight instances where one team was unlucky to lose.

The other scenario is where scores were close at the final siren but the losing side was flattered by the result. An example of this is the preliminary final from 2004 when Port Adelaide was successful in a close game by six points. The simulation for this match shows that they should have won by more than they did and had about a 63% chance of winning. To illustrate this further, the expected score of each team can be extracted from the simulation program. In the match between Port Adelaide and St. Kilda the expected scores were 97.1 and 85.9 respectively. The expected score line reflects the notion that the losing side could be considered lucky to get as close as they did.

This application could be valuable in a bookmaking environment for deriving probabilities of victory in a live environment and subsequent betting prices for each team. The opportunity to develop this further is available for the following AFL season and this technique will be used to do so.

9.4.3 Adjusting transition probabilities to improve chances of victory

The previous section looked at matches as they happened and assigned winning probabilities based on the observed transition matrices. This type of application would be useful to the media but a slight variation may make it attractive to the AFL clubs as well. This variation is the adjustment of probabilities based on the observed count data to identify how, and by what magnitude, a team could improve its chances of victory. Examples of areas that could be addressed by this application could be reducing error rate (i.e. Team Possession → Opposition Possession), better efficiency at stoppages (i.e. more clearances), improved accuracy on goal or improved ability to extract the ball from dispute. Whilst this list is by no means exhaustive they give an idea of the major areas that a team might try to adjust to gain better control within a match.

For the purposes of displaying this application in a working environment, we concentrate on the matches that were looked at above as the chances of victory of each team have already been derived. Two matches that have been chosen as examples are the preliminary finals from 2004. In these games the winning margins were six points and nine points, however it was shown in 9.4.2 that the losing sides were perhaps lucky to get as close as they did.

9.4.3.1 Port Adelaide v St. Kilda

This match was played at Football Park in Adelaide on Friday, 17th September, 2004 in front of 46,978 spectators. It was a game that Port Adelaide was heavily favoured to win, with the bookmakers pricing them a $1.27 favourite against St. Kilda at $4.10. The final margin of six left the Saints disappointed that they had got as close as they did but at the same time seen their season ended only one win away from the grand final. Tables 9.11 and 9.12 contain the observed count data for the match and the transition matrix respectively, with Port Adelaide Team A, and St Kilda Team B.

Table 9.11: Observed count data: Port Adelaide v St. Kilda

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH	Total
CEBO	0	7	0	1	10	13	0	0	31
BUBO	0	6	0	1	13	18	0	0	38
THIN	0	5	1	0	14	12	0	0	32
DISP	0	18	27	6	57	50	0	0	158
APOS	14	0	2	69	114	20	10	0	229
BPOS	13	2	2	80	18	159	0	10	284
ABEH	0	0	0	1	0	8	0	0	9
BBEH	0	0	0	0	8	1	0	0	9

Table 9.12: Observed transition matrix: Port Adelaide v St. Kilda

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH
CEBO	0.00%	22.58%	0.00%	3.23%	32.26%	41.94%	0.00%	0.00%
BUBO	0.00%	15.79%	0.00%	2.63%	34.21%	47.37%	0.00%	0.00%
THIN	0.00%	15.63%	3.13%	0.00%	43.75%	37.50%	0.00%	0.00%
DISP	0.00%	11.39%	17.09%	3.80%	36.08%	31.65%	0.00%	0.00%
APOS	6.11%	0.00%	0.87%	30.13%	49.78%	8.73%	4.37%	0.00%
BPOS	4.58%	0.70%	0.70%	28.17%	6.34%	55.99%	0.00%	3.52%
ABEH	0.00%	0.00%	0.00%	11.11%	0.00%	88.89%	0.00%	0.00%
BBEH	0.00%	0.00%	0.00%	0.00%	88.89%	11.11%	0.00%	0.00%

From this match, there is nothing of note that stands out for St. Kilda, except for maybe their lack of efficiency with the ball in kicking goals. They had a lot more of the ball than Port Adelaide, committed fewer errors and outpointed their opposition around stoppages. To illustrate the power of the model, one of St. Kilda's behinds will be replaced by a goal, which, on the scoreboard would have made them lose by one point instead of six and at the same time, their error rate will be reduced by six, with these errors being replaced by disposals that retained the ball. This adjustment results in the transition matrix in Table 9.12 where it can be seen that their efficiency on goal is slightly increased along with their ability to retain possession, whilst their behind rate has decreased along with their rate of turning the ball over to Port Adelaide.

Table 9.13: Adjusted transition matrix: Port Adelaide v St. Kilda

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH
CEBO	0.00%	22.58%	0.00%	3.23%	32.26%	41.94%	0.00%	0.00%
BUBO	0.00%	15.79%	0.00%	2.63%	34.21%	47.37%	0.00%	0.00%
THIN	0.00%	15.63%	3.13%	0.00%	43.75%	37.50%	0.00%	0.00%
DISP	0.00%	11.39%	17.09%	3.80%	36.08%	31.65%	0.00%	0.00%
APOS	6.11%	0.00%	0.87%	30.13%	49.78%	8.73%	4.37%	0.00%
BPOS	4.93%	0.70%	0.70%	28.17%	4.23%	58.10%	0.00%	3.17%
ABEH	0.00%	0.00%	0.00%	11.11%	0.00%	88.89%	0.00%	0.00%
BBEH	0.00%	0.00%	0.00%	0.00%	88.89%	11.11%	0.00%	0.00%

Using Table 9.13 in the simulation program with those small adjustments to reality has had the desired effect of giving St. Kilda a better chance of winning than Port Adelaide. In the real game they had a 35.6% chance of winning with an expected score line of 97-

86, however, under the adjusted scenario they have a 52% chance of winning with an expected score line of 92-95.

9.4.3.2 Brisbane v Geelong,

This match was played at the Melbourne Cricket Ground in Melbourne on Saturday, 18th September, 2004 in front of 55,768 spectators. Again, it was a game where the favoured side, Brisbane, was long odds-on with the bookies, however most commentators believed that Geelong had enough possession in the game to actually win the match. The experience of Brisbane in finals football probably made the difference in the end. Tables 9.13 and 9.14 contain the observed count data for the match and the transition matrix respectively, with Brisbane Team A, and Geelong Team B.

Table 9.14: Observed count data: Brisbane v Geelong

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH	Total
CEBO	0	4	0	1	9	12	0	0	26
BUBO	0	0	0	0	12	6	0	1	19
THIN	0	5	2	0	13	9	0	0	29
DISP	0	10	21	4	65	50	0	0	150
APOS	12	0	3	77	152	29	12	1	286
BPOS	10	0	3	65	28	162	0	13	281
ABEH	0	0	0	0	0	12	0	0	12
BBEH	0	0	0	3	10	2	0	0	15

Table 9.15: Observed transition matrix: Brisbane v Geelong

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH
CEBO	0.00%	15.38%	0.00%	3.85%	34.62%	46.15%	0.00%	0.00%
BUBO	0.00%	0.00%	0.00%	0.00%	63.16%	31.58%	0.00%	5.26%
THIN	0.00%	17.24%	6.90%	0.00%	44.83%	31.03%	0.00%	0.00%
DISP	0.00%	6.67%	14.00%	2.67%	43.33%	33.33%	0.00%	0.00%
APOS	4.20%	0.00%	1.05%	26.92%	53.15%	10.14%	4.20%	0.35%
BPOS	3.56%	0.00%	1.07%	23.13%	9.96%	57.65%	0.00%	4.63%
ABEH	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%
BBEH	0.00%	0.00%	0.00%	20.00%	66.67%	13.33%	0.00%	0.00%

In this match, both teams had about the same amount of the ball; however Geelong were clearly outpointed by Brisbane at general play stoppages (BUBO and THIN) and getting the ball out of dispute. Adjustments have been made so that they broke even with their opponents in these three areas and Table 9.16 reflects these changes.

Table 9.16: Observed count data: Brisbane v Geelong

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH
CEBO	0.00%	15.38%	0.00%	3.85%	34.62%	46.15%	0.00%	0.00%
BUBO	0.00%	0.00%	0.00%	0.00%	47.37%	47.37%	0.00%	5.26%
THIN	0.00%	17.24%	6.90%	0.00%	37.93%	37.93%	0.00%	0.00%
DISP	0.00%	6.67%	14.00%	2.67%	38.67%	38.00%	0.00%	0.00%
APOS	4.20%	0.00%	1.05%	26.92%	53.15%	10.14%	4.20%	0.35%
BPOS	3.56%	0.00%	1.07%	23.13%	9.96%	57.65%	0.00%	4.63%
ABEH	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%
BBEH	0.00%	0.00%	0.00%	20.00%	66.67%	13.33%	0.00%	0.00%

These adjustments have lifted Geelong's probability of victory to better than Brisbane's chances. They now have a 51% chance compared with Brisbane's 47% chance and the expected score line is now 79-80 in favour of Geelong.

9.4.4 Summary

The examples given here have provided a good insight into the capabilities of the model in identifying particular areas of a match that are influential to the result. This application is not restricted to matches that were close and perhaps has greater importance in games where the margin was wide as it could help a coach isolate particular areas of his team's play where improvement will make the team more competitive. It is also a powerful tool for quantifying which statistics within a game of football are the most crucial to helping a team to victory.

9.5 Simulation of fantasy matches

The final application of the model to be discussed in this chapter relates to the 'playing' of so called fantasy games. Often commentators muse about the merits of teams from different eras and who were the better side. Previously there has been no way to settle the debate definitively; however, with the construction of this model and the simulation program, a mathematical answer can now be given. Obviously, one has to derive a transition matrix for both sides in order to be able to simulate the match and this is where various assumptions need to be made. These assumptions relate to the amount of data to include in order to derive the matrix. If one was interested in looking at teams from different eras then perhaps the entire season data for each team would be appropriate. Alternatively, if a grand final is to be simulated prior to the match, then perhaps the best approach is to use the season data for each side at the venue. These types of issues would be addressed with reference to the match in question and the following example will discuss this in more detail.

9.5.1 2004 fantasy grand final – Geelong v St. Kilda, M.C.G.

As these two teams have featured heavily in this chapter, it was decided to use them as an example of how a fantasy match could be set up. Both teams played three games at the M.C.G. in 2004 with Geelong winning two and losing one and St. Kilda losing two and winning one. If, as hypothesised in section 9.4 both teams won their preliminary finals then this match would have eventuated as the grand final. It would have been a dream match-up for officials with both teams Melbourne-based and starved of recent premiership success. Geelong has won only six premierships in their history with the last coming in 1963. St. Kilda has won only one premiership in their history and that was in 1966. This match-up would have seen one of these teams emerge from the football wilderness. But which club would have enjoyed the spoils and who would have endured the heartbreak of getting so close again only to fail?

In compiling a transition matrix for each team it was decided to use the three games each team had played at the M.C.G. in 2004 as a good starting point for the styles of these teams at the venue. The count data for each team from these three games is then amalgamated to give an overall matrix for their performance at the M.C.G. The amalgamated totals for Geelong are presented in Table 9.17 and for St. Kilda in Table 9.18. Geelong has been assigned Team A and St. Kilda Team B based solely on alphabetical order. Furthermore, the opposition's transitions are not included as only the competing team's profiles are used in this application.

Table 9.17: Amalgamated count data for Geelong at M.C.G. 2004

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH	Total
CEBO	0	9	0	4	37	30	0	0	80
BUBO	0	3	4	2	29	25	1	0	64
THIN	0	14	4	7	34	31	0	0	90
DISP	0	34	70	29	194	209	0	0	536
APOS	36	3	8	238	543	85	45	1	959
BBEH	0	0	0	4	34	1	0	0	39

Table 9.18: Amalgamated count data for St. Kilda at M.C.G. 2004

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH	Total
CEBO	0	14	0	3	36	43	0	0	96
BUBO	0	24	2	6	50	37	0	0	119
THIN	0	16	3	2	33	39	0	1	94
DISP	0	64	79	16	150	161	0	0	470
BPOS	40	2	3	202	81	450	0	31	809
ABEH	0	0	0	7	11	11	0	0	29

The rows for APOS, BPOS, ABEH and BBEH are simply extracted from each of the matrices presented above. In order to obtain the rows for CEBO, BUBO, THIN and DISP, the rows are added together from each matrix to get an overall profile, however they could be averaged to gain a matrix that reflects the expected cell counts for a match. The combined count matrix and subsequent transition probability matrix are presented in Table 9.19 and 9.20 respectively.

Table 9.19: Amalgamated count data for Geelong and St. Kilda at M.C.G. 2004

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH	Total
CEBO	0	23	0	7	73	73	0	0	176
BUBO	0	27	6	8	79	62	1	0	183
THIN	0	30	7	9	67	70	0	1	184
DISP	0	98	149	45	344	370	0	0	1006
APOS	36	3	8	238	543	85	45	1	959
BPOS	40	2	3	202	81	450	0	31	809
ABEH	0	0	0	7	11	11	0	0	29
BBEH	0	0	0	4	34	1	0	0	39

Table 9.20: Transition probability matrix for fantasy grand final between Geelong and St. Kilda at M.C.G., 2004

state	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH
CEBO	0.00%	13.07%	0.00%	3.98%	41.48%	41.48%	0.00%	0.00%
BUBO	0.00%	14.75%	3.28%	4.37%	43.17%	33.88%	0.55%	0.00%
THIN	0.00%	16.30%	3.80%	4.89%	36.41%	38.04%	0.00%	0.54%
DISP	0.00%	9.74%	14.81%	4.47%	34.19%	36.78%	0.00%	0.00%
APOS	3.75%	0.31%	0.83%	24.82%	56.62%	8.86%	4.69%	0.10%
BPOS	4.94%	0.25%	0.37%	24.97%	10.01%	55.62%	0.00%	3.83%
ABEH	0.00%	0.00%	0.00%	24.14%	37.93%	37.93%	0.00%	0.00%
BBEH	0.00%	0.00%	0.00%	10.26%	87.18%	2.56%	0.00%	0.00%

The total number of transitions for the Geelong matches was 2732 and for the St. Kilda games it was 2466. These combine to give 5198 transitions in six matches at an average of 866 transitions per match. This figure was used, along with the transition matrix in Table 9.20 to simulate a fantasy grand final between the two teams. Long suffering fans of St. Kilda will be disappointed that this is only a hypothetical as St. Kilda were given a 63.4% chance of winning and only a 35.3% chance of losing. The draw was a 1.3% chance of occurring. The expected score line for the match was 95-84 in St. Kilda's favour.

9.5.2 Simulation summary

As evidenced by the example presented, this type of application has far reaching implications and would be very useful for the football media, both for pre-match simulations and comparisons of teams from different eras. Not only is the likelihood of victory and expected score able to be elicited from the simulation, but a match in its entirety can also be extracted. This could allow for ‘phantom’ calls of important matches, prior to their happening, with some degree of accuracy. For instance, a simulated match could be randomly extracted from the 10,000 simulations and converted into a match transaction file, which a media outlet could then use to call the game for its listeners complete with quarter by quarter breakdown and summary. The author was approached by a radio station in Brisbane to do this kind of analysis for the third State of Origin rugby league match in 2005.

9.6 Summary

The applications presented here give a solid introduction to the capabilities of the global Markov process model. These applications have great practical scope for all different facets of the football industry. They are by no means an exhaustive list of applications, with range only limited by what would be considered worthwhile for a potential client. Further advantages of the simulation technique are the ability to track matches throughout their entirety without simply getting a summary of the match upon completion. Furthermore, there is the capability to investigate seasons after their completion and identify who was lucky or unlucky and whether the premiers were worthy of the title. This type of investigation has been performed in other sports using different techniques (Lee, 1999, Lee, 1997) but could be of interest to AFL followers.

Whilst all of these applications pertain to matches after they have been completed, some of them could be applied during a match, e.g. perturbing transition probabilities, to investigate how to maximise one’s chance of victory or predicting the chance of winning at breaks in the game. The following chapter will introduce more applications of the model that relate more closely to using the model in a more dynamic environment such as during the game.

Chapter 10: End game and dynamic applications of a global Markov process model

10.1 Introduction

The previous chapter looked at applications of the global model for analysing matches after they have been completed. This chapter is more concerned with using the model in a dynamic environment, more particularly, when a game is in progress. Initially, an application will be given that relates to completed matches, however, its real use is in end game analysis and the importance of a single piece of play. Such an application is worthwhile for quantifying the effect of an event within a game and can be used to answer the cry of every spectator heard at some stage throughout the season that their team was robbed. Intuitively, as a match draws closer to its end, the ability of a simulation programme to approximate the result improves. This chapter is concerned with two main areas; firstly, altering passages of play and simulating the remainder of a match to forecast a result under different scenarios; and secondly, using observed probabilities at varying stages of matches to predict a final margin.

10.2 End game analysis for investigating different play scenarios

Research has often been tailored to investigate 'what if?' scenarios in sporting contexts and come up with a mathematical answer to a hypothetical scenario. Work in the area goes back to the 1970s with investigations done on how, by adjusting a baseball batting line-up, chances of victory could be improved (Freeze, 1974, Trueman, 1976). Other areas of interest have been route choice in orienteering (Hayes and Norman, 1984), service choice in tennis (Norman, 1985), one or two point play after a touchdown in American football (Boronico, 2000) and countless investigations in cricket relating to scoring rates (Clarke, 1988), protecting the weaker batsman (Clarke and Norman, 1995) and choosing a night watchman (Clarke and Norman, 2003).

The application of dynamic programming in Australian Rules football has been limited with the only instance of it being an investigation of when a team could increase their chances of winning by rushing a behind for their opposition (Clarke and Norman, 1998). This was a rather simplistic approach that looked at end game scenarios with different margins and assessed the improvement in a team's chances of victory if they rushed a behind. Basically the recommendation was that if a team was down by less than five points near the end of a match and was under extreme defensive pressure, they would be better off to concede a point and return the ball via a kick in, in the hope that they could go the length of the field and kick a goal and win the match. Whilst this idea will not be revisited in this thesis, the model proposed herein would be more than capable of handling such an investigation.

This section is based upon offering up different scenarios to what actually took place and arriving at different match outcomes based on the change in scenario. Such an investigation was the subject of a television report screened on Channel Seven in Perth that focussed on a scenario in a match that could have impacted on the finals performance of the West Coast Eagles (Butler, 2005). Different outcomes were arrived at using varying scenarios to what actually took place. This kind of analysis would have a use in many different forums and some of these will be addressed using different examples. Before these examples are given, a description of how the model is used in this manner will be provided.

10.2.1 Using a global Markov model to alter the events of a match

The modern day game of AFL football has never enjoyed the spotlight as much as at present. The coverage of the game nationally is very strong, and with a talkback radio station in Melbourne that concentrates mostly on AFL, the amount of analysis and dissection of games is at an all time high. Quite often discussion will centre on controversial decisions made by umpires or basic skill errors by players that allegedly affect the match result. Much speculation is made that these types of events led to the result, however in reality this is just speculation. The global model can be used to quantify how much an event impacted upon a team’s chances of victory.

Any scenario based around an event accounted for in the model can be investigated. The analysis is done by calculating the transition probabilities up to but not including the event in question. So for instance, if a missed kick at goal was being looked at in the final quarter, the transition probabilities would be calculated for the match using the events up to the kick on goal. Having done this, the number of transitions left in the match is used to simulate the rest of the game 10,000 times and predict an outcome based on the different scenarios. Calculating the number of transitions left in the match is straightforward if the game has been completed, however in a dynamic environment a simplistic approach could be used based on time left and the average number of transitions per period of time that have taken place leading up to the event. To gain a better understanding of how the analysis is performed and areas of application, several examples will be presented below.

10.2.2 Case study 1: Failure to convert a set shot on goal

This example investigates inaccuracy by a player in front of goal and the impact of the miss on the team’s chances. This application would be particularly useful for media outlets analysing a game after the event and also for supporters who are interested in knowing whether the kicker has cost them the game. An example of this event comes from the round 19 match in 2003 between Hawthorn and Port Adelaide, played at the MCG. Hawthorn was making a late season charge and needed to win their last four games to make the finals. With roughly two minutes left in the match, Hawthorn led by one point with their full forward, Jade Rawlings, lining up for a goal from a relatively simple position to put Hawthorn in front by seven points. His kick on goal missed to the right and the resultant point made the lead two in Hawthorn’s favour. Port Adelaide was able to transfer the ball the length of the ground and kick a goal that saw them snatch a four point victory. Two different scenarios arise here and need to be analysed separately.

The first is what actually happened with the kick being missed, whilst the second involves what would have happened if Rawlings had managed to kick the goal.

The miss by Rawlings gave Hawthorn a two point lead with 16 transitions left in the match. To simulate the remainder of the game and ascertain Port Adelaide's chances of stealing an unlikely victory, the transition matrix for the match up to Rawlings's miss has been derived and is contained in Table 10.1. These probabilities are then used in the simulation program with 16 transitions left and a margin of two points in Hawthorn's favour starting from state seven, i.e. a Port Adelaide kick-in after a Hawthorn behind.

Table 10.1: Transition probability matrix, HA v PA including Rawlings' miss

State	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH
CEBO	0.0%	15.4%	0.0%	0.0%	38.5%	46.2%	0.0%	0.0%
BUBO	0.0%	16.7%	5.6%	0.0%	30.6%	44.4%	2.8%	0.0%
THIN	0.0%	31.0%	6.9%	0.0%	31.0%	31.0%	0.0%	0.0%
DISP	0.0%	9.2%	14.8%	3.5%	36.6%	35.9%	0.0%	0.0%
APOS	4.2%	0.4%	1.2%	22.3%	55.0%	11.9%	5.0%	0.0%
BPOS	4.3%	1.2%	0.4%	28.9%	8.7%	51.8%	0.0%	4.7%
ABEH	0.0%	7.7%	0.0%	46.2%	7.7%	38.5%	0.0%	0.0%
BBEH	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%

With Hawthorn leading by two points and only 16 transitions left in the match, the simulation shows that Hawthorn was an 81.8% chance of winning, Port Adelaide a 17.2% chance and the draw was a 1.0% chance.

Alternatively, if Rawlings kicks the goal, Hawthorn has a lead of seven points with only 16 transitions left. The only alteration to the matrix contained in Table 10.1 is that the cell APOS→ CEBO increases from 4.2% to 4.6% whilst the cell APOS → ABEH decreases from 5.0% to 4.6%. With these adjustments and a seven point lead, the simulation is run starting from state one, i.e. a CEBO after a goal. Under these conditions, Hawthorn is a 96.2% chance of victory, Port Adelaide a 1.6% chance and the draw is a 2.2% chance. Therefore, it can be said that the miss by Rawlings increased Port Adelaide's winning chance by nearly 16%. However, it may be unfair to pile the woes of the team's season on that one kick given that they still should have won the game from that position according to their chances of victory.

10.2.3 Case study 2: Wrong decision made by an official

With decisions made by officials open to human error in most competitive sports, there is the chance that a decision can impact on the final result. What irks supporters more than anything is when the decision is shown to be incorrect, however it is impossible for the decision to be reversed and the result must stand. Australian football is no different to other sports and sometimes a decision, whether it is made by a field umpire, boundary umpire or goal umpire can materially affect the outcome of the match. Another ready made application for the global model is the investigation of decisions made by officials during a match and whether these decisions affect the result. Not only could this be used by the AFL for coaching and training of umpires but could also be used where claims were made by a club that the decision cost them the game.

The decision investigated here comes from the 2004, round six match between St. Kilda and Brisbane, which St. Kilda won by a point. The match was highlighted by a goal umpire error which ruled a kick from St. Kilda's Austinn Jones as a behind when it should have been ruled out of bounds on the full. This decision proved the difference between the teams with a goal on the siren resulting in a one point win to St. Kilda. The ramification of this decision carried into the 2005 season when the AFL raised the height of all goal posts, at a considerable cost, to reduce the chance of such a decision. Again we are looking at two scenarios here and they will be presented separately.

Firstly, the play as it happens is looked at with Jones' kick on goal being called a behind. At this stage Brisbane has a lead of five points with 14 transitions left and the simulation starts in state seven i.e. St. Kilda behind. The transition matrix for the play in the match up to the kick in by Brisbane is contained in Table 10.2. The simulation with these initial conditions gives St. Kilda an 11.5% chance of winning, Brisbane an 87.3% chance and the draw is a 1.2% chance.

Table 10.2: Transition probability matrix, ST v BL including Jones' behind

State	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH
CEBO	0.0%	10.0%	0.0%	0.0%	50.0%	40.0%	0.0%	0.0%
BUBO	0.0%	8.3%	0.0%	0.0%	37.5%	54.2%	0.0%	0.0%
THIN	0.0%	12.9%	3.2%	3.2%	35.5%	45.2%	0.0%	0.0%
DISP	0.0%	9.3%	16.0%	4.9%	33.3%	36.4%	0.0%	0.0%
APOS	4.4%	0.0%	0.7%	26.7%	53.8%	9.2%	5.1%	0.0%
BPOS	4.7%	0.0%	0.7%	25.4%	10.5%	56.3%	0.0%	2.4%
ABEH	0.0%	0.0%	0.0%	15.4%	23.1%	61.5%	0.0%	0.0%
BBEH	0.0%	0.0%	0.0%	28.6%	71.4%	0.0%	0.0%	0.0%

If Jones' kick is called correctly as out of bounds on the full, the adjustment to the matrix from Table 10.1 would see APOS → ABEH decrease from 5.1% to 4.8% and APOS → BPOS increase from 9.2% to 9.5%. The initial conditions for the simulation program are a margin of six points in Brisbane's favour and the starting state is six, i.e. Brisbane possession as the ball was out on the full. Running the program with 14 transitions left gives St. Kilda a 2.3% chance of winning, Brisbane an 89.7% chance and the draw is an 8.0% chance. In light of this analysis it would seem that the Saints were on the right side of some divine intervention. More importantly, Brisbane could argue rightly that the decision made by the goal umpire has caused them to lose a match they would not otherwise have lost.

10.2.4 Case study 3: Decision making by a player late in a game

Decision making by players can be crucial towards the end of a tight match. A simple misdirected disposal or not disposing of the ball instinctively, and choosing to hold on to it, may prove to be the match-turning mistake with the opposition seizing on the error and stealing the match as a result. The model can be used to investigate the likely results for a particular decision or other scenarios that the player could have pursued. This could then become a useful tool in coaching players and refining their decision making process so that they instinctively make correct choices.

The scenario chosen for analysis comes from the story run by Channel Seven in Perth and centres on the round 3 match from 2004 between Essendon and West Coast. Late in the game a ball up bounce occurred deep in Essendon's attacking zone and the resultant possession was gained by Chris Judd of the West Coast. In heavy traffic, Judd feigned a handball before being forced to handball whilst being tackled. The ball spilled to an Essendon player, with his handball finding a team-mate who snapped an unlikely goal from the boundary to snatch a six point victory. Three scenarios were looked at in this situation using the transition probability matrix resulting from the match up to Judd's ineffective handball. This matrix is presented in Table 10.3

Table 10.3: Transition probability matrix, ES v WC including Judd handball

State	CEBO	BUBO	THIN	DISP	APOS	BPOS	ABEH	BBEH
CEBO	0.0%	4.4%	0.0%	6.7%	48.9%	40.0%	0.0%	0.0%
BUBO	0.0%	9.5%	0.0%	4.8%	47.6%	38.1%	0.0%	0.0%
THIN	0.0%	11.1%	11.1%	3.7%	40.7%	33.3%	0.0%	0.0%
DISP	0.0%	8.4%	13.3%	2.4%	35.5%	40.4%	0.0%	0.0%
APOS	6.7%	0.0%	0.6%	27.2%	54.6%	9.3%	1.6%	0.0%
BPOS	7.3%	0.0%	0.0%	26.2%	11.6%	50.9%	0.0%	4.0%
ABEH	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%
BBEH	0.0%	0.0%	0.0%	9.1%	81.8%	9.1%	0.0%	0.0%

The first scenario simulates play as is. There are 10 transitions left from Judd's ineffective handball and the starting state is four, ball in dispute, with scores level. Under these conditions, Essendon has a 20.7% chance of winning, West Coast a 25.4% chance, and the draw has a 53.9% chance. These probabilities reflect a drawback of the global model in that no reference is made to location on the ground. The piece of play has occurred deep in Essendon's attacking zone and with this information included in the simulation the probabilities would be more accurate. Chapter 11 reproduces this analysis using a zone Markov process model and the difference in probabilities will be discussed there.

Secondly, it has been assumed that Judd has handballed instinctively when he had the chance and has hit a team mate. The only condition that changes is the starting state, which becomes six, West Coast possession. With a handball that hits the target, Essendon has a 16.9% chance of winning, West Coast a 34.8% chance and the draw has a 48.3% chance. Lastly, the scenario of the ball being tied up for a ball up has been looked at and the starting state becomes two. Under these conditions, Essendon has a 21.9% chance of winning with West Coast a 24.4% chance. The draw has a 53.7% chance of occurring. Simple adjustments of the probabilities reflect just how important a single passage of play can be. In the first week of the finals, the West Coast had to travel to Sydney due to inferior percentage. A draw in the Essendon game and the resulting two points would have given the Eagles home ground advantage in the first week. Such a powerful analysis tool can be used to assist and prepare players for the pressure moments of a game. For example, in tight situations like the Judd example, players could be trained in the art of making decisions such as tying the ball up and resetting at a stoppage, rather than appearing to hit and hope, like Judd did.

10.3 Using 'in game' transition probabilities to predict final margin

The initial aim of this research was to derive a dynamic prediction model for AFL football. The bulk of the applications presented so far have revolved around post match analysis. This section investigates the predictive nature of the transition probabilities from within a match. It is hoped that a dynamic model can be constructed to assist coaches during the game with the decisions they make. This chapter has already presented an application for the model in a dynamic environment by simulating the remainder of the match based on what has already taken place to assign probabilities to the outcomes of a match. Unfortunately, this approach could prove time consuming in a game environment and therefore would not be suitable for predicting the outcome of a match at regular intervals. Furthermore, it is accepted that whilst this technique is extremely powerful towards the end of the game, it would struggle to accurately predict outcomes from earlier positions within a match. For this reason, another model had to be developed that could quickly and easily compute expected margins from any point in a match with some accuracy. It was decided to investigate a multiple linear regression model that uses margin of victory (MOV) as the dependent variable and the transition probabilities from the match as the independent variables. The distribution of MOV in Australian football is well approximated by the normal distribution (Bailey and Clarke, 2004), allowing for a multiple linear regression model using significant transition probabilities to derive a prediction equation for final margin.

In the dynamic environment of an AFL match, coaches would appreciate the analytical tool to understand where the match is headed based on what has taken place. In order to provide this kind of tool to coaches, it was important to ascertain the predictive capabilities of the transition probabilities at various times during the match. To do this, the recognised breaks in the game i.e. quarter time, half time and three-quarter time were used along with full time. The transition probabilities from each of the 370 matches were extracted for each of these time periods.

Four models were then constructed that built up the transition probabilities depending on which time period was being looked at. For instance the quarter time model used only the transition probabilities from the first quarter whereas the three-quarter time model used the transition probabilities from the first three quarters of the match. Although the global model is an eight state Markov model, giving rise to 64 possible transitions, a game of Australian football can only throw up 47 possible transitions so each quarter model could have at most 47 significant transitions. To try and improve the predictive power of the model for the early parts of the match, ratings derived from Chapter 7 have been included as predictors. There are eight ratings, being home and away team goals and behinds, on attack and on defence. It is hypothesised that early in the match these ratings may play a more important role than later in the match when the in-game transition probabilities take over. The four models will be looked at individually and it should be noted that 0.15 has been used as the level of significance for a variable to be included in the model. This is the default level from SAS and was retained in this analysis. Included in the summary tables for each model are the standardised estimates for determining the relative importance of each transition in the prediction.

10.3.1 Using first quarter transition probabilities to predict final margin

A model was developed to predict MOV using the transition probabilities from the first quarter of the match as well as the ratings used for the pre-match prediction model presented in Chapter 7. Using stepwise selection in SAS, the final model for MOV_1 included pre-match goal ratings and 11 of the 47 transition probabilities from first quarter data. Details of the quarter one model are included in Table 10.4

Table 10.4: Parameter estimates for first quarter model

Variable	Parameter Estimate	Partial R^2	Model R^2	P-value	Standardised Estimate
Intercept	18.08			0.64	0
BPOSCEBO	-539.22	15.60%	15.60%	<.0001	-0.272
APOSCEBO	423.86	9.90%	25.50%	<.0001	0.241
b_gl_def	-3.53	4.50%	30.00%	0.001	-0.156
a_gl_def	2.99	3.20%	33.20%	0.006	0.119
b_gl_att	-3.54	1.40%	34.60%	0.006	-0.124
a_gl_att	2.70	1.30%	35.90%	0.024	0.099
BPOSDISP	87.63	1.20%	37.10%	0.001	0.149
APOSDISP	-72.08	0.90%	38.00%	0.008	-0.115
BUBOAPOS	22.68	0.70%	38.70%	0.006	0.115
BPOSAPOS	131.28	0.60%	39.30%	0.025	0.095
BUBOABEH	224.23	0.50%	39.80%	0.026	0.096
DISPBPOS	-52.66	0.50%	40.30%	0.039	-0.089
ABEHBUBO	-57.88	0.40%	40.70%	0.117	-0.068
CEBOAPOS	15.44	0.40%	41.10%	0.092	0.070
THINBUBO	-22.48	0.40%	41.50%	0.085	-0.072

Not surprisingly, both Team A and B's ability to kick goals (APOSCEBO, BPOSCEBO) are the most significant variables in the model. The four pre-match ratings for goals are the next most important variables indicating that their inclusion is crucial to final margin prediction. To highlight this, the adjusted R^2 for this model is 39.6% of the variation in final margin whereas a model that does not include pre-match ratings explains only 31.7% of the variation in final margin. The ratings will be included in the prediction models until they no longer become significant.

10.3.2 Using second quarter transition probabilities to predict final margin

This model builds on the information from the first quarter by including the transition probabilities for the second quarter and thus the transition probabilities from the first half are used. These probabilities from the first half are used to try and predict the final margin of the match. It is expected that, as the probabilities are built on throughout the game, the accuracy of the model will improve. Again, using stepwise selection in SAS, the final model for MOV_2, the margin of victory based on quarter one and two transition probabilities and pre-match ratings, contains 12 of the 47 transition probabilities and four of the pre-match ratings. Interestingly, the pre-match ratings that are included are different to those included in the quarter one model. Furthermore, three of the four ratings are for Team B with the rating for B's attacking goals replaced with Team A's defensive goal rating. Details of the quarter two model are presented in Table 10.5.

Table 10.5: Parameter estimates for second quarter model

Variable	Parameter Estimate	Partial R^2	Model R^2	P-value	Standardised Estimate
Intercept	88.49			0.006	0
APOSCEBO	1114.95	26.20%	26.20%	<.0001	0.447
BPOSCEBO	-1120.04	26.20%	52.40%	<.0001	-0.445
b_gl_def	-2.73	3.00%	55.40%	0.002	-0.121
APOSDISP	-113.09	1.90%	57.30%	<.0001	-0.148
APOSABEH	265.73	1.60%	58.90%	0.002	0.101
a_gl_def	2.34	1.40%	60.30%	0.008	0.093
BPOSDISP	86.17	1.20%	61.50%	0.001	0.117
APOSTHIN	-545.03	0.90%	62.40%	0.014	-0.080
ABEHBPOS	-14.00	0.50%	62.90%	0.011	-0.083
THINBUBO	-30.47	0.50%	63.40%	0.039	-0.068
BPOSAPOS	116.84	0.40%	63.80%	0.058	0.063
DISPBPOS	-45.96	0.40%	64.20%	0.086	-0.058
BUBOTHIN	-42.73	0.40%	64.60%	0.055	-0.061
BPOSBUBO	-756.91	0.30%	64.90%	0.037	-0.068
b_bh_att	-2.29	0.30%	65.20%	0.088	-0.061
b_bh_def	-2.37	0.20%	65.40%	0.116	-0.060

Again, both Team A and B's ability to kick goals (APOSCEBO, BPOSCEBO) are the most significant variables in the model with the largest absolute standardised estimates. The model that uses first half match data and pre-match ratings has an adjusted R^2 of

65.3% of the variation in the final margin. If the pre-match ratings are excluded from the model the adjusted R^2 is 60.2%. This first half model is an impressive result and reinforces the theory that as the match progresses the strength of models in predicting final margin improves. Such a tool would be a very useful aid for coaches at half time to help with their strategic decisions.

10.3.3 Using third quarter transition probabilities to predict final margin

This model builds on the information from the first half by including the transition probabilities for the third quarter. A model was investigated that included the pre-match ratings from above and one that excluded those ratings. Although two of the ratings were significant in the first model, the presence of these ratings in the model did not improve the amount of variation explained significantly, relative to the second model. Using stepwise selection in SAS, the second model for MOV_3, the margin of victory based on quarters one, two and three transition probabilities, used 12 of the 47 transition probabilities. Details of this quarter three model are presented in Table 10.6.

Table 10.6: Parameter estimates for third quarter model

Variable	Parameter Estimate	Partial R^2	Model R^2	P-value	Standardised Estimate
Intercept	-126.22			0.018	0
BPOSCEBO	-1537.64	39.90%	39.90%	<.0001	-0.555
APOSCEBO	1748.21	34.00%	73.90%	<.0001	0.610
BPOSBPOS	-142.72	2.60%	76.50%	<.0001	-0.199
APOSAPOS	249.67	1.60%	78.10%	<.0001	0.318
APOSABEH	454.71	1.60%	79.70%	<.0001	0.139
BPOSBBEH	-191.73	1.00%	80.70%	0.009	-0.065
CEBOAPOS	32.54	0.60%	81.30%	0.000	0.085
DISPBPOS	-53.58	0.40%	81.70%	0.045	-0.057
DISPAPOS	54.15	0.20%	81.90%	0.050	0.057
APOSDISP	105.80	0.10%	82.00%	0.059	0.127
BPOSAPOS	95.20	0.10%	82.10%	0.103	0.044
BUBOTHIN	-33.01	0.10%	82.20%	0.080	-0.040

As expected from the earlier models, Team A and B's ability to kick goals (APOSCEBO, BPOSCEBO) are the most significant variables in the model. The adjusted R^2 that this third quarter model explains for the final margin is 81.6% and this compares well to the model that included the pre-match ratings, which had an adjusted R^2 of 82.2%. Based on this result we would be extremely confident of being able to predict the final result of a match based on what has happened in the first three quarters of the game.

10.3.4 Using match transition probabilities to predict final margin

The purpose of this model is to show how powerful the predictive capabilities of the transition probabilities are towards the end of the match. Obviously, there is not a lot of value to be gained out of predicting the final margin of a match after it has been completed. However, in the final quarter of a match, an analytical tool to aid coaches with their decision making would be invaluable especially in close matches. Using the transition probabilities for the four quarters of the match in the same manner as the previous three models produces the final model for MOV_4 contained 22 of the 47 transition probabilities. The following eight transitions have not previously appeared in any of the previous models: ABEHDISP, BBEHAPOS, BPOSABEH, BUBOBBEH, BUBOBPOS, CEBOBPOS, THINAPOS and THINBPOS. Details of the final quarter model are contained in Table 10.7.

This model shows that the transition probabilities derived from the match are extremely good predictors of the final margin of a match. Hence, towards the end of a game when every decision becomes important, especially in close games, such a tool would assist coaches with how best to utilise their resources. For example, if a coach was able to identify crucial areas where his team was being outpointed, e.g. around the stoppages, he may then be able to make positional moves which will maximise the team’s chances of winning, or at least breaking even, at the transition in question.

Table 10.7: Parameter estimates for final quarter model

Variable	Parameter Estimate	Partial R^2	Model R^2	P-value	Standardised Estimate
Intercept	-148.79			<.0001	0
BPOSCEBO	-1771.38	46.20%	46.20%	<.0001	-0.576
APOSCEBO	1974.96	43.10%	89.30%	<.0001	0.635
APOSAPOS	298.40	2.20%	91.50%	<.0001	0.368
BPOSBPOS	-155.69	2.10%	93.60%	<.0001	-0.211
DISPAPOS	113.06	1.30%	94.90%	<.0001	0.108
BPOSBBEH	-166.41	0.90%	95.80%	<.0001	-0.051
APOSABEH	355.14	0.60%	96.40%	<.0001	0.099
THINBPOS	-15.99	0.40%	96.80%	<.0001	-0.037
DISPBPOS	-97.43	0.40%	97.20%	<.0001	-0.091
BUBOAPOS	16.64	0.30%	97.50%	<.0001	0.041
CEBOAPOS	16.42	0.30%	97.80%	<.0001	0.037
APOSDISP	145.44	0.30%	98.10%	<.0001	0.170
BPOSAPOS	150.01	0.20%	98.30%	<.0001	0.064
CEBOBPOS	-15.93	0.10%	98.40%	<.0001	-0.036
BBEHAPOS	5.42	0.00%	98.40%	0.001	0.023
THINAPOS	13.80	0.00%	98.40%	0.003	0.033
BUBOBBEH	-63.51	0.00%	98.40%	0.011	-0.017
BUBOBPOS	-9.08	0.00%	98.40%	0.012	-0.022
ABEHBPOS	-12.40	0.00%	98.40%	0.028	-0.054
ABEHDISP	-9.52	0.00%	98.40%	0.108	-0.039
APOSTHIN	95.94	0.00%	98.40%	0.130	0.011
BPOSABEH	299.79	0.00%	98.40%	0.149	0.010

10.3.5 Using regression models in a game environment

It has been shown in the second part of this chapter that the transition probabilities that occur in a match of AFL football are very good predictors of final margin, particularly as the game progresses. To highlight the expected use of such a model, the 2005 grand final will be looked at as a case study. The game was played between the Sydney Swans and West Coast Eagles with Sydney winning by four points. Table 10.8 displays the predicted margin using the pre-match model from Chapter 7 and the four models presented in this chapter. The actual margin column is the margin at that point of the game while the predicted margin is for the whole match.

Table 10.8: Expected margins for 2005 Grand Final

Model	Actual		Predicted	
	Winner	Margin	Winner	Margin
Ch. 7 Pre-match			West Coast	2
Qtr 1	Sydney	2	Sydney	7
Qtr 2	Sydney	20	Sydney	38
Qtr 3	Sydney	2	Sydney	1
Qtr 4	Sydney	4	Sydney	3

The pre-match model indicates that the Grand Final was always destined to be a tight one with the West Coast given a slight predicted advantage of two points. It is interesting to note that by quarter time, when Sydney led by two points, Sydney has become the predicted winner by seven points. At half time, Sydney held a big lead of twenty points and they were predicted to go on with it and win by 38 points. At this stage, such a model could be a useful tool for the coach of the Eagles to investigate the areas he could best target in order to rein Sydney back in. Target areas could be adjusted, such as stoppage set up or strategy for forward entry, with simulation used to gauge the effect.

10.4 Summary

It has been seen from the applications presented in this chapter that the global model could be utilised in a dynamic game environment with success. The first application was concerned with altering match events that have taken place to come up with different scenarios. These scenarios can then be simulated to gauge the effect that the events had on a team's probability of victory. Such an analytical tool could have wide and varied applications across the football industry as previously discussed. Furthermore, the zone model that will be presented in the next chapter could improve the analysis by paying regard to location on the field where the event takes place. Some of the examples presented in this chapter will be looked at again using the zone model, to see if there is any difference in the expected probabilities of victory for the global and zone models.

The second application presented in this chapter presents a powerful analytical tool for coaches in a game environment. This tool will assist coaches in their decision making to maximise the chance of winning. The regression models presented showed great promise in being able to predict the final margin of a match using the transition probabilities derived from the match, and in the case of the early quarters only, some pre-match ratings. It is hoped that CD will develop an interface for this application which can be used by the coaches on game day. Ideally, this application will allow coaches to adjust transition probabilities by varying degrees producing updated projections on the expected margin of victory. This would enable the coaches to identify key areas that they can address through strategic decisions to maximise their chances of victory.

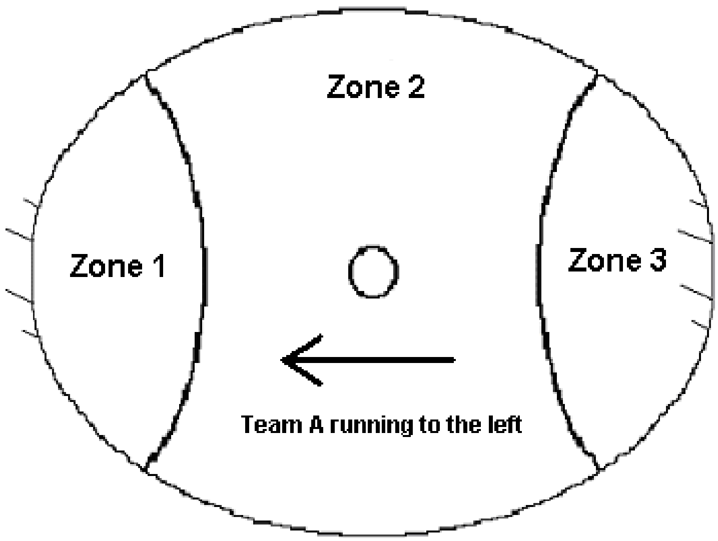
Chapter 11: A zone Markov process model to approximate Australian Rules football

11.1 Introduction

Chapter 8 introduced an eight state Markov process model which was shown to provide a close approximation to a game of AFL football. Various applications of this model were presented in Chapters 9 and 10 and this model was used by the Adelaide Football Club in the latter stages of the 2005 season with great effect. The drawback of this model is that it pays no regard to location on the field. The transition probabilities do not give as accurate an insight into the events of a match as they would if location was taken into account. For this reason a model has been developed with extra states that reflect the location of the transition in computing the probability.

An AFL football ground is broken into three zones, as can be seen in Figure 11.1, namely an attacking zone, a midfield zone and a defensive zone. Obviously one team's attacking zone is their opponent's defensive zone and vice-versa. With the richness of information that CD collects for each match, a model has been developed that utilises location information for each transition. This information has been used to improve upon the model presented in Chapter 8 and this model will be described in this Chapter.

Figure 11.1: AFL playing zones



11.2 Background to the zone model

In order to automate the computation of transition probabilities, certain assumptions have been put in place that hold whenever the model is used. The most important of these relates to the location coding of events and how they are implemented in the model. The locations of events within a match are coded by CD and this is done with reference to two values. One is known as the ‘physical zone’ and the other the ‘logical zone’. The physical zone takes on a numerical value of ‘1’, ‘2’ or ‘3’ and are delineated by the 50m lines marked on the playing surface at all AFL venues. The midfield is always ‘2’ and the other two values depend on the alignment of the playing arena. In almost all cases zone ‘1’ will refer to the end of the ground that is at the left of screen if one was watching the match on television. By default, zone ‘3’ is the area of the ground opposite to zone ‘1’ and on the right side of the screen if watching on TV. The other indicator is the logical zone and it is a character value of ‘F’, ‘M’ or ‘D’, which correspond to forward, midfield and defence respectively. This information is coupled with the physical zone to ascertain where the event takes place with reference to the layout of the ground. For the purposes of this model, there had to be consistency in the approach to how each zone was interpreted for either team. To this end, Team A's attacking zone is always Zone 1 with Team B's attacking zone always Zone 3. By definition, Team A's defensive area will be Zone 3 and Team B's defensive area is Zone 1. By taking this approach in the zone model, the changing of ends after each quarter creates no confusion for the interpretation of the transition probabilities. Unfortunately, the data available for the 2003 season did not contain the physical and logical zones making it impossible to accurately code the data for this model. As a result, the input data for this model is for the 185 games from season 2004 only.

A zonal approach to modeling a game of AFL football needs more states than the global model presented in Chapter 8. At first, one might expect there to be 24 states in the model (3 zones x 8 states), however there are six states that aren't possible. These states are a centre bounce in either zone 1 or 3, Team B scoring a behind in zone 1 or 2 and Team A scoring a behind in zone 3 or 2. This leaves 18 states and they are contained, with a brief description in Table 11.1

Table 11.1: Description of the 18 states contained in zone model

State	Description
BUBO1	Ball Up bounce in zone 1
THIN1	Throw in in zone 1
DISP1	Disputed ball in zone 1
APOS1	Team A possession in zone 1
BPOS1	Team B possession in zone 1
ABEH1	Team A behind
CEBO2	Centre Bounce after a goal or start of qtr
BUBO2	Ball Up bounce in zone 2
THIN2	Throw in in zone 2
DISP2	Disputed ball in zone 2
APOS2	Team A possession in zone 2
BPOS2	Team B possession in zone 2
BUBO3	Ball Up bounce in zone 3
THIN3	Throw in in zone 3
DISP3	Disputed ball in zone 3
APOS3	Team A possession in zone 3
BPOS3	Team B possession in zone 3
BBEH3	Team B behind

The input statistics for the zone model are exactly the same as the global model seen earlier, with the only difference being the inclusion of the zone in which the event took place. The SAS program used for the global model required some adjustments in order to derive the probabilities for the zone model; however, overall the process used for both models is fairly similar. One important difference between the models was the prospect of a player traversing across a zone without causing a transition. An example of this is a player taking an uncontested mark in defence and then running the ball into the midfield before disposing of the ball. This has the effect of increasing the number of transitions in the match under this model compared to the global model presented in Chapter 7. To highlight this, the mean number of transitions under this model is 880 for a match compared to 847, the mean number of transitions under the global model. This indicates that a zone approach to allocating transition probabilities involves an extra 33 transitions per match on average where the zones are crossed without a disposal.

11.3 The 18-state model

The main improvements expected from this model would be in some of the applications that have been put forward in the last two chapters. Applications where this occurs will be addressed later in this chapter and the comparison made with earlier analysis. We have already seen that the 8-state model provided an excellent approximation to an AFL game and it is not expected that the zone model will improve greatly upon this accuracy. Using similar techniques to those described in Chapter 8, the fit of the zone model on the 2004 season can be seen in Table 11.2

Table 11.2: Mean frequency error for each state and 95% confidence interval for mean error

State	Mean Error	Std Dev	Lower Bound	Upper Bound	Mean Squared Error
bubo1	-0.02	0.055	-0.03	-0.01	0.003
thin1	-0.03	0.061	-0.04	-0.02	0.005
disp1	-0.11	0.248	-0.15	-0.08	0.074
apos1	-0.17	0.347	-0.22	-0.12	0.149
bpos1	-0.64	0.765	-0.75	-0.53	0.995
abeh1	-0.06	0.115	-0.08	-0.04	0.017
cebo2	-0.14	0.160	-0.17	-0.12	0.045
bubo2	0.50	0.322	0.46	0.55	0.354
thin2	-0.05	0.074	-0.06	-0.04	0.008
disp2	-0.20	0.248	-0.24	-0.17	0.102
apos2	0.99	0.792	0.87	1.10	1.607
bpos2	0.87	0.773	0.76	0.99	1.354
bubo3	-0.03	0.082	-0.04	-0.02	0.008
thin3	-0.02	0.037	-0.02	-0.01	0.002
disp3	-0.09	0.208	-0.12	-0.06	0.051
apos3	-0.59	0.683	-0.69	-0.49	0.815
bpos3	-0.15	0.316	-0.20	-0.11	0.122
bbeh3	-0.05	0.122	-0.07	-0.04	0.017

Table 11.2 shows that the zone model that has been set up for approximating AFL is well constructed with all 18 states having a mean absolute error less than one. This confirms that the techniques that have been put in place are well founded and provide robust results. To further emphasise the accuracy of the zone model, chi-square goodness of fit tests have been performed similar to those done in Chapter 8. The results for the 18 states display a good fit with the lowest p-value of 0.63 being associated with the state Team B possession in zone 1. Taking these results into account, it can be safely assumed that the zone model presented in this chapter provides an adequate approximation of AFL football. The rest of this chapter will revisit some of the applications from Chapter 9 and 10 using the zone model.

11.4 Using simulation to investigate matches after their completion

The simulation process for the 18 state model is very similar to the 8 state simulation process. Adjustments had to be made to the simulation program to reflect the inclusion of the extra states and the progression from one state to another. As a result of including ten extra states, the computational time for the program is greatly increased when simulating matches 10,000 times. The new program allows for comparisons to be made between both the global and zone models.

11.4.1 Comparison of close match analysis from 2004

The first application investigated is the close matches that were analysed in section 4.2 of Chapter 9. As a result of the lack of data from 2003, only the seven matches that were analysed from 2004 can be compared. The transition matrices from the 18-state zone model were used in the simulation program and the number of transitions was derived from the zone model. As mentioned above, this number differed from the number of transitions used for the global model. Table 11.3 contains the information from Chapter 9 for the 2004 games whilst Table 11.4 contains the same information using the zone model. A margin comparison is made between the two models in Table 11.5 by subtracting the score of Team B from Team A's score.

Table 11.3: Global model analysis of 2004 close matches

Round	Team A	Team B	Actual Margin	Pr(A win) %	Pr(B Win) %	Pr(Draw) %
3	ES	WC	6	54.70	44.30	1.00
6	RI	HA	1	48.40	50.10	1.50
11	AD	CA	-4	44.30	54.40	1.30
22	CA	CO	1	49.40	49.60	1.00
EF	ME	ES	-5	41.60	57.20	1.20
PF	BL	GE	9	62.90	35.70	1.40
PF	PA	ST	6	63.30	35.60	1.20

Table 11.4: Zone model analysis of 2004 close matches

Round	Team A	Team B	Actual Margin	Pr(A win) %	Pr(B Win %	Pr(Draw) %
3	ES	WC	6	56.87	42.16	0.97
6	RI	HA	1	50.35	48.37	1.28
11	AD	CA	-4	42.69	55.98	1.33
22	CA	CO	1	47.93	50.99	1.08
EF	ME	ES	-5	42.85	56.02	1.13
PF	BL	GE	9	62.46	36.26	1.28
PF	PA	ST	6	61.62	37.11	1.27

Table 11.5: Comparison of zone model and global model

				Global Model			Zone Model		
Round	Team A	Team B	Actual Margin	Expected Margin	Winner	Error	Expected Margin	Winner	Error
3	ES	WC	6	6	ES	0.3	7	ES	-0.9
6	RI	HA	1	-1	HA	2.0	1	RI	0.5
11	AD	CA	-4	-4	CA	0.0	-5	CA	0.9
22	CA	CO	1	0	CO	0.9	-1	CO	2.0
EF	ME	ES	-5	-7	ES	1.6	-5	ES	0.5
PF	BL	GE	9	10	BL	-0.7	10	BL	-0.5
PF	PA	ST	6	11	PA	-5.1	10	PA	-4.0

The first thing that is noticeable from Table 11.4 is the error associated with the Port Adelaide/St. Kilda match. In both cases the expected margin exceeds the actual margin, indicating that St. Kilda were lucky to get as close as they did and that the scoreboard flattered them in the end. The global model had the correct winner in five of the seven matches and an average absolute error (AAE) of 1.5 points. The zone model had six of seven winners with the different result coming on the Richmond/Hawthorn game. The AAE for the zone model was 1.3 points. It seems from this result that the zone model does provide a slightly better approximation of matches after their completion than the global model from Chapter 9. This is not an overly surprising result as it was expected that a model that included location on the ground as a factor would perform better than a model that did not. It is reassuring to see just how well the global model does perform considering it does not include location on the ground. As an initial model, it has done extremely well in approximating AFL matches after their completion. The next section follows on from section 4.3 in Chapter 9, in which transition probabilities were adjusted in the hope of improving a team's chances of victory.

11.4.2 Adjusting transition probabilities to improve chances of victory

This application was seen in Chapter 9 and will be revisited here in a similar manner. Making adjustments using the global model meant that individual transitions could not be isolated for their importance. For instance, in section 4.3.1 of Chapter 9, six of St. Kilda's errors were removed and replaced by disposals that found the target. In removing these errors, no regard is paid to where they took place on the field. So, for instance, an error that occurs in St. Kilda's attacking zone may not be as crucial as an error that is committed in defence. To illustrate this, the Port Adelaide/St. Kilda match has been looked at, with the transition matrix for the match contained in Table 11.5 (due to size restrictions the matrix has been broken down by zone). By definition, Zone 1 is Port Adelaide's attacking zone as they are Team A and therefore Zone 3 is St. Kilda's attacking zone. With this in mind, the distribution of St. Kilda's possession coming out of defence is of most interest, and it can be seen that they handed the ball directly to Port Adelaide in their attacking zone 1.6% of the time. They turned it over to Port Adelaide in the midfield 3.2% of the time when coming out of defence. Whilst this error rate was below Port Adelaide's rate coming out of defence (13.0%), how much did these errors contribute to St. Kilda's defeat? To test this, the errors St. Kilda made have been converted into disposals that found the intended target and the resulting transition matrix has been used to simulate the match 10,000 times. From the previous section, we know that St. Kilda was a 37% chance of winning with an expected margin of defeat of 10 points. With the removal of three errors, as discussed above, St. Kilda becomes a 45% chance and the margin of defeat reduces to three points. This is an impressive result, given that in Chapter 9, six errors were removed and a behind was converted into a goal for St. Kilda, in order to give them a 52% chance of winning the match.

Table 11.6: Observed transition matrix: Port Adelaide v St. Kilda

state	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1
BUBO1	12.5%	0.0%	0.0%	12.5%	62.5%	0.0%
THIN1	14.3%	0.0%	0.0%	57.1%	28.6%	0.0%
DISP1	10.4%	12.5%	2.1%	45.8%	29.2%	0.0%
APOS1	0.0%	0.0%	8.6%	11.4%	17.1%	20.0%
BPOS1	1.6%	1.6%	16.1%	1.6%	35.5%	0.0%
ABEH1	0.0%	0.0%	20.0%	0.0%	70.0%	0.0%
CEBO2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
BUBO2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
THIN2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
DISP2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
APOS2	0.0%	0.0%	19.5%	2.5%	3.8%	2.5%
BPOS2	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%
state	**CEBO2**	**BUBO2**	**THIN2**	**DISP2**	**APOS2**	**BPOS2**
BUBO1	0.0%	0.0%	0.0%	0.0%	12.5%	0.0%
THIN1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
DISP1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
APOS1	37.1%	0.0%	0.0%	2.9%	0.0%	2.9%
BPOS1	0.0%	0.0%	0.0%	16.1%	3.2%	24.2%
ABEH1	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%
CEBO2	0.0%	22.6%	0.0%	3.2%	32.3%	41.9%
BUBO2	0.0%	11.1%	0.0%	5.6%	22.2%	61.1%
THIN2	0.0%	11.8%	5.9%	0.0%	35.3%	47.1%
DISP2	0.0%	9.3%	18.7%	5.3%	38.7%	28.0%
APOS2	0.6%	0.0%	0.0%	13.8%	53.5%	3.8%
BPOS2	0.5%	0.0%	0.5%	16.1%	5.4%	54.8%
BUBO3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
THIN3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
DISP3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
APOS3	0.0%	0.0%	2.2%	13.0%	21.7%	13.0%
BPOS3	26.7%	0.0%	0.0%	0.0%	0.0%	2.2%
BBEH3	0.0%	0.0%	0.0%	0.0%	22.2%	11.1%
state	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
CEBO2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
BUBO2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
THIN2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
DISP2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
APOS2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
BPOS2	0.0%	0.0%	11.8%	2.2%	7.5%	0.5%
BUBO3	25.0%	0.0%	0.0%	58.3%	16.7%	0.0%
THIN3	25.0%	0.0%	0.0%	50.0%	25.0%	0.0%
DISP3	16.2%	18.9%	2.7%	18.9%	43.2%	0.0%
APOS3	0.0%	2.2%	13.0%	34.8%	0.0%	0.0%
BPOS3	2.2%	0.0%	17.8%	6.7%	24.4%	20.0%
BBEH3	0.0%	0.0%	0.0%	66.7%	0.0%	0.0%

11.5 End game analysis for investigating different play scenarios

In Chapter 10 several scenarios were presented where pieces of play were altered in a match and the remainder of the game simulated to ascertain the effect of the piece of play. One of the drawbacks of this analysis was the fact that the simulation had no way of knowing where the play was starting from on the ground. For instance, in the third case study involving the West Coast Eagles, the play was deep in Essendon's attacking zone and any simulations that are run should be improved by factoring in this information. Therefore, the three scenarios from section 10.2.3 have been revisited here using the zone model. When play is simulated as is, we saw from Chapter 10 that Essendon had a 20.7% chance of winning, West Coast a 25.4% chance, and the draw was a 53.9% chance. Running the same scenario with the ball starting in state three, i.e. ball in dispute in Zone 1, the benefit of the zone model can be seen immediately with Essendon a 40.8% chance of winning, West Coast is a 15.6% chance and the draw is a 43.6% chance. These probabilities differ markedly from the probabilities derived under the global model.

In the second situation it was assumed Judd had managed to get off a handball when he had the opportunity to do so that hit a team mate and therefore the starting state is five, (West Coast possession in zone1). Using the global model, the relevant probabilities of victory were: Essendon was a 16.9% chance of winning, West Coast a 34.8% chance and the draw was a 48.3% chance. The zone model produced Essendon as a 17.8% chance of winning, West Coast was a 27.2% chance and the draw was a 55.0% chance. Although these probabilities are closer to the global model than the previous scenario, the advantage of the zone model can be seen in West Coast's probability of victory. The lower probability of the zone model reflects the fact that the West Coast has to take the ball from one end to the other to score and win the game.

Finally, the alternative situation is considered where Judd held on to the ball and forced play to start with a ball up in zone1. The probabilities associated with the global model were Essendon a 21.9% chance of winning, West Coast a 24.4% chance and the draw was a 53.7% chance of occurring. Simulating the remainder of the match starting from state one i.e. ball up in Zone 1, gives Essendon a 40.5% chance of winning with West Coast a 16.8% chance and the draw is a 42.7% chance. Again, the value of including the location on the ground is evident with either team's probability of victory markedly affected.

It is hoped that this type of application can be set up and used in the dynamic environment of the game, most probably by media outlets. During the game commentators could isolate passages of play and easily compute what the effect of the passage is on the chances of victory or what an alternative passage of play would yield. For this reason, the transition matrix up to the play in question, has been used in the examples given, whereas an after match analysis could be better approximated with the transition matrix for the entire match. In the examples given for this application, the passages of play analysed were extremely close to the end of the match and the transition matrix used would be very close to the post match matrix. However, for a situation where the play occurred in an earlier quarter, with a significant number of transitions left, the overall match matrix would provide the best approximation for the remainder of the match.

11.6 Using match transition probabilities from the zone model to predict final margin

The final section of this chapter revisits the regression models developed in the previous chapter for dynamic prediction within a match. The predictive capacities of the models in Chapter 10 were encouraging with the strength of the models improving as the match progressed. Similar analysis has been performed using the zone model and it is expected that, with the introduction of more variables, the models from Chapter 10 will be improved. Furthermore, it is expected that the extra advantage of a zone approach for prediction purposes within the game will be beneficial to the coaches. Instead of trying to improve performance in a match using limited transition probabilities, the coach will be able to isolate exact player performance based on position and target useful areas for improvement. The flow-on effect would be that a coach can quantify the effect of making changes in one area, simulating the performance effect in other areas of the ground. This level of detail was lacking from the global models of Chapter 10. The results presented for the four models, one for each quarter, are presented below, with the only difference in analysis to the last chapter being the reduced data set of 185 games from the 2004 season. The data set is still sufficiently large to allow for robust models.

11.6.1 Using first quarter transition probabilities from the zone model to predict final margin

It has already been shown that the play from the first quarter of an AFL game, when coupled with pre-match ratings for the competing teams, gives a reasonable indication as to how the match will progress. Breaking the play down according to location should provide a better indication of match trends. For instance, a team's ability to get the ball into their attacking zone and then use the ball efficiently in front of goal should become more apparent. By using the location probabilities it is expected that important trends will be able to be identified from earlier on in a match. Using the zone probabilities from the first quarter for matches from the 2004 season, a model for predicting final margin from 32 transition probabilities and four pre-match ratings was produced. These variables combined together to explain 58.6% of the variation in final margin and already the benefit of breaking play down by zone can be seen. The model using quarter one global transition probabilities explained only 39.6% of variation in final margin, so there has been a 19.0% improvement with the zone model. Details of the model for predicting MOV from the first quarter transitions and the pre-match ratings is presented in Table 11.7.

Table 11.7: Parameter estimates for first quarter model

Variable	Parameter Estimate	Partial R^2	Model R^2	P-Value	Standardised Estimate
Intercept	-230.89			<.0001	0
b_bh_def	-5.25	11.40%	11.40%	0.021	-0.131
a_gl_def	6.84	6.70%	18.10%	<.0001	0.266
BPOS2DISP1	994.22	3.90%	22.00%	0.010	0.138
BPOS2BPOS3	-160.07	3.70%	25.70%	0.004	-0.154
APOS2APOS1	289.30	3.50%	29.20%	<.0001	0.252
a_bh_att	8.28	3.40%	32.60%	0.000	0.206
BPOS3BBEH3	46.59	3.20%	35.80%	0.000	0.200
BPOS1DISP3	1525.57	2.20%	38.00%	0.000	0.185
BPOS1APOS1	98.52	2.20%	40.20%	0.065	0.095
THIN1APOS2	45.82	2.00%	42.20%	0.002	0.165
THIN1DISP1	56.74	1.70%	43.90%	0.004	0.147
BPOS3APOS2	-209.67	1.70%	45.60%	0.037	-0.109
BPOS3THIN3	172.22	1.50%	47.10%	0.087	0.090
a_bh_def	7.02	1.20%	48.30%	0.004	0.171
DISP3BPOS2	-420.44	1.20%	49.50%	0.003	-0.159
APOS1BUBO1	515.83	1.10%	50.60%	0.002	0.162
BUBO2APOS2	16.33	1.10%	51.70%	0.077	0.092
DISP3APOS3	38.13	1.00%	52.70%	0.002	0.180
BUBO1BPOS2	-50.96	1.00%	53.70%	0.049	-0.104
THIN1THIN1	64.61	1.00%	54.70%	0.019	0.118
BUBO2BUBO2	32.32	0.90%	55.60%	0.053	0.108
BUBO3APOS2	-64.11	0.90%	56.50%	0.019	-0.122
BUBO2APOS1	166.90	0.80%	57.30%	0.019	0.126
APOS3BPOS2	-106.59	0.80%	58.10%	0.007	-0.147
THIN1APOS1	9.30	0.80%	58.90%	0.135	0.078
THIN3THIN3	-90.55	0.80%	59.70%	0.008	-0.139
BUBO3BBEH3	41.18	0.80%	60.50%	0.035	0.110
BPOS1BUBO1	405.98	0.70%	61.20%	0.078	0.094
BUBO3APOS3	17.20	0.70%	61.90%	0.004	0.156
APOS3BPOS3	-231.64	0.70%	62.60%	0.001	-0.203
BBEH3DISP2	16.44	0.70%	63.30%	0.091	0.089
CEBO2BPOS2	-19.88	0.60%	63.90%	0.090	-0.089
APOS3APOS3	-38.13	0.60%	64.50%	0.044	-0.116
BUBO2BUBO1	-369.11	0.60%	65.10%	0.018	-0.125
DISP1BPOS1	-40.54	0.60%	65.70%	0.005	-0.150
THIN2DISP2	103.98	0.60%	66.30%	0.004	0.163

The improvement that the zone model shows over the global model is an expected result; however, what is interesting is that ability to kick goals (APOS2→CEBO2, APOS1→CEBO2, BPOS2→CEBO2 or BPOS3→CEBO2) is not a significant predictor in the model. The ability of either team to get the ball into their attacking zone (APOS2→APOS1 and BPOS2→BPOS3) is very important with both of these probabilities high up in the model for their significance. This will be further investigated in the next chapter when comments made by a football commentator regarding the overrated statistic of "inside 50s" are addressed.

11.6.2 Using second quarter transition probabilities from the zone model to predict final margin

The global model that used first half data explained 63.8% of the variation in final margin. Using the zone model data for the first half, the amount of variation explained by the pre-match ratings and transition probabilities rises to 82.7%, which is quite an impressive result. The regression equation for this model contains 54 variables and the details of the model are contained in Table 11.8.

Table 11.8: Parameter estimates for second quarter model

Variable	Parameter Estimate	Partial R^2	Model R^2	P-Value	Standardised Estimate
Intercept	-85.35			0.071	0
APOS2APOS1	548.37	9.60%	9.60%	<.0001	0.366
BPOS2BPOS3	-110.40	7.80%	17.40%	0.039	-0.080
a_gl_def	1.70	5.90%	23.30%	0.115	0.066
DISP3APOS3	57.52	4.80%	28.10%	<.0001	0.172
THIN3APOS3	37.00	2.60%	30.70%	<.0001	0.302
BPOS3CEBO2	-102.71	2.20%	32.90%	<.0001	-0.272
APOS3APOS2	65.35	2.00%	34.90%	0.001	0.132
APOS1CEBO2	127.71	2.00%	36.90%	<.0001	0.313
b_gl_def	-4.35	1.90%	38.80%	<.0001	-0.190
THIN2APOS3	-118.64	1.90%	40.70%	0.031	-0.075
a_bh_att	4.40	1.60%	42.30%	0.007	0.109
BPOS2APOS3	326.65	1.60%	43.90%	<.0001	0.165
BPOS1BPOS2	-107.66	1.50%	45.40%	<.0001	-0.219
THIN1DISP1	79.23	1.30%	46.70%	0.001	0.126
BUBO2BUBO3	454.57	1.20%	47.90%	<.0001	0.151
DISP1BPOS1	-88.80	1.20%	49.10%	<.0001	-0.258
APOS2APOS2	81.78	1.10%	50.20%	0.005	0.120
BPOS2THIN2	1249.83	1.10%	51.30%	<.0001	0.194
THIN1THIN1	55.40	0.90%	52.20%	0.002	0.111
DISP2APOS3	1051.97	0.90%	53.10%	<.0001	0.165
CEBO2BUBO2	74.47	0.90%	54.00%	<.0001	0.178
THIN2THIN1	828.37	0.90%	54.90%	0.000	0.140
BPOS2CEBO2	-728.61	0.70%	55.60%	0.004	-0.103
APOS3THIN2	-520.57	0.60%	56.20%	<.0001	-0.163

Table 11.8: Parameter estimates for second quarter model (cont.)

Variable	Parameter Estimate	Partial R^2	Model R^2	P-Value	Standardised Estimate
a_bh_def	3.40	0.60%	56.80%	0.037	0.083
THIN1ABEH1	261.93	0.60%	57.40%	0.002	0.111
THIN3BBEH3	158.98	0.60%	58.00%	<.0001	0.145
APOS3BUBO3	847.26	0.60%	58.60%	0.002	0.116
DISP1DISP2	-366.56	0.60%	59.20%	0.001	-0.118
BUBO2BPOS1	320.78	0.60%	59.80%	<.0001	0.154
THIN2DISP1	482.42	0.60%	60.40%	0.055	0.073
APOS2THIN2	-926.03	0.60%	61.00%	0.001	-0.131
BUBO1APOS1	20.56	0.50%	61.50%	<.0001	0.162
BPOS1ABEH1	-528.76	0.50%	62.00%	0.004	-0.099
DISP1BUBO1	-90.95	0.50%	62.50%	0.000	-0.157
BUBO2DISP2	-109.77	0.50%	63.00%	0.000	-0.143
DISP3BUBO3	48.59	0.50%	63.50%	0.032	0.082
BPOS2APOS1	-1574.12	0.40%	63.90%	0.008	-0.094
THIN1APOS2	35.52	0.40%	64.30%	0.007	0.105
BUBO2BUBO2	-49.41	0.40%	64.70%	0.002	-0.131
APOS1BUBO2	-1315.39	0.40%	65.10%	0.001	-0.128
THIN3BPOS2	51.71	0.40%	65.50%	0.000	0.128
THIN1THIN2	180.54	0.40%	65.90%	0.004	0.102
APOS2BPOS3	-1209.05	0.40%	66.30%	0.053	-0.069
BUBO2BPOS2	-23.12	0.40%	66.70%	0.013	-0.097
BUBO2BPOS3	-213.87	0.40%	67.10%	0.014	-0.099
b_gl_att	-3.63	0.30%	67.40%	0.003	-0.124
DISP1BPOS2	-230.24	0.30%	67.70%	0.053	-0.068
APOS2THIN1	1470.53	0.30%	68.00%	0.051	0.073
BUBO3BUBO3	26.02	0.30%	68.30%	0.027	0.082
THIN1BUBO2	-98.68	0.30%	68.60%	0.058	-0.065
BPOS1THIN2	-327.12	0.30%	68.90%	0.001	-0.125
THIN2APOS1	-132.37	0.20%	69.10%	0.018	-0.087
BPOS2DISP1	584.25	0.20%	69.30%	0.121	0.056

The second quarter model has seen the inclusion of each team's ability to score goals from their attacking zone as highly significant. Furthermore, the defensive rating for each team in terms of goals they concede is highly significant as well. The upside of this model is the inclusion of many transition probabilities. Any dynamic application that computed match margins from somewhere within the game would give plenty of opportunity for coaches to make changes that they knew could have an effect on these probabilities.

11.6.3 Using third quarter transition probabilities from the zone model to predict final margin

By the end of quarter three, the global model was explaining 81.6% of the variation in final margin. By this stage the pre-match ratings offer little to the model and are not available for selection. Following this logic, the pre-match ratings have been omitted from the quarter three data set as well. The zone model that uses the transition probabilities from the first three quarters explains 92.5% of the variation in final margin, which is encouraging. The details of the third quarter model are contained in Table 11.9.

Table 11.9: Parameter estimates for third quarter model

Variable	Parameter Estimate	Partial R^2	Model R^2	P-Value	Standardised Estimate
Intercept	-103.12			<.0001	0
APOS2APOS1	397.03	21.50%	21.50%	<.0001	0.234
BPOS2BPOS3	-498.31	17.30%	38.80%	<.0001	-0.316
BPOS3CEBO2	-114.01	7.10%	45.90%	<.0001	-0.237
APOS1CEBO2	136.06	6.50%	52.40%	<.0001	0.280
BPOS1APOS1	179.85	3.50%	55.90%	0.000	0.097
DISP1BPOS1	-120.33	2.60%	58.50%	<.0001	-0.282
DISP3APOS3	136.54	2.20%	60.70%	<.0001	0.347
DISP1BPOS2	-500.86	1.80%	62.50%	<.0001	-0.124
BPOS1BPOS2	-59.75	1.80%	64.30%	0.000	-0.103
APOS3APOS2	59.09	1.80%	66.10%	0.000	0.099
BPOS2CEBO2	-1529.63	1.70%	67.80%	<.0001	-0.160
THIN3BPOS3	-31.71	1.60%	69.40%	<.0001	-0.230
CEBO2APOS2	56.65	1.20%	70.60%	<.0001	0.139
BPOS1DISP3	1594.67	1.10%	71.70%	<.0001	0.105
BPOS2BPOS2	-98.44	1.00%	72.70%	<.0001	-0.138
BPOS2DISP3	-155.60	1.00%	73.70%	0.000	-0.098
DISP3BUBO3	94.49	0.90%	74.60%	<.0001	0.143
APOS2CEBO2	1242.18	0.80%	75.40%	<.0001	0.137
THIN2BPOS3	-276.20	0.70%	76.10%	<.0001	-0.162
BPOS1DISP1	261.53	0.70%	76.80%	<.0001	0.215
DISP2APOS2	113.44	0.60%	77.40%	<.0001	0.148
BPOS3BPOS3	-61.47	0.60%	78.00%	<.0001	-0.116
BUBO2BPOS3	-186.80	0.60%	78.60%	<.0001	-0.104
DISP1DISP2	-430.69	0.40%	79.00%	0.000	-0.094
APOS1DISP1	-54.32	0.40%	79.40%	0.003	-0.084
THIN3THIN3	-51.12	0.40%	79.80%	<.0001	-0.097
CEBO2DISP2	101.98	0.40%	80.20%	0.000	0.102
APOS3BPOS2	-145.34	0.40%	80.60%	<.0001	-0.138

Table 11.9: Parameter estimates for third quarter model (cont.)

Variable	Parameter Estimate	Partial R^2	Model R^2	P-Value	Standardised Estimate
APOS2DISP1	426.67	0.40%	81.00%	<.0001	0.276
BBEH3THIN3	577.03	0.40%	81.40%	<.0001	0.140
BUBO2BPOS1	190.02	0.40%	81.80%	0.001	0.085
APOS2APOS2	164.61	0.40%	82.20%	<.0001	0.217
BUBO3BBEH3	17.05	0.40%	82.60%	0.032	0.054
APOS1THIN1	-370.80	0.40%	83.00%	0.001	-0.079
ABEH1APOS2	83.26	0.30%	83.30%	<.0001	0.153
BPOS1APOS2	150.94	0.30%	83.60%	<.0001	0.134
THIN3APOS2	-27.24	0.30%	83.90%	0.005	-0.069
BUBO1ABEH1	-48.57	0.30%	84.20%	0.002	-0.079
THIN1APOS1	9.45	0.30%	84.50%	0.017	0.058
THIN2BPOS1	-189.63	0.30%	84.80%	0.000	-0.101
DISP3THIN3	66.37	0.30%	85.10%	0.001	0.101
DISP1DISP1	-132.42	0.30%	85.40%	<.0001	-0.120
BPOS2DISP1	1706.90	0.20%	85.60%	<.0001	0.132
BPOS2ABEH1	6057.65	0.20%	85.80%	0.012	0.061
THIN2THIN2	-118.09	0.20%	86.00%	<.0001	-0.148
APOS2ABEH1	-415.46	0.20%	86.20%	0.006	-0.073
APOS3BBEH3	-506.62	0.20%	86.40%	0.003	-0.070
BUBO3THIN2	32.05	0.20%	86.60%	0.015	0.059
THIN2DISP1	-1571.76	0.20%	86.80%	<.0001	-0.227
CEBO2BPOS1	378.71	0.20%	87.00%	0.053	0.047
ABEH1BPOS2	-26.24	0.10%	87.10%	0.001	-0.089
BUBO2APOS3	-149.09	0.10%	87.20%	0.001	-0.086
DISP2BPOS3	473.91	0.10%	87.30%	0.037	0.055
APOS2BUBO1	1949.34	0.10%	87.40%	0.034	0.056
BUBO3APOS3	-9.48	0.10%	87.50%	0.004	-0.074
THIN3THIN2	43.22	0.10%	87.60%	0.003	0.074
THIN3BUBO3	11.62	0.10%	87.70%	0.071	0.048
BUBO1DISP1	-20.91	0.10%	87.80%	0.081	-0.046
BPOS3DISP2	183.27	0.10%	87.90%	0.008	0.068
BUBO2THIN1	2703.34	0.10%	88.00%	<.0001	0.199
BUBO2BUBO2	-27.21	0.10%	88.10%	0.011	-0.062
THIN2BPOS2	-15.25	0.10%	88.20%	0.063	-0.050
DISP2DISP3	-950.82	0.10%	88.30%	0.017	-0.057
BPOS2THIN3	4071.49	0.10%	88.40%	0.000	0.109
BUBO3BPOS2	-45.36	0.10%	88.50%	0.011	-0.060
BUBO1THIN2	91.70	0.10%	88.60%	0.129	0.041

By dropping out the pre-match ratings from the analysis, the ability of teams to get the ball into their attacking zone and then convert it into six points become the most significant predictors in the model.

11.6.4 Using match transition probabilities from the zone model to predict final margin

This model has been included again for completeness and to indicate how accurate predictions can be obtained using the zone model data, towards the end of a match. It was seen in the last chapter that 98.5% of the variation in final margin is explained with the inclusion of last quarter data. The amount of variation explained by the zone model is slightly higher at 98.7% and this is consistent with the earlier models in this chapter that outperform their global counterpart. The final quarter model is made up of 84 variables with the details of these variables included in Table 11.9

Table 11.10: Parameter estimates for final quarter model

Variable	Parameter Estimate	Partial R^2	Model R^2	P-Value	Standardised Estimate
Intercept	-267.19			<.0001	0
APOS2APOS1	780.64	28.70%	28.70%	<.0001	0.428
BPOS2BPOS3	-515.97	17.00%	45.70%	<.0001	-0.289
BPOS3CEBO2	-194.21	8.50%	54.20%	<.0001	-0.340
APOS1CEBO2	229.05	7.30%	61.50%	<.0001	0.380
APOS2DISP2	204.03	4.90%	66.40%	<.0001	0.200
APOS1APOS1	69.85	4.10%	70.50%	<.0001	0.105
BPOS2APOS2	149.87	3.50%	74.00%	<.0001	0.078
DISP3APOS3	35.09	2.00%	76.00%	<.0001	0.080
THIN1APOS1	18.44	1.60%	77.60%	<.0001	0.099
DISP1BPOS1	-57.16	1.60%	79.20%	<.0001	-0.116
APOS3BPOS3	-153.80	1.50%	80.70%	<.0001	-0.076
BPOS1APOS1	149.74	1.40%	82.10%	<.0001	0.072
CEBO2APOS2	27.27	1.10%	83.20%	<.0001	0.057
DISP2BPOS2	-38.67	0.90%	84.10%	0.012	-0.051
THIN3BPOS3	-12.67	0.80%	84.90%	<.0001	-0.079
APOS3APOS2	18.57	0.80%	85.70%	0.038	0.029
BPOS2CEBO2	-1462.08	0.70%	86.40%	<.0001	-0.134
BPOS2APOS3	207.38	0.70%	87.10%	<.0001	0.073
BPOS3BPOS3	-55.04	0.70%	87.80%	<.0001	-0.087
APOS2CEBO2	1059.37	0.50%	88.30%	<.0001	0.096
APOS2BPOS1	208.23	0.40%	88.70%	<.0001	0.070
BUBO1APOS1	5.81	0.40%	89.10%	0.010	0.037
ABEH1BPOS2	-20.22	0.30%	89.40%	<.0001	-0.060
THIN2BUBO1	125.93	0.30%	89.70%	0.000	0.043
DISP1APOS1	47.00	0.30%	90.00%	<.0001	0.095
DISP3BPOS3	-53.28	0.20%	90.20%	<.0001	-0.101
BPOS1DISP1	107.88	0.20%	90.40%	<.0001	0.082
THIN2APOS2	15.78	0.20%	90.60%	0.000	0.047
BBEH3BPOS2	-37.19	0.20%	90.80%	<.0001	-0.064

Table 11.10: Parameter estimates for final quarter model (cont.)

Variable	Parameter Estimate	Partial R^2	Model R^2	P-Value	Standardised Estimate
BUBO2APOS2	14.50	0.20%	91.00%	0.002	0.037
THIN2BPOS3	-138.41	0.20%	91.20%	<.0001	-0.068
APOS2APOS2	327.96	0.20%	91.40%	<.0001	0.404
BPOS1BPOS2	-46.14	0.20%	91.60%	<.0001	-0.072
APOS3DISP3	-57.83	0.20%	91.80%	0.001	-0.048
BPOS1APOS2	75.34	0.20%	92.00%	<.0001	0.059
DISP2APOS2	113.48	0.10%	92.10%	<.0001	0.139
BPOS2DISP3	-197.99	0.10%	92.20%	<.0001	-0.117
APOS2DISP1	459.54	0.10%	92.30%	<.0001	0.279
THIN3THIN3	-33.56	0.10%	92.40%	<.0001	-0.066
DISP3THIN3	-23.74	0.10%	92.50%	0.013	-0.036
APOS1BUBO1	94.80	0.10%	92.60%	0.108	0.018
CEBO2APOS3	-513.76	0.10%	92.70%	0.001	-0.038
BUBO1BUBO2	31.11	0.10%	92.80%	0.002	0.039
BUBO1BPOS2	-41.60	0.10%	92.90%	<.0001	-0.050
APOS1DISP2	145.03	0.10%	93.00%	0.002	0.034
CEBO2DISP2	56.96	0.10%	93.10%	0.000	0.052
BUBO2APOS1	120.13	0.10%	93.20%	<.0001	0.048
DISP2BUBO2	48.91	0.10%	93.30%	0.011	0.042
BPOS3BPOS2	-202.68	0.10%	93.40%	0.004	-0.036
ABEH1DISP1	-37.95	0.10%	93.50%	<.0001	-0.063
CEBO2DISP3	224.69	0.10%	93.60%	0.001	0.042
BUBO2BPOS3	-56.14	0.10%	93.70%	0.036	-0.023
THIN2BUBO3	-169.77	0.10%	93.80%	0.000	-0.042
BPOS2BPOS1	5415.27	0.10%	93.90%	<.0001	0.073
THIN1ABEH1	18.52	0.00%	93.90%	0.008	0.033
BPOS1BUBO1	447.39	0.00%	93.90%	0.005	0.031
ABEH1DISP2	-8.83	0.00%	93.90%	0.063	-0.023
ABEH1APOS2	28.22	0.00%	93.90%	0.000	0.046
BUBO2BUBO3	141.15	0.00%	93.90%	0.003	0.033
DISP2APOS1	370.22	0.00%	93.90%	0.002	0.037
BPOS2BUBO2	402.66	0.00%	93.90%	0.017	0.032
BPOS3APOS3	65.34	0.00%	93.90%	<.0001	0.071
BBEH3BPOS3	-61.74	0.00%	93.90%	0.001	-0.047
BUBO1BPOS1	-5.56	0.00%	93.90%	0.021	-0.034
THIN1BPOS1	4.63	0.00%	93.90%	0.077	0.025
APOS1BPOS1	-47.75	0.00%	93.90%	0.000	-0.049
ABEH1THIN1	-174.65	0.00%	93.90%	0.022	-0.030
BUBO2DISP2	-22.73	0.00%	93.90%	0.052	-0.023
DISP2DISP2	85.96	0.00%	93.90%	0.002	0.054
DISP2APOS3	412.84	0.00%	93.90%	0.000	0.044
APOS2BUBO1	1381.95	0.00%	93.90%	0.009	0.029
BPOS2BPOS2	-121.41	0.00%	93.90%	<.0001	-0.162
BUBO3THIN2	16.93	0.00%	93.90%	0.073	0.021
BUBO3BBEH3	13.43	0.00%	93.90%	0.024	0.030
APOS3BPOS2	-40.21	0.00%	93.90%	0.005	-0.035
APOS3THIN3	-183.64	0.00%	93.90%	0.011	-0.031

Table 11.10: Parameter estimates for final quarter model (cont.)

Variable	Parameter Estimate	Partial R^2	Model R^2	P-Value	Standardised Estimate
BBEH3BUBO3	-81.92	0.00%	93.90%	0.003	-0.035
DISP1DISP2	-178.50	0.00%	93.90%	0.009	-0.036
THIN2THIN1	-130.31	0.00%	93.90%	0.012	-0.027
THIN2THIN2	19.06	0.00%	93.90%	0.086	0.020
APOS2THIN2	-353.48	0.00%	93.90%	0.005	-0.035
APOS2APOS3	1528.05	0.00%	93.90%	0.004	0.032
APOS3APOS1	560.00	0.00%	93.90%	0.069	0.019
BBEH3APOS2	7.44	0.00%	93.90%	0.058	0.022

It is appropriate here to highlight that the diagnostic measures of appropriateness for all four models suggest that the linear approach is satisfactory and justified. Plots of the residual values against the predicted values were randomly distributed for all four models. Furthermore, normal probability plots of the residuals form a nearly linear pattern. Coupling this information together we can assume that the model assumptions are satisfied in each case.

11.7 Summary

This chapter has shown the advantage of breaking transition probabilities up according to location. It was always believed that the location of the ball on the ground would be an important feature in terms of being able to accurately approximate AFL matches. This has been reinforced with the introduction of the extra states in the zone model and the excellent approximations that it provides. The errors associated with each state were comparable to the global model, which is reassuring, indicating that the systems and processes in place to handle the movement of the ball across the 50m arcs are sound. The accuracy of the zone model was shown in section 11.4 to be slightly better in correctly approximating results than the global model, albeit on a very small sample. The zone model was also shown to have wider application in analysing matches after their completion in section 11.5, where pieces of play could be isolated and their numerical effect on a team's chances of victory more accurately identified. The final part of this chapter revisited the dynamic, 'in-game' regression models that rely on the data that has happened in the match. These zone models clearly outperformed the global data models from the last chapter with high amounts of the variation in final margin of victory being explained even from first quarter data. This suggests that the zone model can be used in-game by coaches to identify key areas that they either need to keep dominating or, alternatively, should look to be improving to maximise their chances of winning the match. It is also believed that such a model could be used by media outlets to gain accurate prediction of what the match outcome will be, from very early on in the game. The next chapter will present some applications that are possible only with the zone model.

Chapter 12: Applications of 18 state zone model to AFL football

12.1 Introduction

As already seen, there is no reason why the zone model that was introduced in the previous chapter can't be used for all of the applications that were highlighted for the global model in Chapter's 9 and 10. Of more interest are applications of the zone model that were not possible using the global model and this chapter will present these applications. The introduction of location into the model enables analysis revolving around match strategy such as player movement or style of play. Markov models have been used with great success for investigating strategy issues in a number of sports. The optimum batting line-up for a baseball team was investigated through simulation using a Markov model, maximising the expected run return for each line-up to ascertain which was the best order (Bukiet, Harold and Palacious, 1997). Also in baseball, analysis was done on pinch hitting strategy, substitution of pitchers and substituting batters according to handedness (Hirotsu, 2002). The same author also looked at soccer and used a Markov model for ascertaining the optimum time for making a substitution. Other strategic sporting decisions have been analysed using mathematical techniques other than Markov models such as when to pull the goalie in ice hockey (Morrison, 1976), the optimal play selection in first down and goal situations for American football (Boronico, 2000) and whether rushing a behind in AFL football can increase a team's chances of victory (Clarke and Norman, 1998).

The game of AFL lends itself to analysis using the zone model for a number of issues. Much has been made in the past few seasons about how the style of play in the competition has changed. Critics are concerned about the speed of the game brought about by the introduction of unlimited interchange (Blainey, 2003) or concerned that the style of play in the modern game causes too many stoppages, which was looked at in Chapter 9. This model can be used to identify styles of play, (a team may kick the ball long into their attacking zone or prefer to chip the ball around waiting for an open team mate to present) and to compare one style to another. Furthermore, analysis can be carried out on playing styles to ascertain whether they are advantageous to a team's chances of victory. These types of investigations will be analysed in this Chapter.

12.2 Importance of inside 50s as an AFL match statistic

Champion Data, as part of their match information, record the number of times each team gets the ball into their attacking zone. This statistic, known as the number of inside 50s, is included in the zone model by any transition that starts in zone 2 and ends in zone 1 (for team A) or zone 3 (for team B). A newspaper article questioned the importance of this statistic for predicting success in terms of giving an indication as to who was more likely to win (Sheahan, 2005). Former AFL player, Wayne Carey, described the statistic as "irrelevant" and "the worst stat in footy". Several matches were mentioned in the article to try and highlight the "irrelevance" of the inside 50 statistic and these matches have been analysed using the zone model to gain a better understanding of how the teams distributed the ball into their attacking zone. Analysis will also be done using a competition transition matrix, and adjustments made to attacking zone entry values, to gauge the importance of this statistic to margin and likelihood of victory.

12.2.1 Selected games from 2005 season

The games from the article will be looked at individually using the zone model. Particular reference will be made to how the ball was distributed into the attacking zone and then how it was used once in there. The games looked at are the round 11 match between the Kangaroos and Melbourne, and the round 12 games played between Brisbane and Carlton, Hawthorn and St. Kilda, and Kangaroos and Richmond.

12.2.1.1 Kangaroos v Melbourne

In this round 11 match played at Docklands, Melbourne beat the Kangaroos by 36 points, 17.14.116 to 11.14.80. The official inside 50s favoured the Kangaroos 55 to 51 and the argument is that if inside 50s were so important, then the Kangaroos would have won the game or been much closer to Melbourne instead of losing by six goals. After coding the match using the zone model, the forward 50m entries were 55-50, favouring the Kangaroos, but the secret lies in how the ball was distributed into the attacking 50s. Table 12.1 contains the distribution profile for each team highlighting how they disposed of the ball into the forward zone when they had possession in the midfield.

Table 12.1: Distribution of ball into attacking zones, Kangaroos v Melbourne, round 11 2005

Team	Variable	BUBO	THIN	DISP	APOS	BPOS	BEHI	CEBO	Total
Kangaroos	Count	0	0	29	11	10	4	1	55
Kangaroos	%	0.0	0.0	52.7	20.0	18.2	7.3	1.8	100.0
Melbourne	Count	0	0	21	6	19	3	0	49
Melbourne	%	0.0	0.0	42.9	12.2	38.8	6.1	0.0	100.0

It must be noted that one of Melbourne's inside 50s came from the ball being in dispute in Zone 2 and crossing the 50m line where the Kangaroos took possession, hence Melbourne's total is 49. From Table 12.1 it is clear that Melbourne dominated because of their far superior ability to hit targets. The Kangaroos were guilty of bombing the ball in (DISP = 52.7%) and turning it over easily to their opponents (BPOS = 18.2%). With Melbourne kicking a goal 36.6%[4] of the time they had possession in their attacking zone, it is easy to see why they beat the Kangaroos by six goals.

[4] Derived from match transition probability matrix

12.2.1.2 Carlton v Brisbane Lions

This round 12 match was also played at Docklands with the Lions running out comfortable winners by 58 points. The official inside 50s had each team entering their attacking zone 48 times and again the discrepancy in score is seen to add weight to Carey's argument. The coding of the data using the zone model had the forward 50 entries, 46 to 47 in Brisbane's favour. Table 12.2 contains the distribution profile for both teams.

Table 12.2: Distribution of ball into attacking zones, Carlton v Brisbane, round 12 2005

Team	Variable	BUBO	THIN	DISP	APOS	BPOS	BEHI	CEBO	Total
Carlton	Count	0	0	23	14	5	1	0	43
Carlton	%	0.0	0.0	53.5	32.6	11.6	2.3	0.0	100.0
Brisbane	Count	0	0	14	3	17	7	4	45
Brisbane	%	0.0	0.0	31.1	6.7	37.8	15.6	8.9	100.0

Carlton's additional three entries were not a result of their disposal and Brisbane had two entries as a result of Carlton's disposal. Again, the ability of the winning team to find the target in their forward 50 as well as limiting the number of times they turned it over to their opponent is evident. Brisbane dominated Carlton in these areas and was also able to kick four goals from the midfield, whilst Carlton kicked none. The fact that Carlton were forced to bomb the ball in long (DISP = 53.5%) coupled with their inability to beat Brisbane at winning the ball from dispute in Zone 1 highlights why Carlton found it so hard to score, even though they matched Brisbane at getting the ball into attack.

12.2.1.3 Hawthorn v St. Kilda

This was a match where the losing side had more forward entries than the winning side. St. Kilda won by 46 points yet they had less entries, 49, than Hawthorn, 52. After coding the match using the zone model, the forward entries came out as 51 to 49 in Hawthorn's favour. The profile of ball distribution is contained in Table 12.3.

Table 12.3: Distribution of ball into attacking zones, Hawthorn v St. Kilda, round 12 2005

Team	Variable	BUBO	THIN	DISP	APOS	BPOS	BEHI	CEBO	Total
Hawthorn	Count	0	0	24	13	12	0	0	49
Hawthorn	%	0.0	0.0	49.0	26.5	24.5	0.0	0.0	100.0
St. Kilda	Count	0	0	20	12	12	2	2	48
St. Kilda	%	0.0	0.0	41.7	25.0	25.0	4.2	4.2	100.0

In this game there is no clear answer as to why St. Kilda dominated the scoreboard as they did. Certainly the distribution from each team into attack was very similar with St. Kilda having a slight advantage in being able to score from outside 50. Both teams 'bombed' the ball into their attacking zone at roughly the same rate (49.0% c.f. 41.7%) and this is where St. Kilda's dominance can be traced to. In Hawthorn's attacking zone, St. Kilda extracted the ball from dispute at a rate of 55.2% compared to 27.6% for Hawthorn. Similarly, in St. Kilda's attacking zone, when the ball was in dispute St. Kilda were winning it 46.3% and Hawthorn only 28.6% of the time[5]. It must be assumed from this that St. Kilda was able to repel Hawthorn's forward charges due to their dominance at getting the ball out of dispute in defence, and at the other end of the ground they were able to convert their dominance at getting the ball out of dispute into goals.

[5] Derived from match transition probability matrix

12.2.1.4 Kangaroos v Richmond

In the final game used in the article, the Kangaroos beat Richmond by 29 points with 14 less inside 50 entries. Having coded the match using the zone model, the Kangaroos had 46 inside 50s to Richmond's 57. The profile of each team's ball distribution into attack is contained in Table 12.4.

Table 12.4: Distribution of ball into attacking zones, Kangaroos v Richmond, round 12 2005

Team	Variable	BUBO	THIN	DISP	APOS	BPOS	BEHI	CEBO	Total
Kangaroos	Count	0	0	21	17	5	0	3	46
Kangaroos	%	0.0	0.0	45.7	37.0	10.9	0.0	6.5	100.0
Richmond	Count	0	0	22	10	20	3	0	55
Richmond	%	0.0	0.0	40.0	18.2	36.4	5.5	0.0	100.0

Like the previous match between Hawthorn and St. Kilda, there is no clear indication from the ball distribution profile why the Kangaroos were easy victors. In their favour was the ability to kick long goals from outside 50m. Aside from this, the profiles are very similar. Again, deeper investigation has to be carried out on what happened when the ball was in the attacking zones to see why there was such a discrepancy on the scoreboard. In this match, Richmond did better than the Kangaroos at getting the ball from dispute in the Kangaroos attacking zone. In Richmond's attacking zone the teams broke even in this area. The lopsided scoreboard can only be put down to Richmond's inaccuracy both from inside and outside 50. Table 12.4 showed that the Kangaroos kicked three goals from beyond 50m whilst Richmond could manage only three behinds. When the ball got into the attacking zones, the Kangaroos kicked a goal 40.0% of the time and a behind only 20.0% of the time, whereas, Richmond could manage a goal only 16.8% of the time and a behind 30.4% of the time. This is a clear case where the relationships in the game are relatively equal; however, poor conversion by one side has led to a comprehensive defeat. This type of scenario would be an excellent situation to be simulated with the conversion rates evened up to gauge the effect that the randomness of goal kicking plays in the final result.

12.2.2 Summary

In investigating the importance of inside 50s to a team's chance of victory, it has been found that the numbers by themselves can be misleading and sometimes deeper investigation is required. In Sheahan's article, Wayne Carey was quoted as saying "it's how it comes into the forward 50 that matters", and the analysis done above has shown this to not always be the case. In the Kangaroos/Melbourne and Carlton/Brisbane games this was certainly the case, with the winning sides ability to find team mates unopposed in their attacking zone the major contributor to easy victories. In the last two games that were analysed, the delivery into the attacking zone for each of the teams was very similar. In the Hawthorn/St. Kilda match the distribution profile gave no hints as to why St. Kilda won easily but deeper investigation showed that it was the Saint's ability to dominate when the ball was in dispute at both ends of the ground that saw them win comfortably. For the Kangaroos/Richmond game the lopsided nature of the scoreboard can only be explained by Richmond's inaccuracy and this is a clear case of the random effects that are present in football. Without their poor kicking, this game would have been much closer with a Richmond victory not out of the question, making the article's argument redundant for this match. This analysis has shown that sometimes, by itself, the inside 50 can be slightly misleading and other analysis is needed to explore the transitions which take place in the attacking zones. However, it is argued that the inside 50 is an important statistic in AFL football for the simple reason that if you aren't getting the ball into attack then you aren't going to score, and scoring is what wins games. To investigate this further, the next section will look at a transition probability profile for the competition and will consider the effect of profile adjustments on expected score and probability of victory.

12.3 The importance of inside 50 entries to a team's winning chances

To further investigate the article referred to above, another application of the zone model will be introduced combining simulation and adjustments to transition probabilities. The article stated that "inside 50s were grossly misleading" and this analysis will test the importance of inside 50s to a team's chances of winning and expected score. To do this a 'competition' transition probability matrix has been derived using the 185 games from the 2005 season. This uses the data from every Team A and Team B in these games to come up with a profile for the season and this matrix is presented in Table 12.5.

In order to be able to have something to use as a control for comparison purposes, this matrix has been used to simulate an 'average' match for 2005 with 223 transitions per quarter. This simulation produced a score line of 99-90 in Team A's favour. This translates to Team A having a 59.8% chance of victory, with Team B a 39.0% chance and the draw a 1.2% chance. These results were calibrated against actual 2005 season data, which showed that the average number of points scored in a match for the season was 189 points and this is exactly what the simulation has returned. Varying situations will be investigated by adjusting Team A's inside 50 percentages and simulating with the adjusted probability matrix. These situations will be described briefly and the results will be presented in Table 12.6.

- Scenario One: Removal of inside 50s from ball being forced in from a stoppage – from Table 12.5, 0.2% of CEBO2 1.9% of BUBO2 and 1.9% of THIN2 end up in zone 1. The Zone 1 states these percentages have finished in have been replaced by their counterpart from Zone 2. The reasoning behind this stems directly from the above article where Carey complained that 'teams were credited with as many as four successive entries into the 50 when the ball went back and forth across the arc during play'.

- Scenario Two: Removal of disputed ball inside 50s where the ball is in dispute and crosses the arc. It is assumed that the ball goes from dispute in Zone 2 into the corresponding state that it entered in Zone 1 but still in Zone 2. The logic behind this again relates to Carey's complaints from scenario one.

- Scenario Three: Scenario one and two combined to totally remove Carey's 'junk' inside 50s.

- Scenario Four: 5% point reduction in Team A's inside 50s resulting from Team A possession in Zone 2. Each of the transitions that result from Team A possession in the midfield (except CEBO2) will be reduced by 5% points, with that reduction being given to the corresponding state in Zone 2. This reduction is a way of quantifying the importance of getting the ball into your attacking zone.

- Scenario Five: 7.5% point reduction in Team A's inside 50s.

- Scenario Six: 10% point reduction in Team A's inside 50s.

- Scenario Seven: 15% point reduction in Team A's inside 50s.

- Scenario Eight: 25% point reduction in Team A's inside 50s.

Table 12.5: Transition probability profile for season 2005

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1
BUBO1	9.4%	2.8%	1.9%	36.2%	40.9%	2.3%
THIN1	7.6%	4.5%	2.2%	38.1%	42.6%	0.3%
DISP1	6.4%	14.8%	4.0%	40.4%	33.5%	0.0%
APOS1	0.2%	0.2%	13.6%	25.7%	8.5%	21.1%
BPOS1	0.1%	0.2%	4.8%	1.9%	43.9%	0.3%
ABEH1	0.5%	0.0%	2.5%	2.6%	84.2%	0.0%
CEBO2	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%
BUBO2	0.4%	0.0%	0.0%	0.8%	0.7%	0.0%
THIN2	0.1%	0.0%	0.1%	1.0%	0.6%	0.0%
DISP2	0.0%	0.0%	0.1%	0.2%	0.2%	0.0%
APOS2	0.0%	0.0%	10.2%	7.5%	3.6%	1.0%
BPOS2	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%
BUBO3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
THIN3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
DISP3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
APOS3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
BPOS3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
BBEH3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
State	**CEBO2**	**BUBO2**	**THIN2**	**DISP2**	**APOS2**	**BPOS2**
BUBO1	0.0%	0.7%	0.7%	0.0%	1.7%	3.5%
THIN1	0.0%	0.6%	0.8%	0.1%	1.7%	1.6%
DISP1	0.0%	0.0%	0.0%	0.3%	0.3%	0.3%
APOS1	29.4%	0.0%	0.0%	0.7%	0.1%	0.4%
BPOS1	0.0%	0.0%	0.5%	11.0%	5.0%	32.4%
ABEH1	0.0%	0.0%	0.0%	2.6%	5.1%	2.5%
CEBO2	0.0%	11.9%	0.0%	3.2%	43.2%	41.4%
BUBO2	0.0%	9.5%	2.8%	2.4%	40.5%	41.0%
THIN2	0.0%	9.3%	4.4%	2.7%	41.8%	37.9%
DISP2	0.0%	6.2%	17.7%	3.3%	36.5%	35.4%
APOS2	0.4%	0.1%	0.3%	16.5%	54.2%	5.9%
BPOS2	0.3%	0.1%	0.3%	17.1%	6.4%	53.9%
BUBO3	0.0%	0.2%	1.0%	0.0%	3.3%	1.4%
THIN3	0.0%	0.5%	0.6%	0.1%	1.9%	1.3%
DISP3	0.0%	0.0%	0.0%	0.2%	0.4%	0.3%
APOS3	0.0%	0.0%	0.4%	11.2%	33.3%	4.5%
BPOS3	28.0%	0.0%	0.0%	0.9%	0.2%	0.1%
BBEH3	0.0%	0.0%	0.0%	10.1%	14.9%	3.3%

Table 12.5: Transition probability profile for season 2005 (cont.)

State	BUBO3	THIN3	DISP3	APOS3	BPOS3	BBEH3
BUBO1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%
BUBO2	0.20%	0.00%	0.00%	0.80%	0.80%	0.00%
THIN2	0.30%	0.10%	0.10%	0.80%	0.80%	0.00%
DISP2	0.00%	0.00%	0.10%	0.20%	0.20%	0.00%
APOS2	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	0.00%	0.00%	10.10%	3.70%	7.00%	0.80%
BUBO3	8.40%	3.50%	3.50%	46.00%	30.30%	2.40%
THIN3	9.30%	4.70%	1.90%	40.80%	38.50%	0.40%
DISP3	5.80%	14.70%	4.60%	36.50%	37.60%	0.00%
APOS3	0.10%	0.20%	4.10%	43.80%	2.00%	0.30%
BPOS3	0.10%	0.20%	14.10%	8.60%	27.40%	20.30%
BBEH3	0.30%	0.10%	1.40%	68.50%	1.50%	0.00%

Table 12.6: Comparison of scenarios with adjusted inside 50s for Team A

Scenario	Description	A Score	B Score	A Win	B Win	Draw
	Original	99.0	90.0	59.8%	39.0%	1.2%
1	No Stoppage	97.8	90.5	58.8%	40.1%	1.1%
2	No Dispute	98.3	90.4	59.4%	39.4%	1.3%
3	No Stoppage or Dispute	97.4	90.5	58.2%	40.7%	1.2%
4	Team A reduced by 5%	95.4	90.6	55.1%	43.5%	1.3%
5	Team A reduced by 7.5%	93.8	90.7	52.9%	45.8%	1.2%
6	Team A reduced by 10%	92.2	90.9	50.5%	48.3%	1.3%
7	Team A reduced by 15%	89.0	91.3	46.2%	52.5%	1.3%
8	Team A reduced by 25%	81.8	92.1	36.1%	62.8%	1.2%

The adjustments to Team A's inside 50s have had little effect on Team B's score with only a marginal increase due to more ball in the midfield. This is the desired effect as it allows for a comparison across the scenarios of Team A's expected score and probability of victory. In each scenario, the removal of inside 50s for Team A has reduced their score and subsequently diminished their chances of victory. In scenarios 1, 2 and 3 these reductions are minimal, however they are still noticeable given the low numbers of Inside 50s that were removed. This gives the indication that inside 50s of the nature that Carey alluded to can still be crucial to a team's score with even the slightest edge an advantage in an even competition. The strength of the argument for inside 50s and their significance comes from scenarios 4 to 8 where Team A's forays into attack were reduced by varying amounts by the reduction of inside 50s. Even the smallest reduction of 5% dropped nearly four points from their expected score, and reduced their chances of victory by almost 5%. A drop in inside 50s of 25% saw Team A lose almost three goals from their expected score and drop nearly 25% in terms of their chances of victory. This shows clearly, how important inside 50s are to a team's chances of winning. Therefore inside 50s should not be referred to as the 'worst stat in footy' or 'grossly misleading'. This analysis has shown that by limiting your entries to attack will reduce your chances of victory.

This section has presented an application for investigating the 'science' of football using the opinion of a commentator as the basis for the analysis. It is unfortunate that in the modern era of football commentators and media, comments made by so called 'experts' have no evidence to back them up. Wayne Carey's opinion as espoused by Mike Sheahan appears to be one such comment. Although there was some merit in his argument about the way the ball is distributed into the forward 50, it has been shown in section 12.2 that this is not always the case. Furthermore, the analysis presented in this section shows how important having the ball in your attacking zone is. By reducing the number of entries into the attacking zone, the chances of victory are reduced. Perhaps the last word should be reserved for the football manager who said about inside 50s in the article, "what they do tell you is if you don't get the ball in there enough, you've got no hope".

12.4 Investigating styles of play in the AFL competition

The previous section examined the effect of changes in play by reducing the inside 50 forays for Team A. This kind of application could also be useful for further investigating different strategies and styles within the game. Ultimately, a team wants to maximise its own chances of victory and any edge it can gain in a certain area could be invaluable.

Discussion has been rife in the last few seasons comparing a ball retention style, known as ‘uncontested football’ to the more traditional long kicking style, which is known as ‘contested football’. Before 2004, many pundits believed Port Adelaide’s uncontested style was the reason behind its lack of success in finals football (Ker, 2004). Even though they won the flag in 2004, the numbers will show that they played a different style of football in the 2004 finals series (Champion, 2004). Previous analysis has also shown that every extra long kick a team has over their opponents contributes 1.4 points to the margin. This is by far the most significant and important statistic within the match for explaining margin of victory (Champion, 2005). With the advent of the zone model, analysis along the same lines as above should be able to be used to quantify particular styles of play and to determine where on the ground these styles are at their most effective. To investigate differing strategies, the competition matrix from Table 12.5 will be used with adjustments made to Team A probabilities and simulation used to quantify the effect of these adjustments.

12.4.1 Kicking long out of defence

The first play strategy investigated will be kicking long out of defence compared to retaining the ball via short kicks or handballs to players in the defensive zone. The competition matrix shows that Team A distributes the ball as displayed in Table 12.7 when coming out of defence.

Table 12.7: Team A’s distribution of ball out of defence

State	THIN2	DISP2	APOS2	BPOS2	BUBO3	THIN3	DISP3	APOS3	BPOS3	BBEH3
APOS3	0.4%	11.2%	33.3%	4.5%	0.1%	0.2%	4.1%	43.8%	2.0%	0.3%

It is evident from this profile, that teams are reluctant to kick the ball long out of defence and put the ball into dispute (11.2%). In contrast, the preferred method is to retain possession via short kick or handball either to players in the midfield (33.3%) or defensive zone (43.8%). The issue here is whether teams could be better served by getting the ball out of their defensive zone and putting it into the midfield via long kicks.

The propensity of teams to fiddle around with the ball in defence is as prevalent in the modern game as ever before, so much so that CD introduced a new match statistic for backwards kicks. The assumption has been made that one-quarter of the long kicks put into the midfield would be ones to advantage, guaranteeing that Team A retains possession. The scenarios and their effect are presented in Table 12.8.

Table 12.8: Effect of kicking long out of defence

Reduction in Defensive Possession	A Score	B Score	A Win	B Win	Draw
5%	98.9	90.3	59.5%	39.2%	1.3%
10%	98.9	90.4	59.7%	39.0%	1.2%
15%	99.0	90.4	59.8%	39.1%	1.1%
20%	99.1	90.4	59.6%	39.1%	1.3%
25%	99.0	90.5	59.7%	39.0%	1.3%

The simulated mean scores from the competition matrix were 98.7-90.3 and therefore, it is interesting to note that the expected return for Team A can be increased, albeit slightly, by adopting a long kicking approach out of defence over the more preferred kicking backwards across goal to retain possession. It must be remembered that for this analysis only one in four long kicks were presumed to go to advantage and therefore retain possession. Coming out of defence, the space in front of the kicker is more pronounced and it is believed that it would be possible for a defender to hit a loose target with better than a one in four chance. Any increase in the percentage of long kicks to advantage out of defence would further increase Team A’s expected score. Although the increase in expected scores and winning percentage do not shift greatly with a different style out of defence, this analysis shows how the zone model can be used in this instance and the ability of the user to make their own adjustments and assumptions.

12.4.2 Kicking long in to attack

The second scenario that has been investigated is the usage of the ball in the midfield by Team A and whether getting the ball into attack at any cost is any more advantageous than what takes place at the moment. Team A's usage of the ball in the midfield from Table 12.5 is contained in Table 12.9, with the transitions that amounted to 0.1% or less omitted due to space constraints.

Table 12.9: Team A's distribution of ball in the midfield

State	DISP1	APOS1	BPOS1	ABEH1	CEBO2	THIN2	DISP2	APOS2	BPOS2
APOS2	10.2%	7.5%	3.6%	1.0%	0.4%	0.3%	16.5%	54.2%	5.9%

It is surprising that only 22.3% of ball that Team A has in the midfield makes its way into the attacking zone. Perhaps the size of the midfield when compared to the extremities of the ground is the reason behind this as well as the increased propensity of teams to chip the ball around and increase the likelihood of a turnover. The ability of Team A to increase the entries of the ball into the forward 50 will no doubt increase their chances of victory, but is it worth just blazing away and putting the ball into dispute in the forward zone? To investigate this, reductions have been made to the amount of ball recycled by Team A in the midfield and this extra ball has been attributed under two further scenarios to disputed ball and Team A possession. The results of this analysis are presented in the following three tables.

Table 12.10: Increased forward entries with 100% of the ball into dispute

Increase in Entries	A Score	B Score	A Win	B Win	Draw
5%	103	90	64.4%	34.5%	1.2%
10%	106	91	68.6%	30.2%	1.2%
15%	109	91	71.3%	27.5%	1.2%

Table 12.11: Increased forward entries with 50% to dispute and 50% to possession

Increase in Entries	A Score	B Score	A Win	B Win	Draw
5%	107	90	68.9%	29.9%	1.2%
10%	114	91	76.2%	23.0%	0.8%
15%	120	91	81.2%	18.0%	0.8%

Table 12.12: Increased forward entries with 100% of the ball to possession

Increase in Entries	A Score	B Score	A Win	B Win	Draw
5%	111	91	72.7%	26.3%	1.0%
10%	122	91	82.3%	16.9%	0.8%
15%	131	91	87.9%	11.4%	0.7%

This analysis reinforces that done in section 12.3 regarding inside 50s. It is clear from the three tables above that getting the ball in as often as you can significantly improves the expected score a team will kick and therefore their chances of winning. It is little surprise that the ability to directly find a target inside the attacking zone is more beneficial than blazing the ball in, with the scores in Table 12.12 much higher than their counterparts in Table 12.10. Even so, the amount of fiddling about that is done in the midfield in the modern game is restrictive on teams and their ability to score. It is obvious from this analysis that a team is better served by making small reductions to this retention in the midfield and pumping the ball long into the forward 50. Even a 5% increase with the ball going only into dispute has been shown to benefit a team on the scoreboard by over half a goal.

12.5 Summary

The added advantage of having the location on ground in the zone model has allowed for detailed analysis that would be impossible with the global model. In particular, strategic questions are able to be answered and this was seen in a number of examples. Whilst the assumptions made in this chapter were quite limited, there is nothing stopping more elaborate assumptions and investigations from taking place. It was shown that despite criticism of the inside 50 as a match statistic, their importance to a team's quest for victory can not be denied. In some cases the method of distribution into the attacking zone illustrates why a team, which had similar entries to their opponents, has dominated on the scoreboard. In other instances, deeper investigation is needed due to the similarity in delivery. It was also shown how playing styles can be analysed and adjusted to ensure that they are the optimal strategy for winning. The improvements shown by coming out of defence via a more direct route were minimal but with more realistic assumptions and probability adjustments it is argued that these effects would have a greater impact on improving a team's chances of winning. Finally, the ability to create more scoring options by getting the ball into the forward 50 quicker had pronounced effects on the scoreboard, and although it is more advantageous to directly find a team mate, the impact of getting the ball in at all costs, even if this means blazing away and putting the ball into dispute, cannot be questioned. Having given an overview of the advantages of the zone model, the penultimate chapter of this thesis will investigate the characteristics of teams within the competition with a view to gaining an understanding of the style that team's adopt.

Chapter 13: Team characteristics derived from the Markov process models

13.1 Introduction

Whilst the applications of the Markov models have been well documented in this thesis, little reference has been made to the styles of play that teams display on the field. In the previous chapter, styles of play were investigated but these styles were on a global competition level using a combined matrix from the 185 games in season 2005. This chapter will be devoted to the probability matrices for individual teams and the differences they show from each other and from the competition average. Furthermore, the performance of teams will be compared for home state and interstate matches. There will also be venue analyses to determine whether different probability matrices result for different venues as well as an overview of AFL venues and how play differs amongst them. It is hoped that this analysis will provide an insight into the differing styles of teams and will explain why the stronger teams enjoy more success than the weaker teams. Hirotsu investigated the offensive and defensive capabilities of English Premier League teams at home and away, with reference to the Markov model he developed (Hirotsu and Wright, 2003). He was then able to use this information to suggest how the style of play should change for particular opponents. He believed that his approach "may be useful for extracting the characteristics of teams from a large amount of soccer data" (Hirotsu, 2002) and it is hoped that the approach presented here will be of similar use in AFL football.

13.2 Characteristics of AFL venues

From 2006 the AFL competition used 11 different venues Australia-wide with every state and territory represented. A unique feature of these venues is that the dimensions are all different. Table 13.1 contains the venues used for matches in the 2004 and 2005 seasons and the number of games they hosted.

Table 13.1: AFL venues and number of games hosted, 2004/2005

Venue	State/Territory	Matches
M.C.G.	VIC	90
Docklands	VIC	89
Football Park	SA	49
Subiaco Oval	WA	45
Gabba	QLD	26
S.C.G.	NSW	18
Kardinia Park	VIC	16
Optus Oval	VIC	15
York Park	TAS	8
Olympic Stadium	NSW	8
Manuka Oval	ACT	5
Marrara Oval	NT	1

Using the transition matrices from each game, a log-linear analysis using the CATMOD procedure in SAS was conducted as all the variables are categorical, to test whether the transition probabilities are significantly affected by venue. This analysis indicated that match transition probabilities are significantly affected by venue ($\chi^2 = 706.66$, d.f. = 307, p-val <0.0001). This is not a surprising result given the different sizes of grounds and the different playing styles of clubs that will be addressed in the next section. To add to the overall venues analysis, comparisons have been carried out between venues to discover whether any venues generate a similar style of match. These comparisons are presented in Table 13.2 with the non-significant comparisons at the 5% level of significance in bold and the test statistics will be presented in Appendix 2. A Bonferroni correction has been applied in order to ensure a realistic overall probability of a Type I error (Yes: p-value < 0.001; No: p-value ≥ 0.001). The venues are coded as follows: Football Park (FP), Gabba (GA), Optus Oval (OO), Subiaco (SU), Manuka Oval (MAN), Marrara Oval (MAR), Docklands (DO), York Park (YP) and Olympic Stadium (OLY).

Table 13.2: Comparison of transition probabilities between AFL venues

Venue	**GA**	**KP**	**MCG**	**OO**	**SCG**	**SU**	**MAN**	**MAR**	**DO**	**YP**	**OLY**
FP	**No**	Yes	Yes	Yes	Yes	Yes	**No**	**No**	Yes	**No**	Yes
GA		Yes	Yes	Yes	**No**	Yes	**No**	**No**	Yes	**No**	Yes
KP			**No**	Yes	**No**	**No**	**No**	**No**	Yes	**No**	**No**
MCG				Yes	**No**	Yes	**No**	**No**	Yes	**No**	**No**
OO					Yes	Yes	**No**	**No**	Yes	Yes	Yes
SCG						**No**	**No**	**No**	**No**	**No**	**No**
SU							**No**	**No**	Yes	**No**	**No**
MAN								**No**	**No**	**No**	Yes
MAR									**No**	**No**	**No**
DO										Yes	Yes
YP											**No**

From Table 13.2, there are seven relevant comparisons where venues do not display a dissimilar style of play. The other comparisons that were not significant came from venues (Manuka, Marrara, York and Olympic) who have hosted less than ten games for the period of analysis and therefore the results should be treated with some caution. Of the venues who have hosted a reasonable number of games, the M.C.G. produces a not dissimilar style to both Kardinia Park and the S.C.G. Given the very small dimensions of the S.C.G. (149m x 136m) when compared to the M.C.G. (160m x 141m) this result is surprising. Similarly, Kardinia Park (170m x 115m) has very different dimensions to the M.C.G.; however the style of play at these venues is not significantly different. Kardinia Park is not significantly different to either the S.C.G. or Subiaco (177m x 122m) with its similarity to Subiaco's dimensions the likely reasons for this comparison's result. Subiaco and the S.C.G. do not produce a significantly dissimilar style of play. The Gabba's (156m x 138m) style of play is not dissimilar to both Football Park (165m x 133m) and the S.C.G., and these grounds have similar width dimensions although they differ in length.

Even though AFL grounds have very different dimensions, there is some evidence to suggest that certain grounds are more likely to produce a certain style of play than others. It is interesting to note that the smallest ground, the S.C.G., does not produce a dissimilar style to a number of other grounds, particularly the bigger grounds such as Kardinia Park and Subiaco. The focus of this chapter will now shift to the clubs and perhaps explain why some of these venues produce a similar style of play.

13.3 Characteristics of AFL teams compared to the competition average

The initial analysis of this section will concentrate on individual teams and will determine whether something meaningful can be gleaned from their match statistics. To do this, each team has had their matches from the 2004 and 2005 season combined to form a 'club' matrix, similar to what was used for the competition in Chapter 12. This club matrix will then be used to compare clubs, both to each other, and to the competition as a whole. Initially though, some reference has to be made to the number of transitions that each team is likely to have in a game. It was seen in Chapter 11 that the average number of transitions for a match under the zone model was 880 or around 220 per quarter. Table 13.3 displays the average number of transitions for each club together with 95% confidence intervals for the mean. Also included is the average number of general play stoppages per team per match.

Table 13.3: Average match transitions per club 2004/2005

Team	Matches	Mean Transitions	Std Dev	Lower Limit	Upper Limit	Mean Stoppages
Adelaide	47	874.6	55.4	858.8	890.5	56.7
Brisbane	47	889.9	44.4	877.2	902.5	54.0
Carlton	44	875.3	52.7	859.7	890.9	53.9
Collingwood	44	889.0	50.9	874.0	904.1	50.5
Essendon	46	895.7	53.7	880.2	911.2	51.7
Fremantle	44	890.9	43.1	878.1	903.6	51.0
Geelong	49	890.1	59.5	873.4	906.7	56.7
Hawthorn	44	882.5	49.2	868.0	897.1	54.3
Melbourne	46	884.4	46.3	871.0	897.8	50.4
Kangaroos	45	863.2	47.2	849.4	877.0	51.2
Port Adelaide	49	870.4	45.3	857.7	883.1	49.7
Richmond	44	883.2	53.0	867.6	898.9	51.9
St. Kilda	49	866.7	49.5	852.8	880.6	54.5
Western Bulldogs	44	904.6	47.5	890.5	918.6	46.8
West Coast	48	892.5	48.3	878.9	906.2	53.6
Sydney	50	833.4	49.2	819.8	847.1	67.8
Competition	740	880.0	52.1	876.3	883.8	53.5

Only one club, Western Bulldogs, averages over 900 transitions per game, whilst only one club, Sydney, averages less than 850 transitions per game. Sydney are renowned for playing a negating style of football, whilst the Bulldogs are know for their free flowing style of play that involves a lot of running. The table shows that Sydney, on average, has the most stoppages per games, whilst the Bulldogs, on average, have the fewest stoppages per game. This result indicates that the more transitions there are in a match, the more free-flowing the style of the game and the fewer stoppages in the game.

The correlation between general play stoppages and transitions in the match equals a weak -0.20. The sign of this correlation indicates that less general play stoppages lead to more transitions and vice-versa. This reinforces the result from above relating to Sydney and the Western Bulldogs and suggests that something can be learnt about the style of play that took place in a game by looking only at the number of transitions. However, there is much more that can be learnt from the transition probabilities of the zone model. These relationships will be examined at an individual club level in the following sections. Due to the amount of space they require the transition matrices for each team are contained in Appendix 3.

The individual club matrices were compared to the competition matrix and z-tests were carried out to ascertain whether there is any significant difference between the proportions. The level of significance is revised using a Bonferroni correction due to the large number of tests conducted. As the Bonferroni correction is large and makes the significance level very conservative, the overall level of significance has been raised to 0.15. The number of transitions that are possible in the 18x18 zone model is 213 therefore the revised level of significance is 0.0007 (i.e. 0.15/213). 94 transition probabilities showed a significant difference between the club probability and the competition average. To gain a better understanding of the relative strength of teams over the past two seasons, an amalgamated table ranked by winning percentage is contained in Table 13.4 along with the number of significantly different transitions for each club compared to the competition.

Various match statistics were investigated in Chapter 6 to ascertain whether home advantage had an affect on team's hardness at the ball via contested possession or the ability of a team to create extra space at home to gain uncontested possessions. The analysis showed that home advantage only had minimal effects on types of possession, however, with the development of a model that uses transitions from one state to another to approximate an AFL match, this analysis has been revisited using the Markov model. Similar to above, the 2004 and 2005 seasons have been used for the analysis. The matches have been coded into interstate or local depending on whether the team is playing in their home state or interstate. A matrix has been derived for each club for their local and interstate matches and each transition probability proportion from the two matrices have been compared to test whether there is a significant difference between a club's performance in their home state and interstate. Again the level of significance is adjusted for the number of z-tests to 0.0007. The number of significant differences between home state and interstate games for each club is contained in Table 13.4.

Table 13.4: AFL table and significant comparisons of transition probabilities ranked by winning percentage

Team	Played	Win	Loss	Draw	Win %	Significant Comparison to Average	Significant Interstate Comparison
West Coast	48	32	16	0	66.7	4	2
Port Adelaide	49	32	16	1	65.3	5	0
St Kilda	49	32	17	0	65.3	3	4
Sydney	50	32	18	0	64.0	15	0
Brisbane Lions	47	28	19	0	59.6	2	3
Geelong	49	29	20	0	59.2	6	0
Melbourne	46	26	20	0	56.5	4	2
Adelaide	47	26	21	0	55.3	5	2
Kangaroos	45	23	22	0	51.1	3	0
Fremantle	44	22	22	0	50.0	6	3
Essendon	46	21	25	0	45.7	6	2
Bulldogs	44	16	28	0	36.4	8	3
Carlton	44	14	29	1	31.8	10	1
Richmond	44	14	30	0	31.8	7	0
Collingwood	44	13	31	0	29.5	3	0
Hawthorn	44	9	35	0	20.5	7	2

The relevant transitions from Table 13.4 that differ significantly, both to the competition average and for interstate travel, will be displayed graphically on a football field to try and better highlight the style of play that the team in question has adopted. The arrows on each field indicate the significantly different transitions for the club with the red broken arrows representing rates that are significantly below the average and the blue solid arrows representing rates significantly above the average.

13.4 Adelaide Football Club

Figure 13.1: Adelaide: significantly different transitions to competition

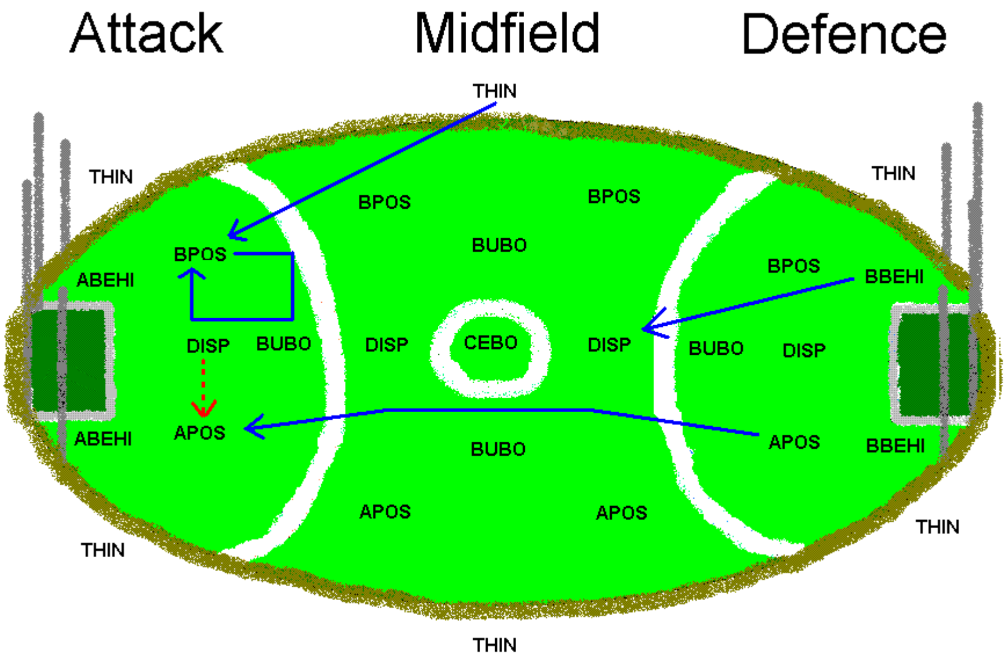
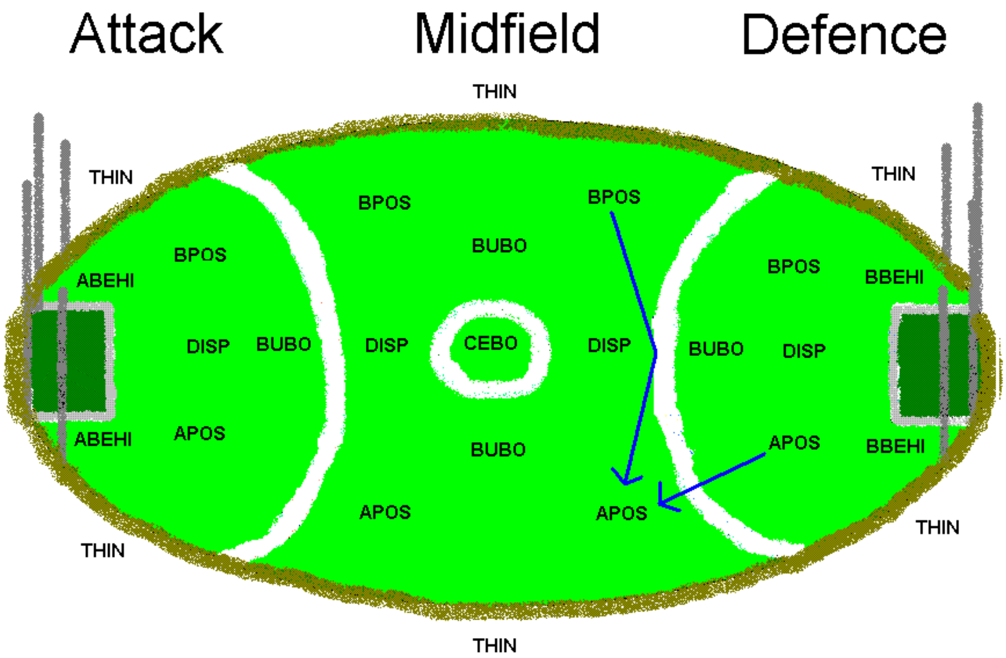


Figure 13.2: Adelaide: significantly different transitions for interstate travel



When compared to the competition, Adelaide struggle at winning the disputed ball in their attacking zone (-4.8%). Their opponent's are able to gain above average possessions in Adelaide's attacking zone from midfield throw-ins (+0.9%). When defending, the opposition is able to find free targets with ease (+3.1%) and Adelaide could look to tighten up on their opposition in their attacking zone. When bringing the ball back into play after an opposition behind, Adelaide kicks long to contests in the midfield at a significantly higher rate than the competition average (+5.9%). Adelaide benefits from opposition turnovers in the midfield at Football Park at a significantly higher rate than when they travel interstate (1.7%). At home they are also able to clear their defensive zone to team mates in the midfield at a much higher rate than when they play matches interstate (6.7%). Adelaide needs to adopt similar midfield pressure when playing interstate in an attempt to force their opponents into error. They could also look to improve their run out of defence when they travel interstate and hope to clear the defensive area as well as they do at home.

13.5 Brisbane Football Club

Figure 13.3: Brisbane: significantly different transitions to competition

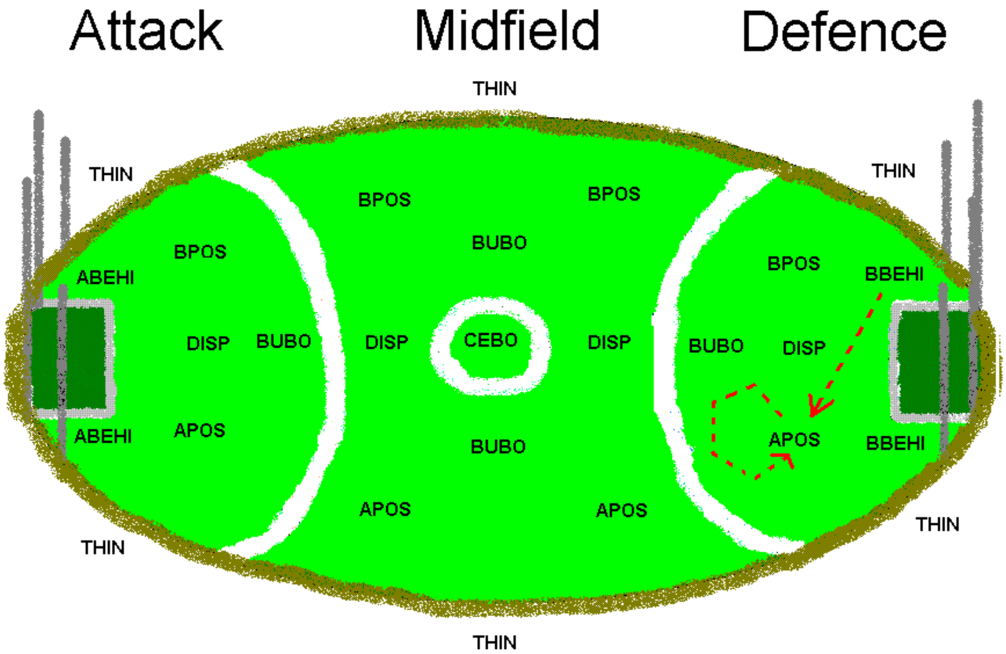
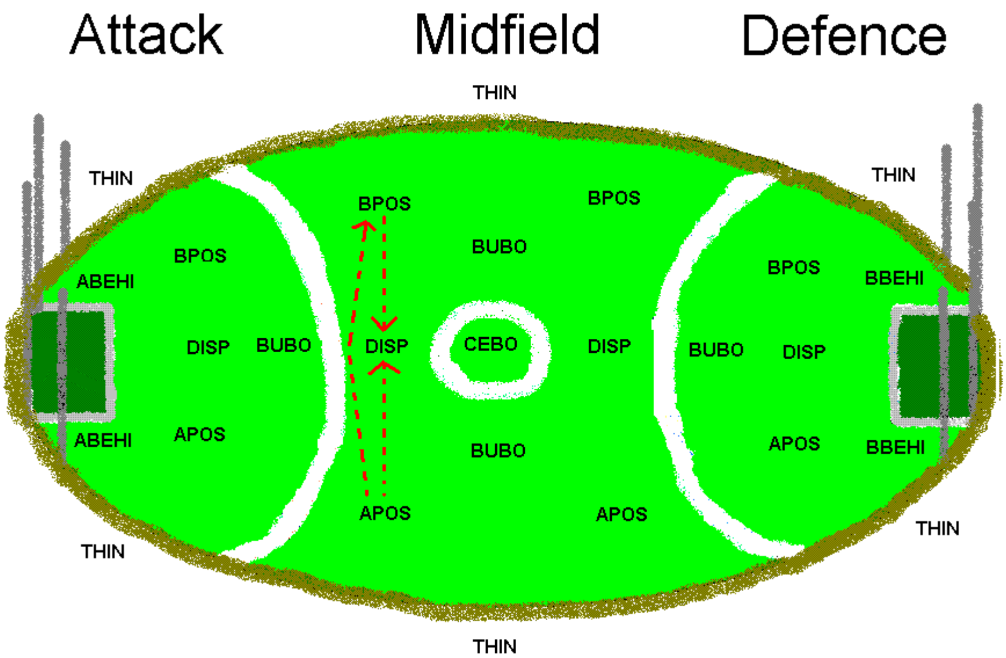


Figure 13.4: Brisbane: significantly different transitions for interstate travel



Brisbane don't do a lot different to the competition average, however, their returning of the ball into play after an opposition behind sees them significantly lower in finding open team mates in zone 3 (-9.8%). Presumably they prefer to kick long into the midfield. They are also well below the competition average for being able to find open team mates when they have the ball in the defensive zone (-4.9%). When Brisbane plays outside of Queensland they have a significantly higher rate of turning the ball over to their opposition in the midfield (1.7%). They also change their style in the midfield by putting the ball into dispute at a higher rate (3.1%), whilst their opponent's interstate put the ball into dispute at a lower rate than games played at the Gabba (2.5%).

13.6 Carlton Football Club

Figure 13.5: Carlton: significantly different transitions to competition

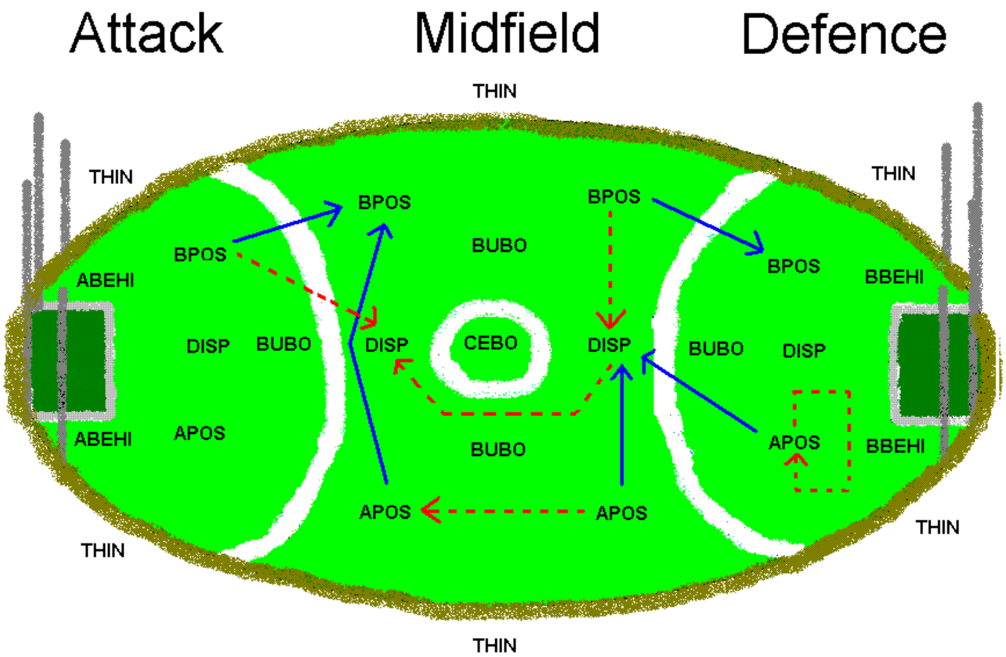
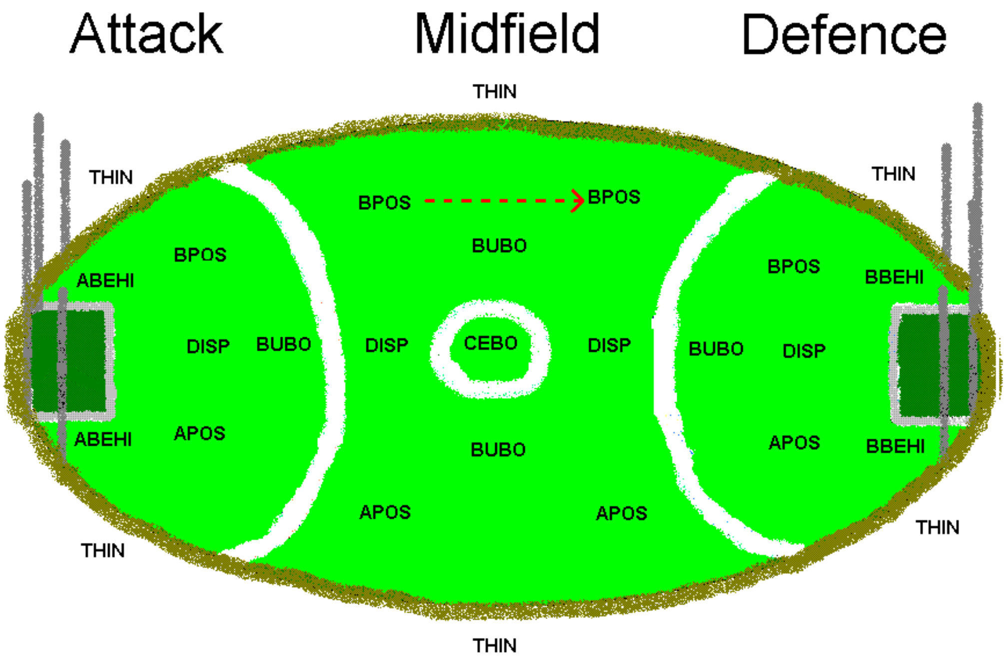


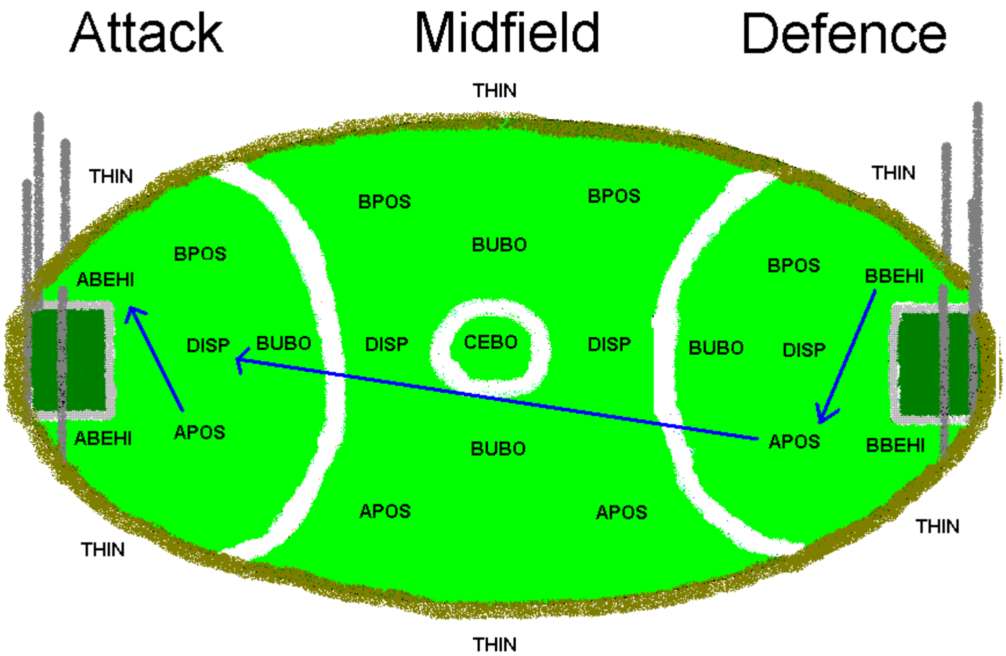
Figure 13.6: Carlton: significantly different transitions for interstate travel



Carlton does a number of things significantly different to the competition average. They allow their opponents too much latitude coming out of defence by letting them find team mates in the midfield easily (+3.4%) instead of having to kick to contests (-2.2%). Once the ball goes into the midfield, Carlton's use of the ball is poor with an increased rate of turnovers (+1.4%) and disputed situations (+2.4%). They have a significantly lower rate of retaining the ball in the midfield (-5.6%). On the other hand, their opponent's are not forced into kicking to contests in the midfield (-1.6%) but are able to find team mates on attack at an increased rate (+1.0%). When in possession of the ball in defence, Carlton have to kick more often to midfield contests (+3.1%) and struggle to find open team mates within zone 3 (-4.3%). These significant differences indicate that Carlton needs to tighten up on their opponent's all over the ground. At the same time their use of the ball needs to improve and they may benefit from reconsidering their game plan of kicking the ball long as the first option. At least until they have the squad whose skills are up to playing this way. The only difference between Victorian and interstate games for Carlton is the increased ability of the home side to find open team mates in the midfield in interstate games (4.4%). Again this seems to point to a lack of close checking by Carlton players when they play games outside of Victoria.

13.7 Collingwood Football Club

Figure 13.7: Collingwood: significantly different transitions to competition



Collingwood would best be described as an average club; although their winning percentage does not reflect this. They have no significant differences between what they do in Victoria compared to interstate. When compared to the average, they have an increased rate of kicking behinds when attacking (+3.6%). When returning the ball from an opposition behind they prefer to take the cheap possession in zone 3 (+7.9%). Perhaps Collingwood need to look at becoming a little more unpredictable in their play with a view to increasing their winning percentage.

13.8 Essendon Football Club

Figure 13.8: Essendon: significantly different transitions to competition

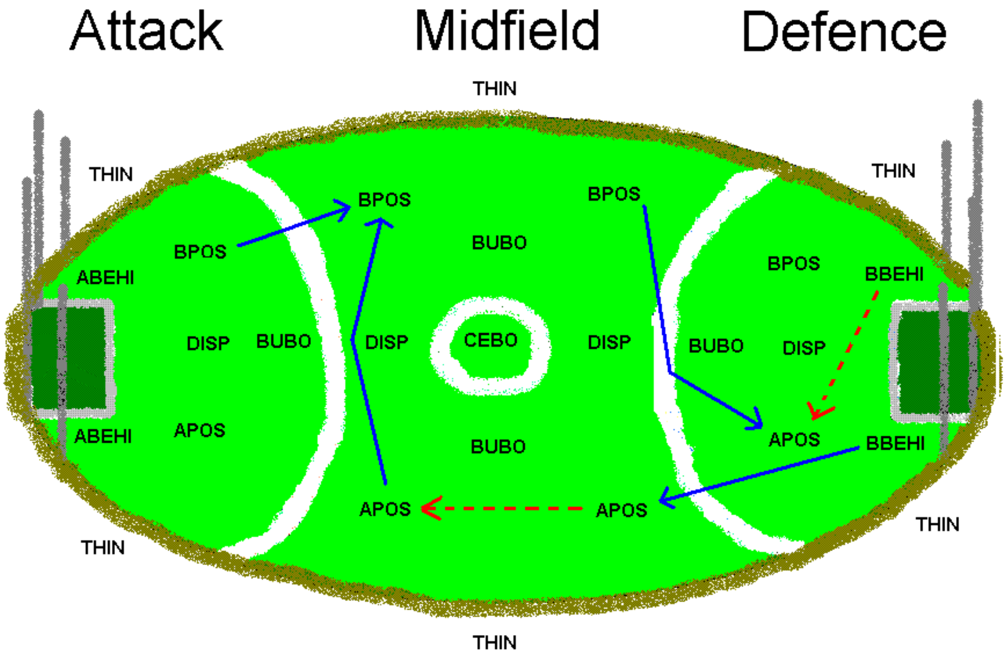
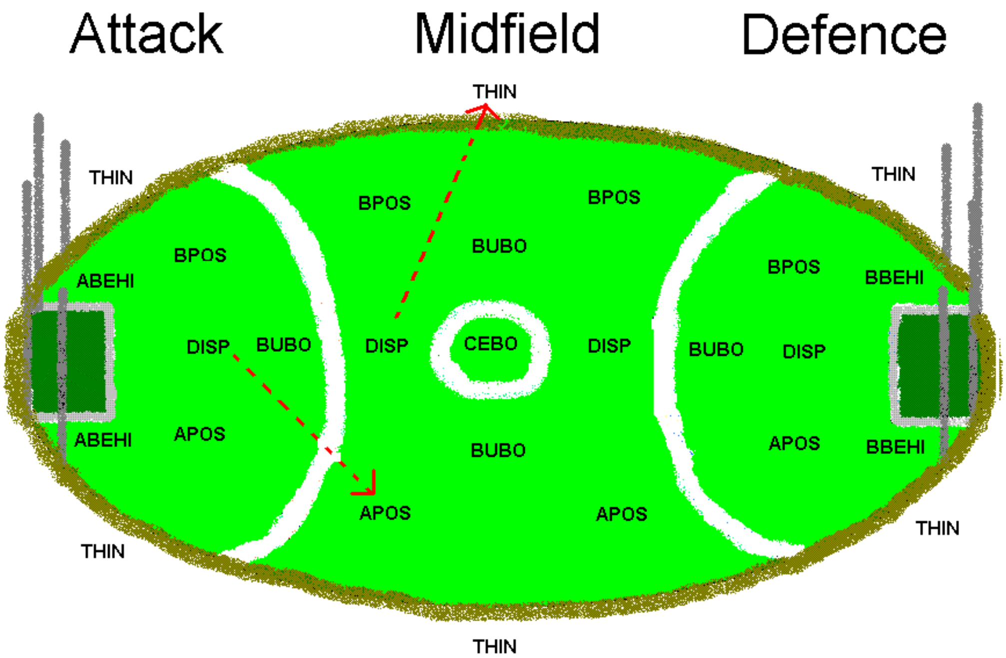


Figure 13.9: Essendon: significantly different transitions for interstate travel



Essendon's opponents clear the ball easily out of defence at an above average rate (+3.3%). In the midfield Essendon's skill lets them down with a below average rate of hitting a team mate (-2.2%) and an above average error rate (+0.9%). The lack of skill is lessened by the opposition's above average rate for turning the ball over when they enter their forward 50 (+0.7%). When Essendon is kicking-in, their long kicking allows them to find a team-mate in the midfield at an above average rate (+5.5%). Long kick-ins means Essendon's rate of finding a team-mate in the defensive zone is much lower than the average (-8.0%). When traveling interstate, Essendon is significantly less likely to win a disputed ball around their forward 50 (1.0%). Interstate, their rate of forcing throw-ins in the midfield is much lower than when they don't have to travel (5.5%).

13.9 Fremantle Football Club

Figure 13.10: Fremantle: significantly different transitions to competition

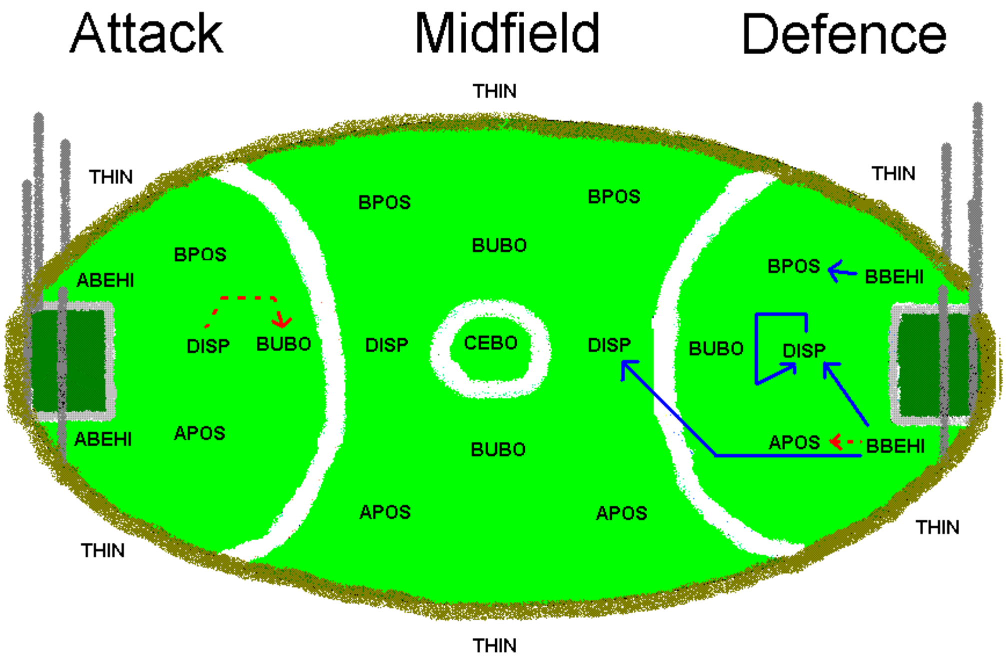
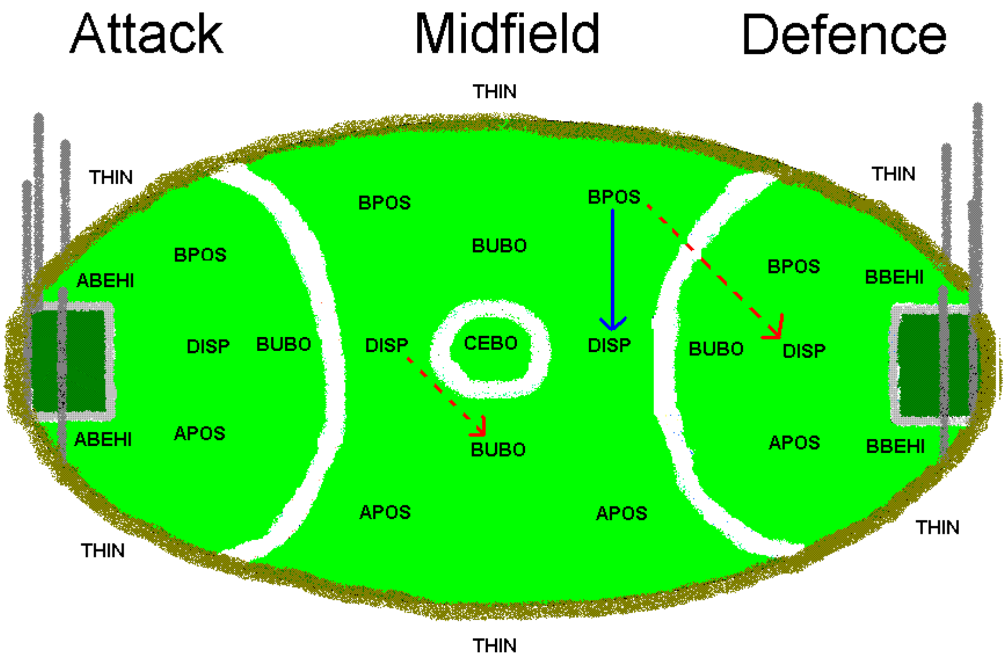


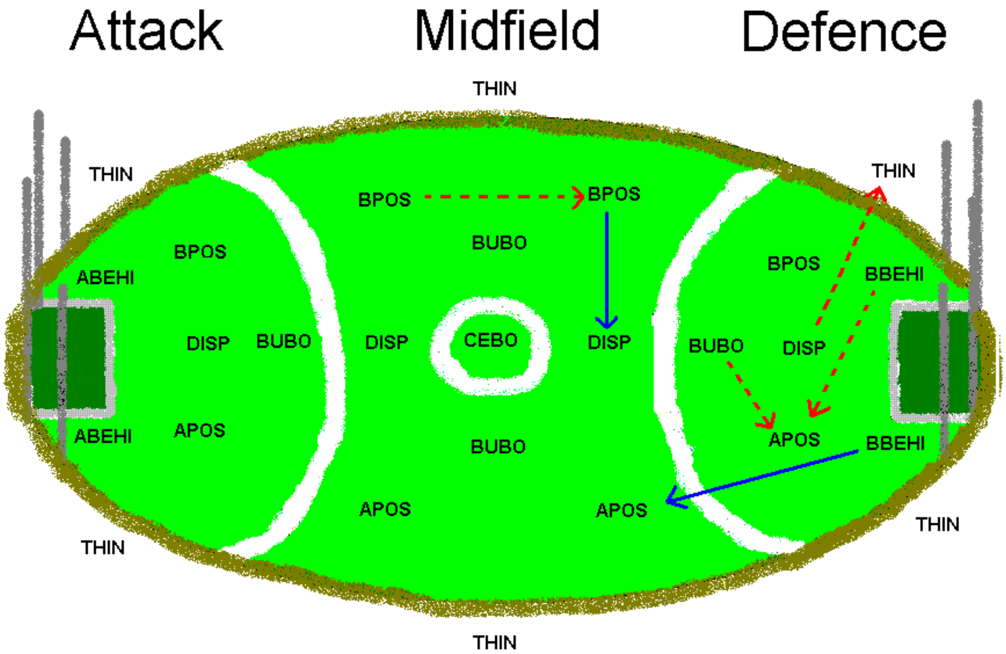
Figure 13.11: Fremantle: significantly different transitions for interstate travel



A key area for Fremantle to address is their kick-ins. When returning the ball into play after an opposition behind, Fremantle return it directly to an opponent on attack at a higher rate (+2.4%). They put the ball into dispute short (+2.3%) and long (+5.8%) at above average rates. As a result of this poor delivery, their rate of finding a team mate in an uncontested area of the defensive 50 is significantly lower than average (-8.5%). Clearly, Fremantle need to address their kicking in set up including the kicker, in order to improve their winning percentage and become a more consistent side. Locally, Fremantle force less ball-ups in the midfield (3.0%) even though their opponents, when in possession in zone 2, put the ball into dispute in zone 2 (2.7%) more than zone 3 (-2.7%).

13.10 Geelong Football Club

Figure 13.12: Geelong: significantly different transitions to competition



Like Essendon, Geelong prefers a long kick in to a team mate in the midfield (+6.2%) over a short kick to a team mate in defence (-8.6%). They really struggle in defence at winning first possession at ball-ups (-12.7%) and have a below average rate for zone 3 throw-ins (-3.0%). In the midfield, Geelong's opponents are forced to kick to contests more often (+2.1%) at the expense of finding open team-mates (-2.0%). Geelong could benefit greatly by improving their work around zone 3 ball-ups and improve their winning percentage as a result. There are no significant differences between Geelong's Victorian style of play and what they do when they travel interstate.

13.11 Hawthorn Football Club

Figure 13.13: Hawthorn: significantly different transitions to competition

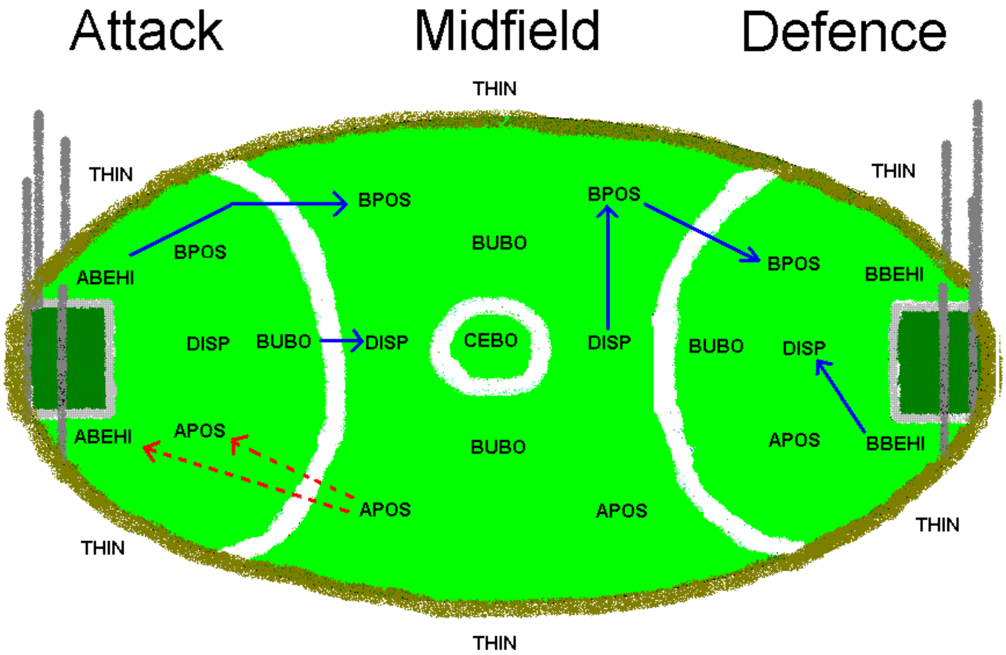
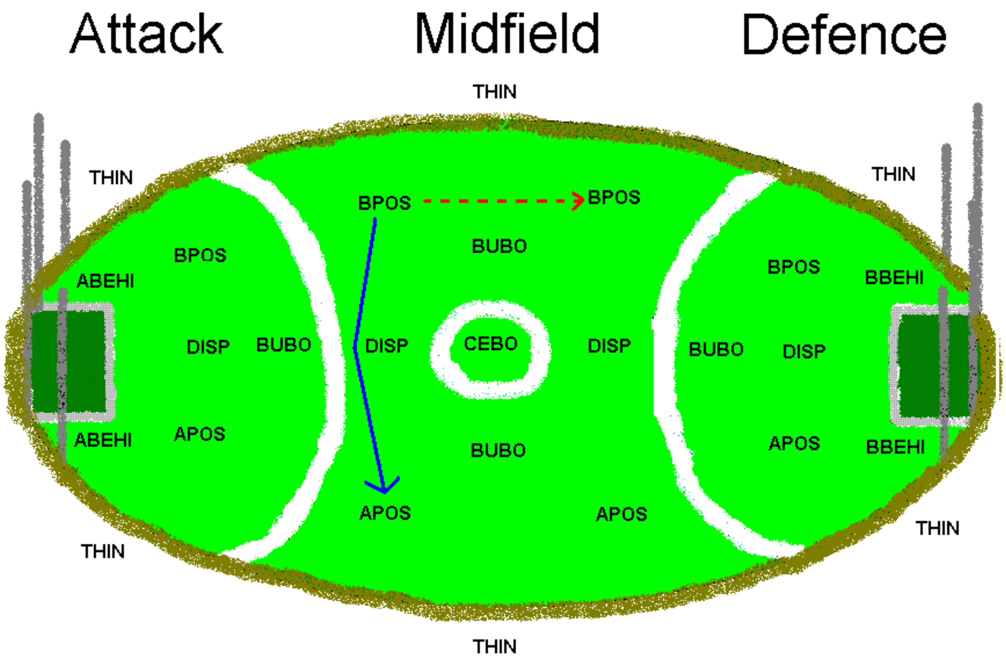


Figure 13.14: Hawthorn: significantly different transitions for interstate travel



Hawthorn's struggle to win games is evidenced by their below average rate for finding a team mate open in attack when kicking from the midfield (-1.1%) and their below average rate for kicking the ball long from the midfield for a score (-0.4%). Their opponent's enjoy an above average ability to find team mates when returning the ball after a Hawthorn behind (+8.1%). Once in possession in the midfield, the opposition enters their attacking zone easily at an above average rate (+1.3%), which comes on the back of a superior advantage at extracting the ball in midfield, disputed situations (+3.0%). Hawthorn struggle on kick-ins with an above average rate for kicking to contests in zone 3 (+2.3%). When playing in Victoria, Hawthorn put more pressure on their opponent's in the midfield and force them into more turnovers (2.0%). When they travel interstate, they allow the opposition to hit open targets at a much increased rate (5.1%). More pressure in the midfield when playing on the road is a definite must for Hawthorn. This increase in possession needs to be better used when being transferred into attack.

13.12 Melbourne Football Club

Figure 13.15: Melbourne: significantly different transitions to competition

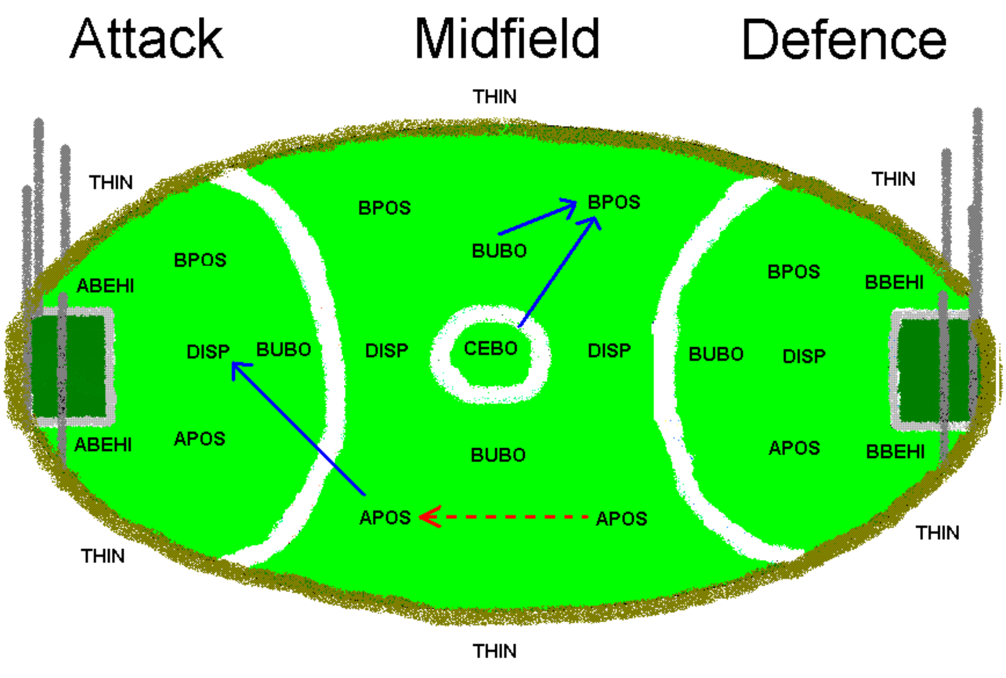
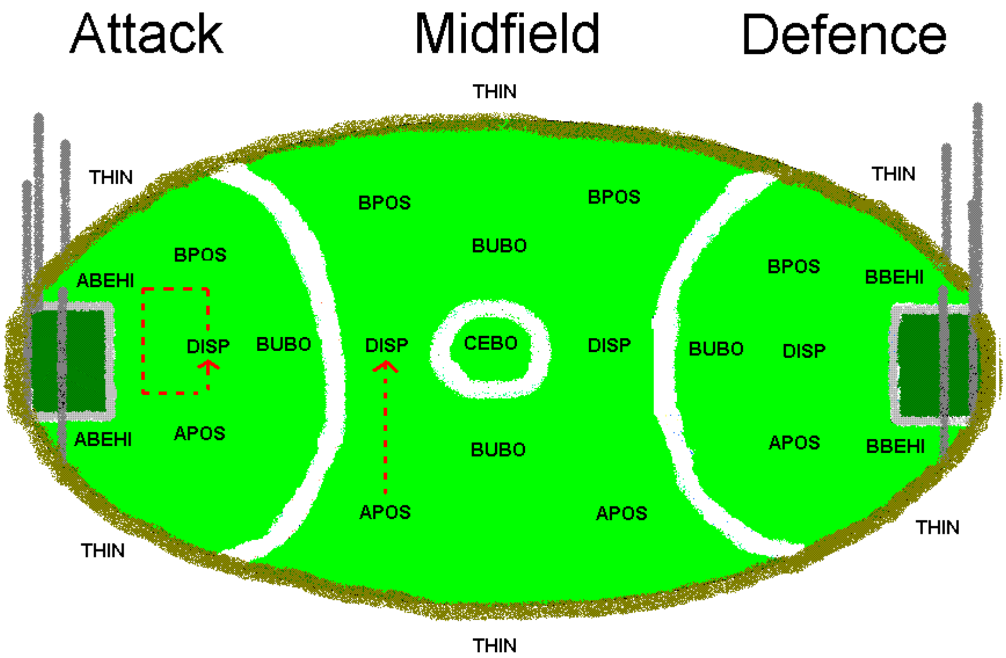


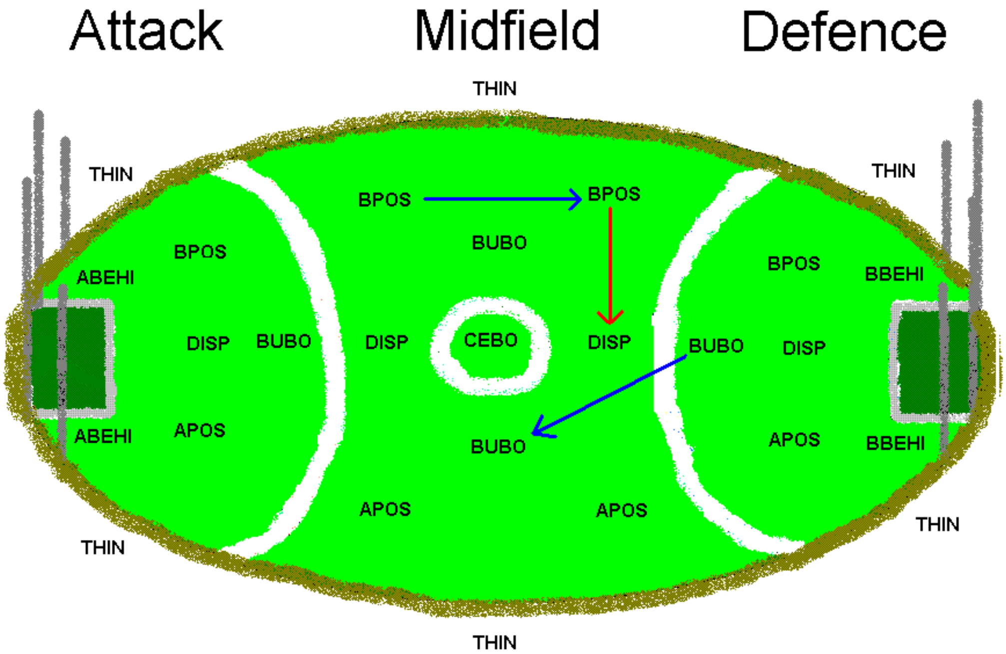
Figure 13.16: Melbourne: significantly different transitions for interstate travel



Melbourne's differences all stem from the midfield with a reduced rate of finding open team mates when in possession (-3.1%) along with the opposition excelling at centre bounces (+5.3%) and ball ups (+6.6%). When Melbourne has the ball in the midfield they are too reliant on kicking the ball into attack via a kick to a contest (+1.5%). On the road, Melbourne increase their rate of putting the ball into dispute in the midfield (3.3%) as well as the level of ball staying in dispute in their attacking zone (4.2%). Melbourne can improve their effectiveness with a stronger attack on the ball leading to an improved rate of extraction from dispute and better disposal in the midfield. Their away form could be aided by cleaner ball in attack, creating more scoring opportunities.

13.13 North Melbourne Football Club

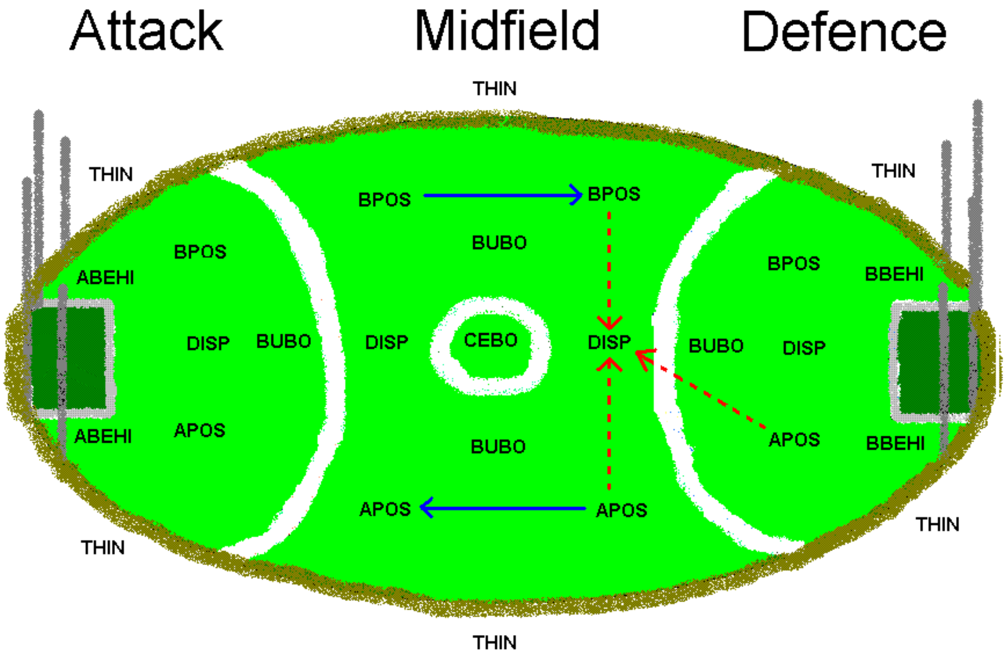
Figure 13.17: North Melbourne: significantly different transitions to competition



The profile of North Melbourne is no different when playing interstate or in Victoria. Furthermore, they stamp themselves as an average side with only their opponent's ability to hit open team mates in the midfield (+2.1%) rather than kick to contests (-1.7%), and an increased propensity to force secondary bounces around zone 3 ball-ups (+2.8%), significantly different to the competition.

13.14 Port Adelaide Football Club

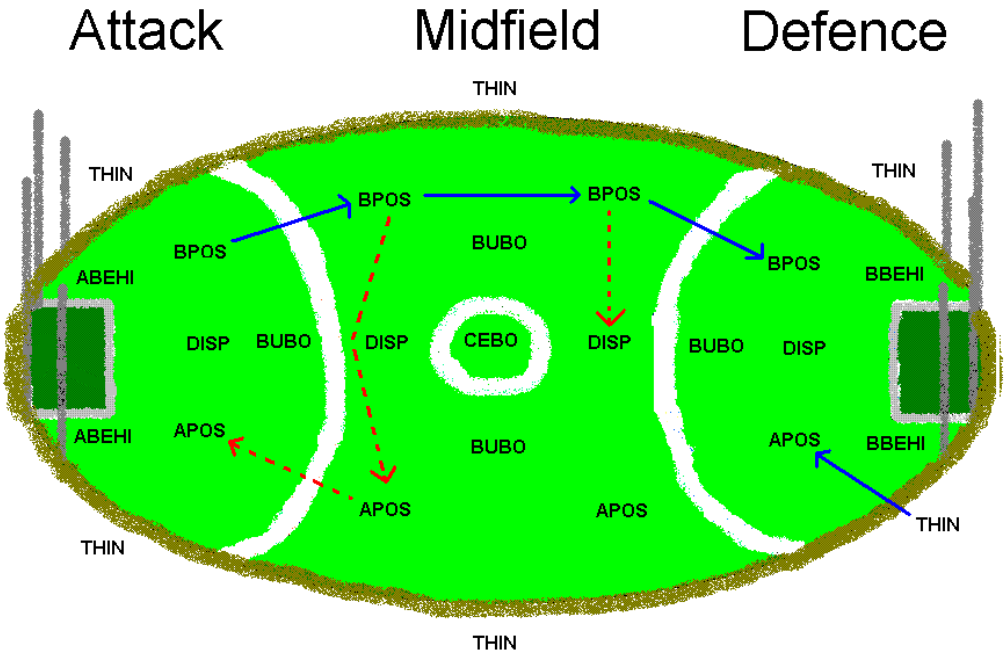
Figure 13.18: Port Adelaide: significantly different transitions to competition



Port Adelaide is another club who play a similar style on the road as they do at Football Park. Where they do differ from the competition is their encouragement of uncontested possession in the midfield. They resist disputed ball with a below average rate of contested ball when coming out of defence (-2.1%) as well as a reduced rate for themselves (-1.8%) and their opposition (-1.8%). At the same time, Port has an above average rate of retaining the ball in the midfield (+2.9%) and so do their opponents (+2.9%). Whilst this openness in the midfield has brought some success, tightening up in the midfield could prove very beneficial to the club.

13.15 Richmond Football Club

Figure 13.19: Richmond: significantly different transitions to competition



Richmond have struggled to win games presumably because they have a below average rate for finding team mates on attack when entering their attacking 50 (-1.1%). Furthermore, their opponents clear easily from defence (+3.4%) and then find open team mates in the midfield (+2.2%) as well as on attack (+1.3%). These increased rates mean that in the midfield, less disposals are put into dispute by the opposition (-2.2%) or turned directly over to Richmond (-1.0%). One positive for Richmond is their increased ability to win the ball at zone 3 throw-ins (+13.2%). Richmond needs to put more pressure on the opposition when the ball is in the midfield. Clearly this could result in them gaining more ball in the midfield than they do at present and therefore lifting their attacking entries.

13.16 St. Kilda Football Club

Figure 13.20: St. Kilda: significantly different transitions to competition

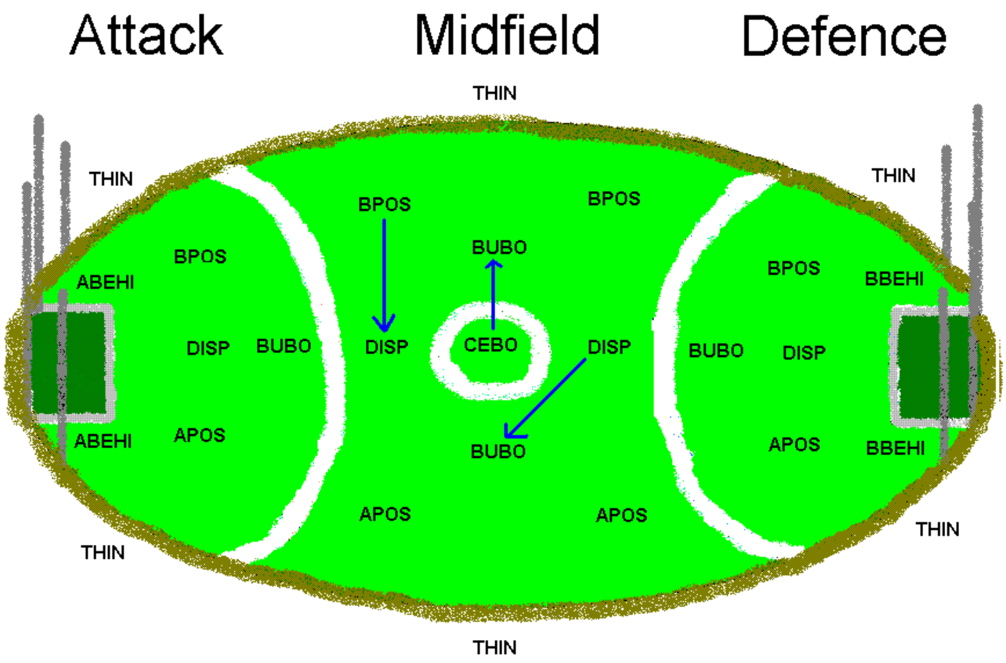
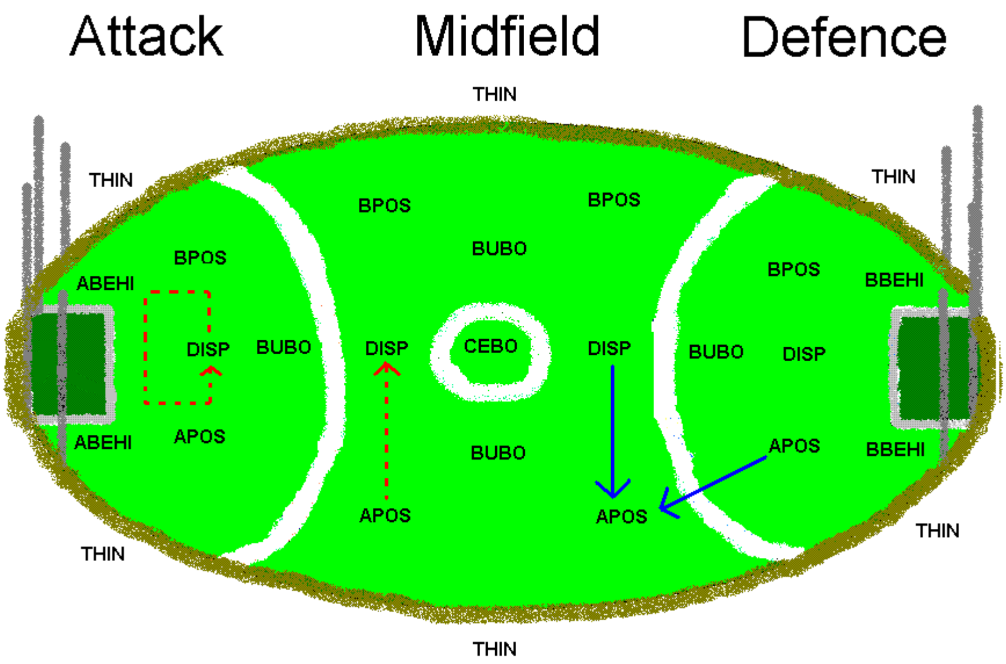


Figure 13.21: St. Kilda: significantly different transitions for interstate travel



St. Kilda plays a style that creates increased ball-ups in the midfield. They are above average for forcing bounces both in general play (+2.5%) and from centre bounces (+3.0%). Their opponents create extra disputed ball in the midfield with an above average rate (+1.4%). When traveling interstate, St. Kilda displays a tendency to increase scrappy ball on attack (3.2%) and less effective disposal in the midfield (3.7%). In Victorian games they are better at winning the ball from dispute in the midfield (5.6%) and finding open team mates when coming out of defence (7.3%).

13.17 Western Bulldogs Football Club

Figure 13.22: Western Bulldogs: significantly different transitions to competition

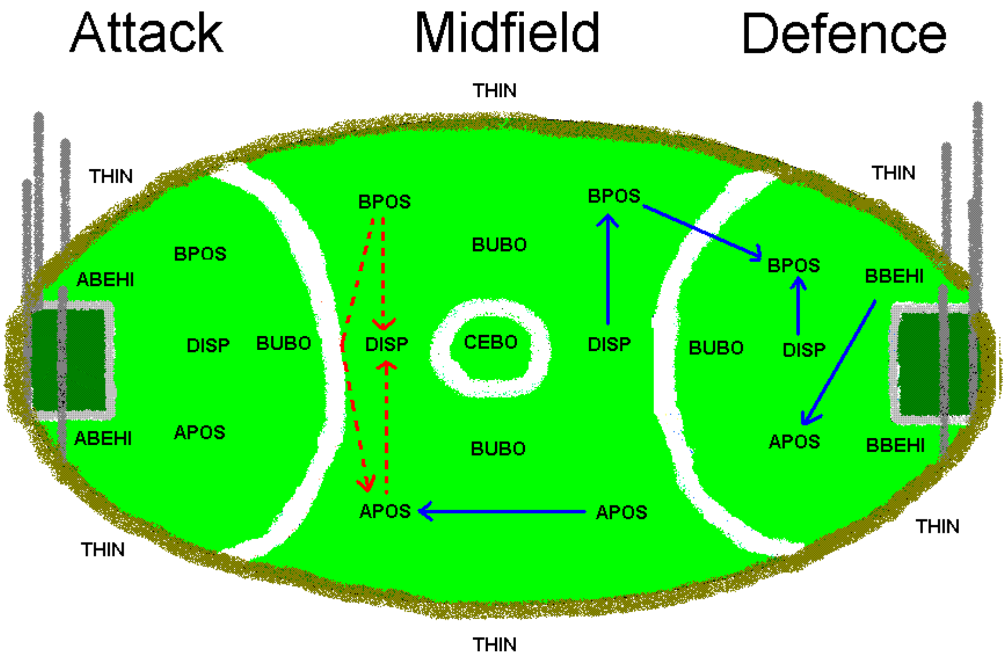
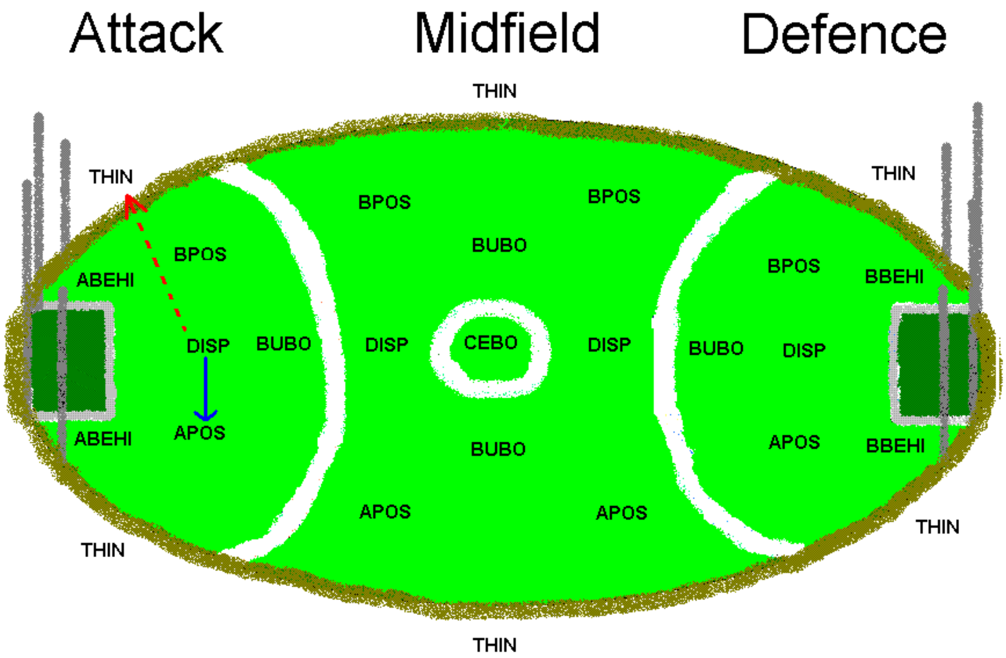


Figure 13.23: Western Bulldogs: significantly different transitions for interstate travel



The Bulldogs enjoy an open style of play and this is reflected in the transition probabilties that are significantly different to the competition. In the midfield, there is a propensity, like Port Adelaide, to discourage contested possession. Both the Bulldogs (-2.7%) and their opponents (-1.6%) have below average rates for putting the ball into dispute whilst the Dogs' rate of retaining the ball in zone 2 is above average (+3.2%). At the same time the rate of error by the opposition in zone 2 is below average (-1.0%) and their ability to win disputed ball is above average (+2.8%). The opposition also have an above average rate for hitting open team mates when going inside 50 (+1.4%) and winning disputed ball in this area (+5.3%). The Bulldogs like to find free targets in defence when kicking the ball in (+6.8%) at an above average rate. When traveling, the Bulldogs are less inclined to win zone 1 disputed ball than in Melbourne (9.1%) and more likely to increase the rate of attacking throw-ins (7.5%). In order to continue to improve their winning percentage, the Bulldogs need to address their defensive issues. Not just in the defensive zone but also in the midfield, they need to play closer and pressure their opponents more than they currently are.

13.18 West Coast Football Club

Figure 13.24: West Coast: significantly different transitions to competition

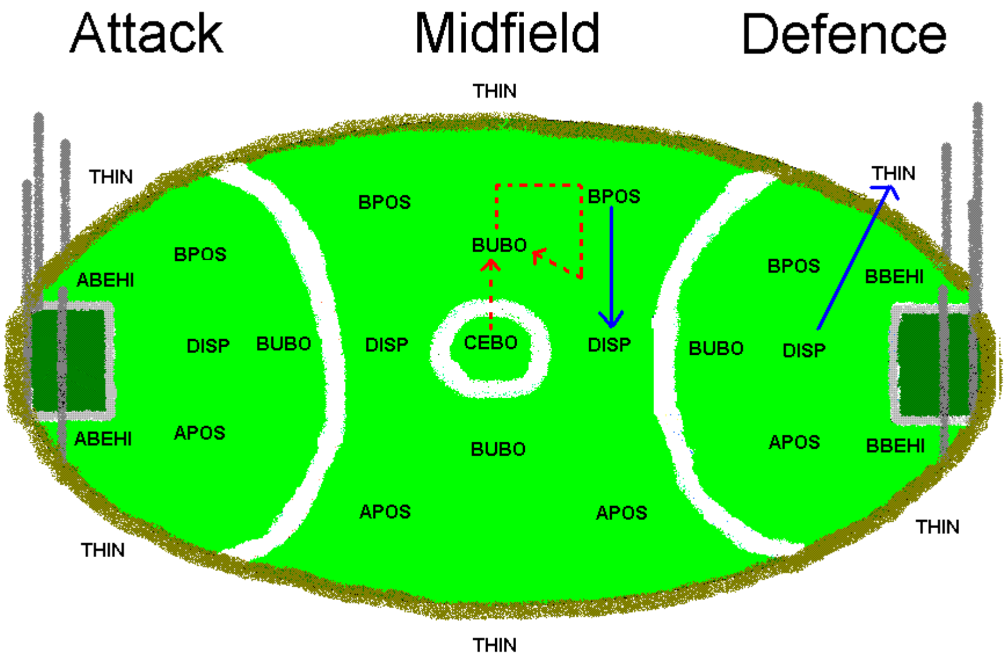
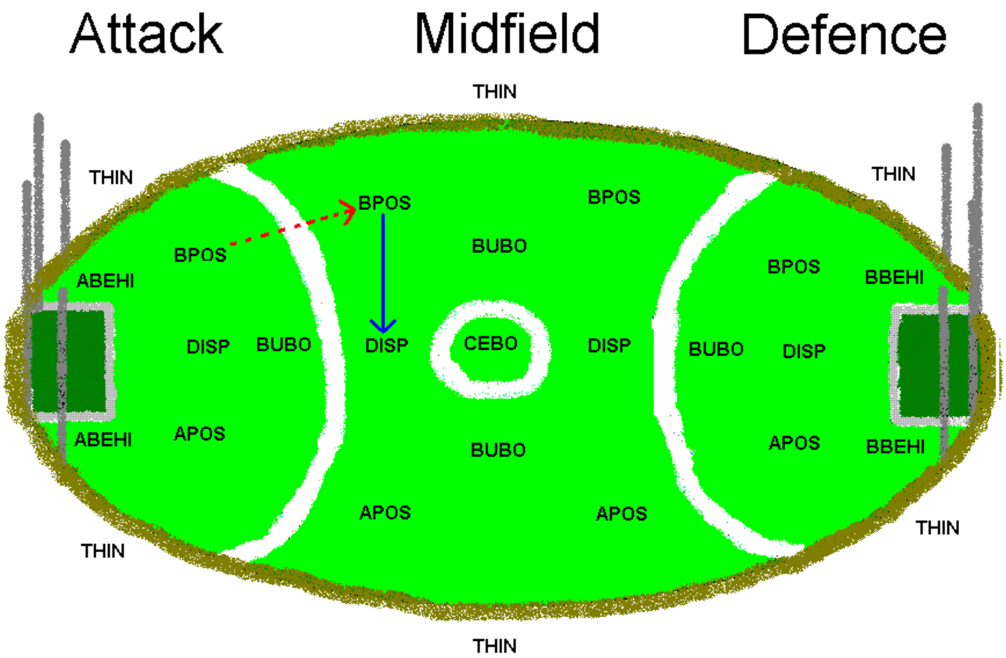


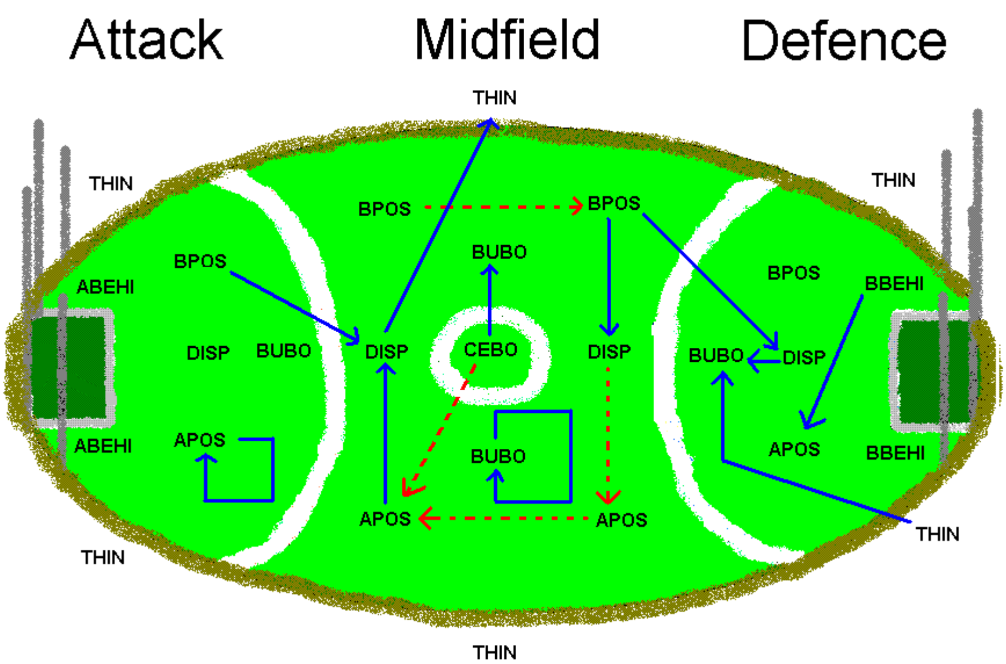
Figure 13.25: West Coast: significantly different transitions for interstate travel



The West Coast likes to play an open style in the midfield that limits secondary stoppages. They are below average in their games for secondary ball-ups following centre bounces (-3.7%) and ball-ups (-4.8%). At the same time they make their opponents kick to contests in zone 2 at an increased rate (+1.3%). In defence, the rate of throw ins forced is above average (+3.8%). When playing at Subiaco, the West Coast's opposition find it hard to hit open team mates in the midfield when coming out of defence (6.6%) as well as when they have possession in zone 2 (2.6%). As the West Coast have the best winning percentage it is hard to suggest improvement. One area they need to address is the lack of pressure that they apply to the opposition in interstate games when they are coming out of defence and in the midfield.

13.19 Sydney Football Club

Figure 13.26: Sydney: significantly different transitions to competition



With a completely unique style, it is little wonder Sydney's profile it significantly different to the competition in a number of areas. When attacking, Sydney likes to recycle the ball around the zone, presumably to players in better scoring positions (+4.2%). They clearly enjoy a game that involves a lot of disputed ball with increased rates for the opposition kicking to contests when coming out of defence (+3.6%), in the midfield (+3.8%) and when going on attack (+2.0%). Sydney like to force stoppages in the midfield with an above average rate of throw-ins (+3.1%), and secondary ball-ups from centre bounces (+5.0%) and bounces (+4.0%). This is also evident on defence with increased ball-ups in general play (+3.6%) and from throw-ins (+7.7%). As a result of this style, the rate of contested ball in the midfield from Sydney possession is above average (+2.3%), whilst the rate of uncontested possession for Sydney (-2.1%) and their opposition (-4.7%) is significantly below the competition average. Despite seeming to enjoy this clustered style of play, Sydney are below average at winning the ball from centre bounces (-4.7%) or when the ball is in dispute in the midfield (-4.2%). Sydney is another team which prefers to take the short, safe option when kicking the ball in (+10.0%).

13.20 Comparison of team styles

With comparisons made for each club to the competition average as well as when they play in their home state and interstate, this section will contrast the style of each club in the competition. In a manner similar to the venue analysis in section 13.2, the CATMOD procedure has been used to perform a log linear analysis. The overall transition matrix used for each club from the past two seasons has been used to compare to every other club's in the competition to see if they are statistically different. As a result 120 comparisons were made matching up every club in the competition with every other club, necessitating the use of the Bonferroni correction. Of the 120 tests, there were 14 comparisons that indicated a similar style of play. Table 13.5 contains the clubs from the AFL competition and the clubs they are not significantly different from.

Table 13.5: AFL clubs and their opponents who play a similar style

Club	Clubs not significantly different ($p \geq 0.0004$)
AFC	BFC, COFC, EFC, FFC, MFC, NMFC, PAFC, RFC
BFC	AFC, EFC, PAFC
CAFC	-
COFC	AFC, NMFC, PAFC, SKFC
EFC	AFC, BFC, FFC, GFC, MFC, NMFC, PAFC, WCFC
FFC	AFC, EFC, GFC, MFC, NMFC, WCFC
GFC	EFC, FFC, HFC, MFC, SKFC, WCFC
HFC	GFC
MFC	AFC, EFC, FFC, GFC, NMFC, PAFC, WCFC
NMFC	AFC, COFC, EFC, FFC, MFC, PAFC
PAFC	AFC, BFC, COFC, EFC, MFC, NMFC
RFC	AFC, WBFC
SKFC	COFC, GFC
WBFC	RFC
WCFC	EFC, FFC, GFC, MFC
SFC	-

Leading on from the venue analysis, the two Perth sides, who share Subiaco, are not significantly different in their playing styles, but perhaps the West Coast are more consistent than Fremantle. Also the two sides from Adelaide who share Football Park do not play dissimilar styles. It seems as though teams who are not significantly different play the majority of their matches at venues which were shown earlier to not be significantly different in the style of play they produce. Obviously, the interaction between team and venue contributes to the not dissimilar style for venues and or teams. What is noticeable about Table 13.5 is that Sydney is significantly different to every other club in the competition. Given that the S.C.G. was not significantly different to a number of other venues, it was expected that Sydney would have been not dissimilar to other clubs in the competition. However, given the number of transitions Sydney had that were significantly different transitions from the competition average in Figure 13.31, perhaps this result is not unexpected.

The clubs ranked in the middle of Table 13.4, and therefore considered to be average, such as Geelong, Melbourne, Adelaide, the Kangaroos, Fremantle and Essendon all have at least six clubs that they are not dissimilar to. Aside from these clubs, only Port Adelaide, West Coast and Collingwood have more than three comparisons that are not significantly different. This suggests that the sides in the middle of the table are there because they play an 'average' style of play that is not dissimilar to numerous other clubs in the competition.

13.21 Summary

This chapter has shown the analytical benefits available to AFL clubs with the advent of the Markov process models presented earlier. These models can be used to amalgamate the performance of a club under differing scenarios and compare these performances to an agreed upon benchmark. In this chapter, analysis was done on the past two seasons performance of every club and comparing their style of play to the competition average. Statistical tests were carried out to see where the clubs differed significantly from the average. AFL clubs could impose their own conditions (e.g. winning v losing games) and analyse their performance identifying where they perform well or poorly and this could be done on different time frames. For instance, individual games could be looked at, as could seasons and longer periods such as a coach's tenure.

Analysis was done building on Chapter 6, looking at home advantage to analyse whether clubs perform differently when they are playing locally compared to when they travel interstate. This type of analytical tool could be very important for AFL clubs to understand their performances by location. Areas could be identified where they need to improve when traveling interstate to play, to ensure the performance on the road is at the same level as their home state performances.

The transition matrices for clubs and venues were compared using log-linear analysis which identified similarities between venues and clubs. This was another example of the power of the model in being able to group clubs according to whether they played a style that was similar. This technique could be used to predict transition probabilities for clubs according to where a match is being played and who the opposition is.

This chapter has presented further applications of the models to the AFL clubs that build on those already presented in previous chapters. This analysis could also be used to thoroughly investigate the psychological effects of home advantage and whether teams are more prepared to defend their territory when playing at home over when they are visiting another team's home venue. It should be noted that the differences presented in this chapter required a very conservative level of significance due to the size of the Bonferroni correction. If a commercial application of this chapter was to eventuate, it would be of more benefit to AFL clubs using a higher level of significance and being able to identify areas that are above or below the relevant benchmark. Furthermore, the results would become even more robust with the inclusion of more data, which, in time will occur. This chapter concludes the development of the Markov process models and the final chapter will provide a conclusion to this thesis suggesting where the research can be taken from here.

Chapter 14: Conclusion

14. 1 Overview

The focus of this thesis has been on trying to forecast the game of Australian Rules football using real time match statistics as collected by the official statisticians for the AFL. As the sport in question, Australian football is a game unique to Australia, the scope of this thesis may be seen as having little or no global application or interest. However, the models presented in this thesis should not necessarily be limited to Australian football and this will be alluded to later in this chapter.

The background for this research was set in Chapter 3 with a detailed account of the systems and processes used by CD to accurately and efficiently collect match information. They have clearly changed the face of information collection of AFL statistics available to AFL clubs and supporters as well as different media outlets. Champion Data's revolutionary approach to AFL information collection has facilitated the unique models developed in this thesis for approximating AFL football.

Using the data that CD collect for AFL matches, a data set was put together containing the match statistics for 1,110 games between 1998 and 2003. This data set was used in Chapter 4 to carry out some preliminary exploration on relationships in the data set. Correlations were investigated between scoring events for teams on attack and defence. This was also done for goals and behinds between clubs and their opposition. These results were used to set the framework for the pre-match prediction model and to decide whether independence or dependence between scoring events would be assumed.

This analysis was built on in Chapter 5 where statistical distributions were fitted to AFL scoring events. This analysis was driven by the body of research relating to scoring events in soccer, particularly the work from the late 1980s (Baxter and Stevenson, 1988). It was found that the negative binomial distribution best approximated overall competition scoring, however it was clearly outperformed by the Poisson distribution for approximating individual team scoring and concession rates. This analysis was the first of its kind applied to AFL scoring rates and was also unique in that concession rates were analysed. This had not previously been done in the literature relating to fitting statistical distributions to sporting scoring events. The discoveries made in this chapter enabled the development of the Markov process models presented later in the thesis, which assumed constant scoring rates in the transition probabilities used for approximating match events.

Home advantage in the AFL competition was investigated in Chapter 6 using traditional measures as well as some different indicators based on match statistics. The advantage home teams enjoy on the scoreboard was quantified along the same lines that Clarke used and not surprisingly the results were very similar (Clarke, 1997). A different approach was taken to home advantage that categorized match statistics and analysed whether teams at home enjoyed an advantage in performance statistics like soft or hard possessions. Whilst there was a home advantage in most cases it was not as substantial as expected. This analysis was a precursor to the capabilities of the Markov model presented later in the thesis that determined in which areas of play teams performed differently when playing at home and away.

14. 2 Modeling AFL football

After using CD’s data to investigate various relationships that exist between variables in the data set, the focus of the thesis shifted to modeling AFL match results and trying to assign a margin of victory to a match and a probability of victory. This was initially done using a pre-match model that had its groundings in the work of Clarke, who later teamed up with Stefani (Stefani and Clarke, 1992) and Bailey (Bailey and Clarke, 2004). This model differed from their work as it relied on the results of Chapter 5 to use a negative binomial regression technique to predict team scores using the interaction between attack and defence of the competing clubs. The models referred to above rely solely on a team’s attacking rating and this is where the present pre-match model differs from what is already present in the literature relating to Australian football. It was also shown that the performance of the model compares favourably with the established models in the literature for Australian football.

As outlined by the title of this thesis, the aim was to produce a model that reacted to game events as they took place. The aim of developing this dynamic application meant that a time-structured model was required as opposed to a static model. Initially, it was thought that multivariate techniques would be pursued in order to develop the model; however, this idea was replaced by a desire to approximate the game using a Markov process model. This desire stemmed from the work of Hirotsu (Hirotsu, 2002), particularly relating to his work on soccer. It was firmly believed that the game of Australian football lent itself perfectly to being broken into states of play, which could then be used to calculate transition probabilities using the data collected by CD.

The first implementation of the model did not allow for location on the ground and was a basic eight state model that included the important states of the game refined by watching games and analyzing CD's transaction files. The results that this model produced were beyond what was expected with good accuracy displayed when the derived transition probabilities were used to simulate matches. It was reassuring to discover that a Markov process did such an accurate job of reflecting match events when applied to AFL football.

In discovering that the simple eight state Markov process model provided an excellent approximation to AFL football matches, the thesis focused on practical applications of the model that could be developed for clients. Two chapters were devoted to various applications of the global model including post-match analysis and in-game prediction, using match events to calculate transition probabilities. It is hoped that these applications can be utilised not only by football clubs but also by the media and other football-related industries, such as sports bookmaking.

The impressive results associated with the global model were built on with the introduction of ground location in a zone model. This increased the number of states to 18, providing more accurate approximations to match results. As well as being able to replicate the applications of the global model, there were further applications relating to strategical decisions according to location on the ground. The ability to quantify the importance of match statistics, such as inside 50s, as well as identifying optimum playing styles was demonstrated in Chapter 12. These modeling capabilities provide invaluable tools for coaches planning their strategy for upcoming matches.

The final analysis that was presented using the Markov model involved various comparisons of team playing styles under differing scenarios. Firstly, each team was compared to the competition average to investigate where the strong teams derived their strength and, similarly, why the weak teams were inferior. This analysis can help coaches and commentators understand the most important facets of the game and suggest where the most attention should be paid in order to improve performance. Secondly, the performance of teams at home and on the road was investigated following on from the work done in Chapter 6. This work identified where and how team styles are affected by travel. Again, this could be an important training and planning tool as it identifies particular areas that can be focused on.

In summary, the models developed in this thesis break new ground in approximating Australian football. Never before has a dynamic model been applied to Australian football that displays the same accuracy as these models and can be used for a number of applications, both when a match is in progress and after its completion. The techniques developed in this thesis should greatly benefit coaches in being able to identify and quantify individual passages of play as well as overall strategies that optimize their team's chances of winning. The work has received substantial media coverage (Butler, 2005) and the Adelaide Football Club has already begun to use the information derived from the global model, with this association to continue in the 2006 season. It is envisaged that the benefits of the zone model will be provided to the Adelaide Football Club on a week-to-week basis for upcoming matches and a revision of completed matches. For the 2006 season, the Collingwood Football Club are to receive weekly reports using the model to help improve their competitiveness on the field. As seen in

Chapter 13, they are a club who has not enjoyed a lot of success in past seasons and would benefit greatly from the analysis contained in that chapter.

14. 3 Limitations of the research

Although the results of the models are extremely promising there are several limitations of the research. The major concern is the human element in the collection of match information. It is accepted that there are going to be errors made in calling matches. These errors are minimal as evidenced by the lack of edits required after the match is completed; however, in the process of coding matches into transition probabilities, which are a completely automated process, there were several instances where transitions took place that were not possible. These events need to be accounted for in the coding process; however in some cases they may have been overlooked. This should be recognized when applying the models and when acting on the results.

The models presented in Chapters 8 and 11 are first-order Markov process models that rely only on the present state to predict upcoming events. Clearly, the bulk of the dynamics of Australian Rules football is captured by the first-order Markov process. It was acknowledged in Chapter 8 that the game of AFL football may be better approximated using a higher order model that took into account play chains. The limitations associated with this approach were the introduction of many extra states and the reduction of data for each cell in the transition matrix. In time, a model that includes more than just the present state for determining future events, may further improve the fit of a Markov model to Australian football.

It must also be noted that no reference was made in this thesis to the length of time spent in a state. A time-structured approach would allow for more accurate attention to the length of matches and the amount of time spent in any state. This could give a better understanding of the various states present in the model by giving an indication as to how long a transition takes from one state to another. Furthermore, the inclusion of time into the analysis would enable for better profiling of team styles and whether they adopt a free-flowing approach or like to shut the game down, in the process slowing down match events.

Finally, although this thesis has presented groundbreaking research relating to AFL football and the appropriateness of the models rests on firm support, the research was limited by time constraints. The next section identifies extensions to this research that were considered for inclusion but could not be contained in this thesis due to time constraints.

14. 4 Extensions of the research

There are several areas where it is hoped this research can be extended in a commercial environment. In this thesis, the progression was made from a ‘whole of ground’ approach in Chapter 7 to a model that used probabilities according to one of three zones on the field in Chapter 11. In constructing the zone model, it was hypothesized that the division of the field into three zones lengthwise, i.e. a left wing, right wing and centre corridor, would strengthen the model. This would see the ground divided into nine zones and an increased number of states, but would give an indication of whether teams had a preference for playing wide or down the middle. A recent introduction to CD’s data collection is the recording of co-ordinates for every event on the field. This kind of information could be used to construct the zones referred to above and derive transition probabilities for a model that used these extra zones.

The models presented herein are based on team statistics and provide an excellent approximation. With the richness of information that CD collect for each match, a model could be developed that makes reference to individual player probabilities. Career data could be used to derive these probabilities according to the states contained in the model. This would be similar to the baseball model developed that optimized batting lineup using player information (Bukiet, Harold and Palacious, 1997). A player based model would be invaluable to coaches as it would identify perceived strengths and weaknesses of individual players. It could also be used to investigate the loss of a player through injury and the affect on the team, or whether recruiting a certain player may be beneficial to the team set up. Obviously, a model that contained individual transition probabilities for every player would contain a very large number of states and computationally may not be worthwhile.

Another area that could be worthwhile investigating is the relative benefits of improving a given transition by a certain percentage. The best way to gauge whether improving a give transition benefited the team would be to measure the increase on the scoreboard and consequently the increase in expected probability of victory. For example, a 1% improvement in one transition may affect the scoreboard as much as 2% improvement in another transition. A systematic study could be used to investigate this based on the work presented in this thesis.

Finally, further work is warranted investigating the ability to develop a technique that predicts the transition probabilities for a Markov process model prior to a match. Although it was shown in Chapter 7 that pre-match predictions reach a ceiling level in terms of the accuracy of predictions, a model that uses transition probabilities predicted using past match data might perform at a level comparable to the model presented in that Chapter. Furthermore, it would be expected that such an approach to match prediction would be improved by adjusting the transition probabilities as match events take place. These updated predictions would be derived in a manner similar to the dynamic models presented in Chapters 10 and 11.

14. 5 Conclusion

This thesis has presented an innovative and revolutionary way to approximate match events and outcomes in the AFL competition. Never before have models been presented that have had the analytical capability to investigate AFL matches in the ways presented in this thesis. It is hoped that over the next few years the techniques and applications presented will find their way into the coaches' box on match day as well as the public arena via media outlets and sports bookmakers. There is no reason why this will not happen given the amount of time and resources that have already been invested in the research and the promising feedback and interest from clients who have been exposed to the research, albeit in a preliminary sense.

Finally, although this analysis has been centered solely on Australian football, there is no reason why the techniques could not be applied to other continuous sports such as rugby league or union. The only requirement is a level of richness in the data collected from these sports that is present in CD's AFL information collection. Thought has already been put into the relevant states for these sports and the author would welcome the opportunity to apply Markov process techniques to data from these sports.

References

Abt, G. A., Dickson, G. and Mummery, W.K. (2002). Goal Scoring Patterns over the course of a match: An analysis of the Australian national soccer league. In W. Spinks, Reilly, T. and Murphy, A. (Ed.). *Science and Football IV.* London, Routledge, 106 - 111.

Agnew, G. A. & A.V., C. (1994). Crowd effects and the home advantage. *International Journal of Sport Psychology,* (25) 53-62.

Australian Bureau of Statistics (2003). *Sport and Recreation: A statistical overview, Australia (Catalogue No. 4156),* Canberra, Australian Government Publishing Service.

Bailey, M. & Clarke, S. R. (2004). Deriving a profit from Australian Rules football: a statistical approach. In H. Morton (Ed). *Seventh Australian Conference on Mathematics and Computers in Sport*. Massey University: Palmerston Nth. 48-56.

Baxter, M. & Stevenson, R. (1988). Discriminating between the Poisson and negative binomial distributions: an application to goal scoring in association football. *Journal of Applied Statistics,* 15 (3) 347-354.

Bellman, R. (1977). Dynamic Programming and Markovian decision processes, with application to baseball. In S. P. Ladany & Machol, R. E. (Eds.). *Optimal strategies in sports.* Amsterdam, North Holland, 77-85.

Blainey, G. (2003). Whither our beloved game? *The Age*. Melbourne. August 29, 2003.

Boronico, J. S. (2000). Play Calling Strategy in American Football: A Game-Theoretic Stochastic Dynamic Programming Approach. *Journal of Sport Management*.

Bracewell, P. (2002). Quantification of Individual Rugby Player Performance through Multivariate Analysis and Data Mining. Doctor of Philosophy. *Information and Mathematical Sciences*. Albany, Massey University.

Bukiet, B., Harold, E. & Palacious, J. (1997). A Markov chain approach to baseball. *Operations Research,* 45 (1) 14 - 23.

Butler, G. (2005). Number crunching in the AFL. *Today Tonight, Channel 7, Perth*.

Champion Data. (2004). Premiership Standards Report II. T. Hopkins. Melbourne.

Champion Data. (2005). Premiership Standards Report I. T. Hopkins. Melbourne.

Clarke, S. R. (1988). Dynamic Programming in one day cricket - Optimal scoring rates. *Journal of the Operational Research Society,* 39 (4) 331-337.

Clarke, S. R. (1993). Computer forecasting of Australian Rules football for a daily newspaper. *Journal of the Operational Research Society,* 44 (8) 753-759.

Clarke, S. R. (1997). Performance modelling in sport. Doctor of Philosophy. *School of Mathematical Sciences*. Hawthorn, Swinburne Univeristy.

Clarke, S. R. (2005). Home advantage in the AFL. *Journal of Sports Sciences,* 23 (4) 375-385.

Clarke, S. R. & Norman, J. M. (1995). Protecting a weak batsman from taking strike at the start of a new over - A simple Dynamic Programming model in cricket. *Asia Pacific Journal of Operational Research,* 15 93-108.

Clarke, S. R. & Norman, J. M. (1998). When to rush a behind in Australian Rules football: a Dynamic Programming approach. *Journal of the Operational Research Society,* 49 (5) 530-536.

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

Cox, D. R. M., H.D. (1965). *Stochastic Processes,* London, Methuen,

Croucher, J. S. (1992). Winning with Science. In N. de Mestre (Ed.). *Mathematics and Computers in Sport.* Gold Coast, Qld, Bond University, 1-22.

Croucher, J. S. (2000). Player ratings in one-day cricket. In G. Cohen & Langtry, T. (Ed). *Fifth Australian conference on Mathematics and Computers in Sport*. Sydney. 28-36.

Devaney, J. (2002). 1970 VFL Grand Final: Carlton vs. Collingwood, 'Handball, Handball, Handball!' *http://www.fullpointsfooty.net/1970_vfl_grand_final.htm*.

Dixon, M. J. & Coles, S. G. (1997). Modelling Association football scores and inefficiencies in the UK football betting market. *Applied Statistics,* 46 265-280.

Dixon, M. J. & Robinson, M. E. (1998). A birth process model for association football matches. *The Statistician,* 47 (3) 523-538.

Dobson, S. & Goddard, J. (2003). Persistence in sequences of football match results: A monte carlo analysis. *European Journal of Operational Research,* 148 247-256.

Dyte, D. (1998). Constructing a plausible test cricket simulation using available real world data. In N. de Mestre & Kumar, K. (Ed). *Fourth Australian Conference on Mathematics and Computers in Sport*. Gold Coast, Qld. 153-159.

Dyte, D. & Clarke, S. R. (2000). A ratings based model for World Cup soccer simulation. *Journal of the Operational Research society,* 51 993-998.

FAPL. (2002). English clubs top European attendances, league crowd figures break new records in the FA Barclaycard Premiership. http://www.premierleague.com/en/files/press_releases/RecordAttendances23_01_02.pdf.

Forbes, D. & Clarke, S. R. (2004). A seven state Markov process for modeling Australian Rules football. In H. Morton (Ed). *Seventh Australian Conference on Mathematics and Computers in Sport*. Palmerston North: Massey University. 148-158.

Freeze, R. A. (1974). An analysis of baseball batting order by Monte Carlo simulation. *Operations Research,* 22 728 - 735.

Glasson, S., Jeremiejczyk, B. & Clarke, S. (2001). Simulation of Women's Beach Volleyball Tournaments. *ASOR Bulletin,* 20 (2) 2-8.

Greenhough, J., Birch, P. C., Chapman, S. C. & Rowlands, G. (2002). Football goal distributions and extremal statistics. *Physica A,* 316 615 - 624.

Harville, D. A. (1980). Predictions for National Football League games via linear-model methodology. *Journal of the American Statistical Association,* 75 516-524.

Hayes, M. & Norman, J. M. (1984). Dynamic Programming in orienteering: Route choice and the siting of controls. *Journal of the Operational Research Society,* 35 791-796.

Hirotsu, N. (2002). A Formulation of Optimal Substitution Strategies Using a Markov Process Model in Baseball and Soccer. Doctor of Philosophy. *Management Science*. Lancaster, Lancaster University.

Hirotsu, N. & Wright, M. (2003). An evaluation of characteristics of teams in association football by using a Markov process model. *The Statistician,* 52 (4) 591 - 602.

Hopkins, T. (1996). A new standard in stats. *The Herald Sun*. Melbourne. April 2.

Hopkins, T. (1997). Statisticians figure out another line of attack. *The Australian Financial Review*. Sept.

Hopkins, T. (1998a). Statistical analysis makes for good sport. *The Australian Financial Review*. December 5-6.

Hopkins, T. (1998b). Swinburne University's AFL tips. *The Australian Financial Review*. April 17.

Howard, R. A. (1960). *Dynamic Programming and Markov Processes,* New York, M.I.T. Technology Press and John Wiley and Sons,

Jinshan, X. (1986). The analysis of the techniques, tactics and scoring situations of the 13th World Cup. *Sandong Sports Science and Technique,* (April) 87-91.

Jinshan, X., Xiaoke, C., Yamanaka, K. and Matsumoto, M. (1993). Analysis of the goals in the 14th World Cup. In T. Reilly, Clarys, J. and Stibbe, A. (Ed.). *Science and Football II.* London, E. and F.N. Spon, 203 - 205.

Johnson N.L. & Kotz, S. (1969). *Discrete Distributions,* New York, Wiley,

Karlis, D. a. N., J. (2003). Analysis of sports data using bivariate Poisson models. *The Statistician,* 52 381-393.

Ker, P. (2004). Swans have the right finals brand: veteran. *The Age*. Melbourne. September 6.

Koning, R. H., Koolhaas, M., Renes, G. & Ridder, G. (2003). A simulation model for football championsips. *European Journal of Operational Research,* 148 268-276.

Leake, R. J. (1976). A method for ranking teams with an application to college football. In R. E. Machol, Ladany, S. P. & Morrison, D. G. (Eds.). *Management Science in sports.* North Holland, 27-46.

Lee, A. (1999). Modelling rugby league data via bivariate negative binomial regression. *Australian and New Zealand Journal of Statistics,* 41 (2) 141-152.

Lee, A. J. (1997). Modelling scores in the Premier League: is Manchester United really the best. *Chance,* 10 (1) 15-19.

Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game,* New York, W. W. Norton & Company, 288.

Lock, R. (2000). CHODR - College Hockey Offensive/Defensive Ratings. http://it.stlawu.edu/~chodr/.

Lonergan, D. (2005). Debate raging about standard of footy. http://www.abc.net.au/sport/features/2004/s1364148.htm.

Lovett, M. (Ed.) (2004). *The official statistical history of the AFL,* Melbourne, AFL Publishing.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica,* 36 109-118.

Moroney (1956). *Facts from Figures,* Middlesex, Penguin Books,

Morrison, D. G. (1976). On the optimal time to pull the goalie: A poisson model applied to a common strategy used in ice hockey. In R. E. Machol, Ladany, S. P. & Morrison, D. G. (Eds.). *Management Science in Sports.* Amsterdam, North-Holland, 137-144.

Norman, J. M. (1985). Dynamic Programming in tennis - when to use a fast serve. *Journal of the Operationsal Research Society,* 36 75-77.

Norman, J. M. (1999). Markov Process Applications in sport. In (Ed). *IFORS conference*. Beijing China.

Patrick, J. (1985).The capture and analysis of football in real-time. *ACS Bulletin*. 9-11.

Patrick, J. (1992). The marriage of mathematics and computer technologies for sport improvement. In N. de Mestre (Ed). *Second Australian Conference on Mathematics and Computers in Sport*. Gold Coast, Qld. 61-70.

Pollard, R. (1985). Goal Scoring and the Negative Binomial Distribution. *Mathematical Gazette,* 69 (9) 45-47.

Pollard, R., Benjamin, B. & Reep, C. (1977). Sport and the negative binomial distribution. In S. P. Ladany & Machol, R. E. (Eds.). *Optimal Strategies in Sports.* Amsterdam, North Holland, 188-195.

Reep, C. & Benjamin, B. (1968). Skill and chance in association football. *Journal of the Royal Statistical Society A,* 131 581-585.

Reep, C., Pollard, R. & Benjamin, B. (1971). Skill and chance in ball games. *Journal of the Royal Statistical Society A,* 134 623-629.

Reilly, T. (1996). Motion analysis and physiological demands. In T. Reilly (Ed.). *Science and Soccer.* London, E. and F.N. Spon, 65 - 81.

Ridder, G., Cramer, J. S. & Hopstaken, P. (1994). Down to Ten: estimating the effect of a red card in soccer. *Journal of the American Statistical Association,* 89 (427) 1124-1127.

Ryan, M. (2005). Sheedy has plan for stoppages. *The Age*. Melbourne. February 17.

Sheahan, M. (2005). Carey sees inside 50s as coach's alibi. *Herald Sun*. Melbourne. June 15.

Stefani, R. T. (1977). Football and basketball predictions using least squares. *IEEE Transactions on systems, man, and cybernetics,* 7 117-121.

Stefani, R. T. (1980). Improved least squares football, basketball and soccer predictions. *IEEE Transactions on Systems, Man and Cybernetics,* SMC -10 (2) 116-123.

Stefani, R. T. & Clarke, S. R. (1992). Predictions and home advantage for Australian rules football. *Journal of Applied Statistics,* 19 (2) 251-261.

Stern, H. S. (1993). Who's Number One? Rating Football Teams. In A. S. Association (Ed). *1992 Proceedings of the Section on Statistics in Sports*. Alexandria. 1-6.

Thomas, A. C. (2006). The Impact of Puck Possession and Location on Ice Hockey Strategy. *Journal of Quantitative Analysis in Sports,* 2 (1).

Tomecko, N. (1999). Player Assignments in Australian Rules Football. Master of Mathematics. *Information of Technology*. Adelaide, University of South Australia.

Trueman, R. E. (1976). A computer simulation model of baseball: with particular application to strategy analysis. In R. E. Machol, Ladany, S. P. & Morrison, D. G. (Eds.). *Management Science in Sports.* Amsterdam, North-Holland, 1-14.

Trueman, R. E. (1977). Analysis of Baseball as a Markov Process. In S. P. a. M. Ladany, R.E. (Ed.). *Optimal Strategies in Sports.* Amsterdam, NorthHolland Publishing Co.,

Wright, C. (1996). Boot-up there, Cazaly! *The Australian Financial Review*. September 20.

Appendix 1 – Publications and presentations relevant to research

Publications

Forbes, D. & Clarke, S.R. (2004). A seven state Markov process for modeling Australian Rules football. In Morton, H. (Ed.) *Seventh Australian Conference on Mathematics and Computers in Sport.* Palmerston North, Massey University. 148-158.

Forbes, D., Clarke, S.R. & Meyer, D. (2006). AFL football – How much is skill and how much is chance? In Hammond, J. (Ed.) *Eight Australian Conference on Mathematics and Computers in Sport.* Gold Coast, Southern Cross University. In Press.

Presentations

Post match analysis of AFL matches using a Markov process model: a case study (2005). *Australian Postgraduate Workshop on Stochastic Processes and Modelling*. Brisbane, University of Queensland.

An eight state Markov process for modeling Australian Rules football (2004). *Seventh Australian Conference on Mathematics and Computers in Sport*. Palmerston North, Massey University.

Post match analysis of AFL matches using a Markov process approach (2004). *ASOR Student Conference.* Melbourne, RMIT.

Exploratory analysis of scoring rates in the AFL (2003). *ASOR Student Conference.* Melbourne, RMIT.

Media exposure

Butler, G. (2005). Number crunching in the AFL. *Today Tonight.* Perth, Channel Seven.

O'Donoghue, C. (2005). Long way home is the best. *The West Australian – Pre Game*. May 13th, 8-9.

Appendix 2 – AFL clubs names and mascots

Table A2-1: AFL club names

Club	Full Name	Short Name	Mascot
AFC	Adelaide Football Club	Adelaide	Crows
BFC	Brisbane Lions Football Club	Brisbane	Lions
CAFC	Carlton Football Club	Carlton	Blues
COFC	Collingwood Football Club	Collingwood	Magpies
EFC	Essendon Football Club	Essendon	Bombers
FFC	Fremantle Football Club	Fremantle	Dockers
GFC	Geelong Football Club	Geelong	Cats
HFC	Hawthorn Football Club	Hawthorn	Hawks
MFC	Melbourne Football Club	Melbourne	Demons
NMFC	North Melbourne Football Club	Kangaroos	Kangaroos
PAFC	Port Adelaide Football Club	Port Adelaide	Power
RFC	Richmond Football Club	Richmond	Tigers
SKFC	St. Kilda Football Club	St. Kilda	Saints
WBFC	Western Bulldogs Football Club	Bulldogs	Bulldogs
WCFC	West Coast Football Club	West Coast	Eagles
SFC	Sydney Football Club	Sydney	Swans

Appendix 3 – AFL venue comparisons

Table A3-1: P-values between venues

Venue	Gabba	K. Park	M.C.G.	Optus	S.C.G.	Subiaco	Manuka	Marrara	Dock.	York	Olympic
F. Park	0.0125	<.0001	<.0001	<.0001	0.0008	<.0001	0.0514	0.133	<.0001	0.076	0.0001
Gabba		<.0001	<.0001	<.0001	0.0018	<.0001	0.0729	0.0428	<.0001	0.0017	<.0001
K. Park			0.1272	<.0001	0.0327	0.0057	0.0016	0.0511	<.0001	0.1382	0.0038
M.C.G.				<.0001	0.2616	<.0001	0.0406	0.0222	<.0001	0.2387	0.0043
Optus					<.0001	<.0001	0.0138	0.0088	<.0001	0.0002	<.0001
S.C.G.						0.0188	0.0024	0.157	0.0013	0.0849	0.0872
Subiaco							0.0231	0.1414	<.0001	0.0191	0.0036
Manuka								0.0808	0.0047	0.0829	0.0001
Marrara									0.0211	0.2156	0.0416
Dock.										0.0005	0.0001
York											0.0293

Table A3-2: Chi-squared statistic between venues

Venue	Gabba	K. Park	M.C.G.	Optus	S.C.G.	Subiaco	Manuka	Marrara	Dock.	York	Olympic
F. Park	47.4	73.6	73.0	136.3	59.1	76.4	40.0	23.5	122.2	38.1	64.0
Gabba		88.5	78.6	92.7	53.4	70.0	37.1	29.5	74.4	55.1	67.2
K. Park			34.3	82.4	40.8	47.8	51.1	27.5	84.7	33.9	47.9
M.C.G.				92.3	33.4	73.3	41.1	32.0	81.6	32.9	53.0
Optus					90.1	140.7	44.4	35.2	85.6	61.0	88.7
S.C.G.						45.7	51.0	22.8	57.3	37.6	38.6
Subiaco							43.5	23.3	91.2	44.3	50.9
Manuka								25.7	49.9	35.3	60.0
Marrara									32.2	22.4	29.6
Dock.										59.3	65.7
York											42.5

Appendix 4 – AFL club's transition matrices 2004/2005

Table A4-1: Adelaide Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	12.10%	2.90%	4.90%	34.50%	42.20%	1.00%	0.00%	0.00%	0.50%
THIN1	8.30%	4.50%	2.60%	38.30%	41.00%	0.00%	0.00%	0.00%	0.40%
DISP1	9.40%	14.90%	4.40%	34.00%	36.40%	0.00%	0.00%	0.00%	0.00%
APOS1	0.00%	0.10%	14.10%	27.10%	8.70%	21.40%	27.10%	0.00%	0.00%
BPOS1	0.10%	0.10%	4.30%	1.90%	45.50%	0.20%	0.00%	0.00%	0.50%
ABEH1	0.00%	0.00%	2.90%	2.70%	61.10%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	14.50%	0.00%
BUBO2	0.60%	0.00%	0.10%	0.90%	0.80%	0.00%	0.00%	15.20%	2.60%
THIN2	0.20%	0.00%	0.00%	0.70%	1.50%	0.00%	0.00%	13.20%	3.40%
DISP2	0.00%	0.00%	0.10%	0.10%	0.10%	0.00%	0.00%	8.80%	18.30%
APOS2	0.00%	0.00%	10.70%	7.60%	3.60%	1.00%	0.30%	0.10%	0.30%
BPOS2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.40%	0.20%	0.30%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.60%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%	0.50%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.10%	0.60%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	29.20%	0.00%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.00%	1.00%	1.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.40%	1.90%	2.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.30%	0.40%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.90%	0.00%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	11.80%	5.90%	29.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	9.80%	7.50%	16.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	3.50%	44.60%	37.10%	0.00%	0.00%	0.20%	0.00%	0.00%	0.00%
BUBO2	2.50%	37.00%	38.80%	0.20%	0.10%	0.00%	0.70%	0.50%	0.00%
THIN2	2.20%	40.10%	37.10%	0.30%	0.10%	0.10%	0.50%	0.40%	0.00%
DISP2	3.80%	34.10%	34.20%	0.00%	0.00%	0.10%	0.20%	0.20%	0.00%
APOS2	16.70%	53.70%	5.80%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	16.90%	7.20%	54.30%	0.00%	0.00%	9.50%	3.60%	6.60%	0.90%
BUBO3	0.00%	4.00%	1.70%	12.60%	4.00%	0.60%	50.00%	22.40%	4.00%
THIN3	0.00%	0.90%	1.80%	9.00%	4.10%	2.70%	49.10%	30.20%	1.40%
DISP3	0.60%	0.60%	0.40%	8.80%	14.70%	4.90%	34.10%	35.90%	0.00%
APOS3	12.20%	33.30%	4.20%	0.00%	0.10%	4.60%	42.70%	1.70%	0.20%
BPOS3	0.30%	0.20%	0.00%	0.20%	0.10%	13.80%	8.90%	25.80%	21.40%
BBEH3	17.50%	14.70%	5.20%	0.60%	0.00%	0.90%	59.10%	2.20%	0.00%

Table A4-2: Brisbane Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	11.80%	4.60%	1.30%	41.40%	32.90%	0.70%	0.00%	2.60%	1.30%
THIN1	9.00%	4.20%	1.60%	41.30%	41.70%	0.30%	0.00%	0.00%	0.60%
DISP1	6.50%	16.90%	4.40%	38.80%	32.50%	0.00%	0.00%	0.00%	0.00%
APOS1	0.00%	0.20%	13.90%	26.10%	9.10%	20.50%	29.40%	0.00%	0.00%
BPOS1	0.00%	0.40%	4.90%	2.40%	43.60%	0.40%	0.00%	0.00%	0.50%
ABEH1	0.00%	0.00%	1.00%	1.40%	68.40%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%	0.00%	10.90%	0.00%
BUBO2	0.00%	0.10%	0.00%	0.90%	0.40%	0.00%	0.00%	11.00%	3.00%
THIN2	0.10%	0.10%	0.00%	0.30%	1.20%	0.00%	0.00%	9.50%	5.30%
DISP2	0.00%	0.00%	0.10%	0.10%	0.10%	0.00%	0.00%	6.90%	18.50%
APOS2	0.00%	0.10%	10.40%	8.30%	4.00%	1.00%	0.30%	0.20%	0.40%
BPOS2	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.30%	0.10%	0.40%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.80%	0.80%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.10%	2.50%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.40%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	28.40%	0.10%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

State	DISP2	APOS2	BPOS2	BUBO3	THIN3	DISP3	APOS3	BPOS3	BBEH3
BUBO1	0.00%	1.30%	2.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	0.30%	1.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.30%	0.30%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.40%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	11.30%	5.90%	30.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	11.10%	3.60%	14.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	2.80%	43.80%	42.10%	0.00%	0.00%	0.10%	0.10%	0.10%	0.00%
BUBO2	1.90%	38.70%	42.60%	0.10%	0.10%	0.00%	0.30%	0.70%	0.00%
THIN2	1.90%	39.00%	40.60%	0.10%	0.20%	0.20%	0.50%	1.00%	0.00%
DISP2	3.60%	36.30%	33.80%	0.00%	0.00%	0.10%	0.20%	0.30%	0.00%
APOS2	17.50%	52.00%	5.50%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	16.80%	6.40%	54.20%	0.00%	0.00%	9.90%	3.60%	7.30%	0.90%
BUBO3	0.00%	6.20%	3.10%	4.60%	3.10%	2.30%	46.90%	29.20%	3.10%
THIN3	0.40%	2.50%	2.90%	9.10%	4.40%	2.20%	36.70%	37.80%	0.40%
DISP3	0.60%	0.50%	0.50%	6.10%	15.90%	5.20%	33.50%	37.60%	0.00%
APOS3	13.80%	34.30%	5.00%	0.10%	0.50%	5.00%	37.90%	2.50%	0.30%
BPOS3	0.80%	0.30%	0.00%	0.00%	0.30%	14.70%	8.30%	24.80%	22.20%
BBEH3	13.70%	20.10%	6.60%	0.40%	0.00%	1.30%	55.00%	3.00%	0.00%

Table A4-3: Carlton Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	13.40%	4.80%	2.10%	34.80%	39.60%	0.00%	0.00%	0.50%	0.50%
THIN1	15.00%	4.10%	2.60%	32.60%	42.00%	0.00%	0.00%	0.00%	1.60%
DISP1	8.40%	12.00%	4.00%	38.70%	36.10%	0.00%	0.00%	0.00%	0.00%
APOS1	0.20%	0.10%	15.00%	24.80%	8.90%	20.80%	28.90%	0.00%	0.00%
BPOS1	0.10%	0.10%	4.50%	2.20%	42.40%	0.30%	0.00%	0.00%	0.40%
ABEH1	0.90%	0.00%	1.70%	2.10%	60.30%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.00%	14.40%	0.00%
BUBO2	0.40%	0.00%	0.00%	0.50%	0.30%	0.00%	0.00%	12.40%	2.60%
THIN2	0.40%	0.10%	0.00%	0.50%	0.70%	0.00%	0.00%	9.70%	4.70%
DISP2	0.00%	0.00%	0.10%	0.10%	0.10%	0.00%	0.00%	8.40%	16.60%
APOS2	0.00%	0.10%	11.60%	7.70%	4.30%	1.10%	0.40%	0.20%	0.30%
BPOS2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.40%	0.10%	0.30%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.60%	0.50%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.90%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	30.00%	0.00%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

State	DISP2	APOS2	BPOS2	BUBO3	THIN3	DISP3	APOS3	BPOS3	BBEH3
BUBO1	0.00%	2.70%	1.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	0.50%	1.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.40%	0.10%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.80%	0.10%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	9.80%	4.50%	35.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	12.00%	4.90%	18.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	2.60%	39.90%	42.80%	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%
BUBO2	2.20%	38.50%	41.60%	0.30%	0.00%	0.00%	0.50%	0.70%	0.00%
THIN2	2.00%	40.00%	39.50%	0.50%	0.10%	0.40%	0.40%	1.00%	0.00%
DISP2	2.50%	35.40%	36.60%	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%
APOS2	19.20%	47.40%	7.40%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	15.70%	6.10%	53.80%	0.00%	0.00%	10.50%	3.70%	8.40%	0.80%
BUBO3	0.00%	3.80%	1.10%	11.40%	3.80%	1.60%	42.90%	31.00%	2.20%
THIN3	0.00%	1.10%	0.40%	10.60%	4.20%	1.50%	40.30%	39.90%	0.00%
DISP3	0.30%	0.10%	0.20%	7.60%	14.50%	3.70%	34.10%	39.60%	0.00%
APOS3	14.90%	31.30%	6.00%	0.20%	0.60%	5.70%	38.50%	2.20%	0.20%
BPOS3	0.30%	0.40%	0.20%	0.20%	0.30%	12.60%	9.00%	27.40%	19.70%
BBEH3	11.60%	12.30%	3.40%	0.20%	0.00%	1.50%	69.10%	1.90%	0.00%

Table A4-4: Collingwood Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	6.80%	3.80%	2.30%	34.10%	43.20%	3.80%	0.00%	1.50%	1.50%
THIN1	7.70%	4.00%	0.80%	38.70%	41.50%	1.20%	0.00%	0.40%	0.40%
DISP1	7.10%	15.90%	4.80%	35.40%	35.90%	0.00%	0.00%	0.00%	0.00%
APOS1	0.10%	0.30%	13.30%	25.10%	7.60%	24.40%	28.50%	0.10%	0.10%
BPOS1	0.10%	0.30%	3.90%	1.80%	43.80%	0.30%	0.00%	0.00%	0.50%
ABEH1	0.40%	0.20%	2.40%	1.70%	63.50%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	12.40%	0.10%
BUBO2	0.30%	0.20%	0.00%	0.30%	0.30%	0.00%	0.00%	11.40%	2.00%
THIN2	0.10%	0.10%	0.10%	1.10%	0.80%	0.00%	0.00%	9.10%	3.70%
DISP2	0.00%	0.00%	0.10%	0.10%	0.30%	0.00%	0.00%	6.80%	17.00%
APOS2	0.00%	0.10%	9.90%	7.10%	3.80%	0.90%	0.30%	0.10%	0.30%
BPOS2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.40%	0.00%	0.30%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.30%	0.00%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%	1.20%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	29.70%	0.00%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.00%	2.30%	0.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	1.60%	3.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.40%	0.30%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.50%	0.00%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	11.90%	4.60%	32.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	13.20%	5.10%	13.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	3.70%	41.90%	41.90%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BUBO2	2.80%	39.20%	42.60%	0.00%	0.00%	0.00%	0.30%	0.70%	0.00%
THIN2	2.70%	38.30%	42.30%	0.10%	0.20%	0.00%	0.50%	0.70%	0.00%
DISP2	4.00%	35.30%	36.00%	0.00%	0.00%	0.10%	0.20%	0.10%	0.00%
APOS2	17.50%	53.40%	6.30%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	16.30%	6.00%	53.40%	0.00%	0.00%	11.00%	3.90%	7.60%	0.90%
BUBO3	0.00%	2.00%	3.30%	7.80%	3.90%	4.60%	38.60%	37.90%	0.70%
THIN3	0.00%	0.80%	1.60%	10.40%	4.40%	3.60%	30.50%	46.20%	0.80%
DISP3	0.30%	0.30%	0.40%	6.80%	13.70%	4.40%	35.00%	39.10%	0.00%
APOS3	12.60%	33.20%	5.70%	0.10%	0.50%	4.70%	40.10%	2.50%	0.20%
BPOS3	0.70%	0.30%	0.00%	0.30%	0.20%	14.00%	7.70%	25.40%	21.60%
BBEH3	9.00%	11.80%	3.30%	0.20%	0.00%	1.70%	72.70%	1.30%	0.00%

Table A4-5: Essendon Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	11.20%	5.10%	2.80%	28.70%	44.90%	1.70%	0.00%	0.60%	1.10%
THIN1	10.60%	5.30%	2.70%	39.00%	39.40%	0.00%	0.00%	0.40%	0.80%
DISP1	7.10%	14.30%	5.30%	37.90%	34.50%	0.00%	0.00%	0.00%	0.00%
APOS1	0.40%	0.30%	15.70%	23.80%	7.10%	20.70%	31.40%	0.00%	0.00%
BPOS1	0.00%	0.30%	5.10%	2.00%	40.00%	0.20%	0.00%	0.00%	0.30%
ABEH1	0.00%	0.20%	1.40%	1.40%	64.60%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	13.90%	0.00%
BUBO2	0.10%	0.00%	0.10%	0.40%	0.40%	0.00%	0.00%	12.60%	2.40%
THIN2	0.90%	0.10%	0.10%	0.60%	0.20%	0.00%	0.00%	10.20%	2.90%
DISP2	0.00%	0.00%	0.10%	0.10%	0.20%	0.00%	0.00%	6.80%	15.70%
APOS2	0.00%	0.00%	11.10%	7.80%	3.80%	0.90%	0.40%	0.10%	0.30%
BPOS2	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.30%	0.10%	0.30%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.60%	0.00%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.40%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	28.90%	0.10%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.00%	1.10%	2.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	1.10%	0.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.30%	0.20%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.30%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	11.40%	5.20%	35.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	10.90%	2.60%	19.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	3.50%	39.80%	42.50%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%
BUBO2	4.60%	41.10%	36.80%	0.40%	0.00%	0.00%	0.60%	0.40%	0.00%
THIN2	3.60%	38.40%	41.60%	0.10%	0.00%	0.10%	0.70%	0.40%	0.00%
DISP2	4.00%	35.70%	36.70%	0.00%	0.00%	0.10%	0.20%	0.30%	0.00%
APOS2	17.60%	50.90%	6.90%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	16.70%	5.90%	52.10%	0.00%	0.00%	11.10%	4.50%	7.70%	1.00%
BUBO3	0.60%	1.20%	0.60%	9.30%	2.90%	2.30%	39.00%	42.40%	1.20%
THIN3	0.00%	2.70%	1.10%	7.60%	3.40%	1.90%	37.50%	45.10%	0.40%
DISP3	0.20%	0.40%	0.40%	7.20%	13.90%	4.80%	34.90%	38.30%	0.00%
APOS3	13.30%	31.90%	4.60%	0.10%	0.30%	4.90%	41.60%	2.40%	0.30%
BPOS3	0.50%	0.30%	0.00%	0.20%	0.20%	15.40%	8.60%	25.60%	20.20%
BBEH3	13.00%	20.90%	5.90%	0.50%	0.00%	1.60%	56.80%	1.30%	0.00%

Table A4-6: Fremantle Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	6.40%	2.70%	3.60%	33.60%	46.40%	4.50%	0.00%	0.90%	0.90%
THIN1	10.00%	3.80%	0.90%	40.30%	39.80%	0.00%	0.00%	0.90%	0.50%
DISP1	5.10%	13.70%	5.20%	37.50%	37.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.30%	0.10%	14.40%	24.50%	9.20%	20.70%	30.00%	0.00%	0.00%
BPOS1	0.00%	0.40%	4.80%	2.00%	42.10%	0.40%	0.00%	0.00%	0.40%
ABEH1	0.50%	0.20%	2.70%	0.50%	67.30%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	11.40%	0.00%
BUBO2	0.50%	0.00%	0.20%	1.60%	0.50%	0.00%	0.00%	12.00%	3.60%
THIN2	0.00%	0.10%	0.00%	1.40%	0.50%	0.00%	0.00%	8.30%	3.70%
DISP2	0.00%	0.00%	0.10%	0.20%	0.20%	0.00%	0.00%	7.40%	16.90%
APOS2	0.00%	0.00%	9.70%	7.20%	3.70%	0.70%	0.30%	0.10%	0.30%
BPOS2	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.30%	0.10%	0.30%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.20%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.20%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	26.40%	0.00%	0.10%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.00%	0.90%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	3.30%	0.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.30%	0.60%	0.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.40%	0.10%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	12.00%	5.70%	32.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	12.20%	5.00%	11.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	3.30%	41.90%	43.20%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%
BUBO2	3.30%	39.30%	37.90%	0.00%	0.00%	0.00%	0.60%	0.50%	0.00%
THIN2	3.50%	38.20%	41.80%	0.60%	0.10%	0.20%	0.90%	0.60%	0.00%
DISP2	4.60%	34.20%	35.80%	0.00%	0.00%	0.20%	0.20%	0.10%	0.00%
APOS2	17.50%	53.60%	6.70%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	17.40%	6.10%	53.10%	0.00%	0.00%	11.20%	3.60%	6.70%	1.10%
BUBO3	0.00%	1.80%	1.80%	11.20%	4.10%	3.60%	41.40%	33.70%	1.20%
THIN3	0.00%	0.40%	1.50%	8.00%	5.70%	3.80%	40.50%	39.40%	0.40%
DISP3	0.60%	0.30%	0.40%	7.20%	14.40%	6.30%	34.80%	36.00%	0.00%
APOS3	11.90%	30.30%	5.20%	0.00%	0.20%	4.10%	45.60%	1.90%	0.50%
BPOS3	0.70%	0.40%	0.10%	0.20%	0.10%	14.70%	8.30%	26.70%	22.50%
BBEH3	17.30%	12.40%	4.90%	0.90%	0.20%	4.20%	56.30%	3.80%	0.00%

Table A4-7: Geelong Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	12.80%	3.70%	1.20%	31.80%	43.40%	1.70%	0.00%	0.40%	0.80%
THIN1	15.60%	7.00%	3.20%	29.20%	37.80%	0.60%	0.00%	1.00%	1.60%
DISP1	8.20%	14.20%	5.60%	36.80%	33.80%	0.00%	0.00%	0.00%	0.00%
APOS1	0.10%	0.60%	14.20%	27.10%	8.00%	21.70%	27.30%	0.00%	0.00%
BPOS1	0.10%	0.40%	5.40%	2.60%	40.90%	0.50%	0.00%	0.10%	0.60%
ABEH1	0.30%	0.00%	3.60%	1.30%	61.50%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	13.90%	0.00%
BUBO2	0.50%	0.10%	0.00%	0.30%	1.60%	0.00%	0.00%	13.30%	3.20%
THIN2	0.50%	0.40%	0.30%	0.90%	0.90%	0.00%	0.00%	11.50%	3.90%
DISP2	0.00%	0.00%	0.20%	0.10%	0.10%	0.00%	0.00%	8.70%	16.30%
APOS2	0.00%	0.00%	10.70%	7.30%	3.70%	1.00%	0.20%	0.20%	0.30%
BPOS2	0.00%	0.00%	0.20%	0.10%	0.00%	0.00%	0.30%	0.20%	0.30%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.10%	1.10%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.30%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	28.30%	0.10%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

State	DISP2	APOS2	BPOS2	BUBO3	THIN3	DISP3	APOS3	BPOS3	BBEH3
BUBO1	0.00%	2.10%	2.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	1.90%	2.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.30%	0.60%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.70%	0.00%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	13.70%	6.20%	29.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	14.00%	5.70%	13.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	2.60%	41.10%	41.90%	0.00%	0.00%	0.10%	0.20%	0.10%	0.00%
BUBO2	3.30%	36.30%	40.00%	0.20%	0.00%	0.00%	0.80%	0.50%	0.00%
THIN2	3.40%	36.40%	39.80%	0.30%	0.00%	0.20%	0.40%	0.90%	0.00%
DISP2	3.40%	36.10%	34.60%	0.00%	0.00%	0.10%	0.10%	0.20%	0.00%
APOS2	17.30%	53.00%	6.10%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%
BPOS2	19.30%	6.90%	50.90%	0.00%	0.00%	10.10%	4.00%	6.80%	0.80%
BUBO3	0.00%	1.60%	1.10%	15.50%	3.70%	3.20%	29.90%	41.70%	1.10%
THIN3	0.00%	3.10%	0.50%	10.20%	2.00%	3.10%	39.80%	40.30%	0.50%
DISP3	0.30%	0.50%	0.50%	7.70%	11.30%	4.50%	38.70%	36.50%	0.00%
APOS3	11.00%	32.50%	4.20%	0.10%	0.10%	4.60%	45.00%	1.80%	0.20%
BPOS3	0.50%	0.50%	0.10%	0.40%	0.30%	16.10%	8.20%	25.00%	20.60%
BBEH3	11.70%	21.60%	5.60%	0.40%	0.00%	2.70%	56.20%	1.90%	0.00%

Table A4-8: Hawthorn Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	10.10%	2.80%	1.70%	34.80%	36.50%	2.20%	0.00%	0.60%	1.10%
THIN1	10.20%	1.10%	3.80%	31.70%	45.70%	1.60%	0.00%	0.50%	1.10%
DISP1	9.90%	12.50%	4.50%	34.40%	36.90%	0.00%	0.00%	0.00%	0.00%
APOS1	0.10%	0.20%	15.00%	27.20%	7.20%	20.10%	28.80%	0.10%	0.00%
BPOS1	0.00%	0.30%	5.40%	2.40%	41.00%	0.20%	0.00%	0.10%	0.50%
ABEH1	0.50%	0.00%	1.20%	1.50%	59.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	14.60%	0.00%
BUBO2	0.30%	0.00%	0.00%	0.40%	0.40%	0.00%	0.00%	14.70%	2.30%
THIN2	0.00%	0.10%	0.10%	0.50%	0.40%	0.00%	0.00%	13.10%	4.50%
DISP2	0.00%	0.00%	0.10%	0.10%	0.10%	0.00%	0.00%	7.60%	17.50%
APOS2	0.00%	0.00%	9.90%	6.70%	4.10%	0.60%	0.30%	0.10%	0.40%
BPOS2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.30%	0.10%	0.30%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%	1.00%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%	1.10%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	28.60%	0.10%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	1.10%	3.40%	5.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	2.20%	2.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.30%	0.80%	0.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	1.00%	0.00%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	11.00%	3.70%	35.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	11.50%	2.70%	23.30%	0.00%	0.00%	0.20%	0.00%	0.00%	0.00%
CEBO2	2.90%	41.10%	41.40%	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%
BUBO2	2.90%	37.20%	39.80%	0.70%	0.10%	0.10%	0.70%	0.40%	0.00%
THIN2	1.70%	37.60%	39.60%	0.50%	0.20%	0.00%	0.90%	0.90%	0.00%
DISP2	3.10%	33.50%	37.70%	0.00%	0.00%	0.10%	0.20%	0.00%	0.00%
APOS2	16.70%	54.60%	6.40%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	16.20%	6.20%	52.30%	0.00%	0.00%	10.80%	3.80%	8.70%	1.00%
BUBO3	0.00%	1.00%	0.50%	11.90%	3.00%	2.50%	38.30%	37.30%	4.00%
THIN3	0.00%	1.10%	1.80%	9.90%	4.40%	4.00%	39.90%	36.60%	0.70%
DISP3	0.20%	0.10%	0.30%	8.00%	15.20%	3.80%	34.20%	38.10%	0.00%
APOS3	10.80%	30.60%	5.80%	0.10%	0.50%	5.00%	44.30%	2.20%	0.20%
BPOS3	0.70%	0.10%	0.10%	0.50%	0.00%	13.50%	8.20%	28.00%	20.30%
BBEH3	10.50%	12.30%	3.30%	0.40%	0.00%	4.20%	68.40%	0.90%	0.00%

Table A4-9: Melbourne Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	8.50%	6.30%	3.40%	29.50%	41.50%	4.00%	0.00%	1.10%	0.60%
THIN1	8.40%	6.90%	3.60%	37.20%	38.30%	0.70%	0.00%	0.40%	1.10%
DISP1	7.40%	14.30%	4.80%	38.40%	34.50%	0.00%	0.00%	0.00%	0.00%
APOS1	0.30%	0.10%	15.50%	22.60%	8.40%	23.00%	29.00%	0.00%	0.00%
BPOS1	0.10%	0.20%	4.70%	2.30%	42.00%	0.20%	0.00%	0.00%	0.20%
ABEH1	0.70%	0.00%	1.40%	1.40%	66.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	10.80%	0.10%
BUBO2	0.00%	0.00%	0.20%	0.80%	1.00%	0.00%	0.00%	9.80%	2.70%
THIN2	0.10%	0.10%	0.20%	1.00%	0.40%	0.00%	0.00%	9.10%	4.90%
DISP2	0.00%	0.00%	0.10%	0.20%	0.20%	0.00%	0.00%	7.30%	15.60%
APOS2	0.00%	0.00%	12.00%	8.50%	4.00%	0.90%	0.30%	0.10%	0.40%
BPOS2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.30%	0.10%	0.30%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.60%	0.60%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.80%	0.40%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	28.40%	0.00%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.00%	1.10%	4.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	1.50%	1.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.10%	0.40%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.90%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	11.80%	4.20%	34.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	10.60%	4.70%	15.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	3.50%	39.80%	45.50%	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%
BUBO2	2.10%	35.70%	45.40%	0.50%	0.20%	0.00%	1.40%	0.30%	0.00%
THIN2	4.20%	35.70%	42.60%	0.40%	0.10%	0.00%	0.50%	0.60%	0.00%
DISP2	4.10%	36.40%	35.60%	0.00%	0.00%	0.10%	0.20%	0.10%	0.00%
APOS2	17.30%	49.90%	6.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS2	17.30%	6.40%	52.30%	0.00%	0.00%	10.50%	3.90%	7.60%	1.00%
BUBO3	1.10%	1.70%	2.30%	9.10%	2.30%	2.80%	47.20%	31.30%	1.10%
THIN3	0.40%	2.40%	1.60%	12.40%	5.20%	2.00%	35.60%	38.80%	0.40%
DISP3	0.30%	0.70%	0.20%	7.70%	14.90%	4.00%	34.20%	38.10%	0.00%
APOS3	11.60%	34.40%	4.90%	0.10%	0.10%	4.70%	41.30%	2.10%	0.30%
BPOS3	0.80%	0.30%	0.00%	0.10%	0.10%	14.20%	8.20%	26.50%	21.40%
BBEH3	11.70%	13.20%	5.70%	0.20%	0.00%	2.00%	65.40%	1.80%	0.00%

Table A4-10: North Melbourne Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	6.20%	2.30%	2.30%	38.00%	41.90%	3.90%	0.00%	0.00%	0.80%
THIN1	7.60%	3.80%	3.40%	37.10%	42.20%	0.40%	0.00%	0.80%	0.80%
DISP1	6.70%	15.80%	3.40%	39.20%	34.10%	0.00%	0.00%	0.00%	0.00%
APOS1	0.30%	0.30%	13.70%	26.60%	8.00%	20.20%	29.60%	0.00%	0.00%
BPOS1	0.10%	0.20%	4.40%	2.40%	43.90%	0.10%	0.00%	0.00%	0.20%
ABEH1	0.60%	0.00%	1.60%	0.40%	69.20%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	12.70%	0.00%
BUBO2	0.00%	0.20%	0.00%	0.20%	0.80%	0.00%	0.00%	11.00%	4.40%
THIN2	0.00%	0.10%	0.20%	0.90%	0.20%	0.00%	0.00%	10.20%	4.90%
DISP2	0.00%	0.00%	0.20%	0.30%	0.30%	0.00%	0.00%	6.90%	18.60%
APOS2	0.00%	0.00%	9.80%	8.00%	3.90%	1.10%	0.40%	0.10%	0.30%
BPOS2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.30%	0.20%	0.20%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.40%	1.40%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%	0.80%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	29.10%	0.00%	0.10%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.00%	1.60%	3.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	2.50%	1.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.10%	0.20%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	1.00%	0.10%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	11.40%	5.00%	32.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	12.00%	4.40%	11.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	3.40%	42.70%	40.90%	0.00%	0.00%	0.10%	0.00%	0.20%	0.00%
BUBO2	3.40%	40.20%	38.80%	0.30%	0.00%	0.00%	0.50%	0.50%	0.00%
THIN2	2.50%	36.70%	41.90%	0.10%	0.10%	0.00%	0.90%	1.20%	0.00%
DISP2	3.70%	34.60%	35.00%	0.00%	0.00%	0.10%	0.10%	0.20%	0.00%
APOS2	17.60%	52.00%	6.60%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	15.50%	6.00%	55.00%	0.00%	0.00%	9.60%	3.90%	8.20%	0.70%
BUBO3	0.70%	3.40%	1.40%	10.90%	4.10%	3.40%	42.20%	27.20%	2.00%
THIN3	0.00%	3.60%	1.20%	11.60%	5.60%	1.60%	39.40%	35.30%	0.40%
DISP3	0.50%	0.30%	0.40%	6.60%	15.90%	3.50%	34.50%	38.30%	0.00%
APOS3	12.70%	34.50%	5.50%	0.10%	0.10%	4.20%	40.30%	1.90%	0.40%
BPOS3	0.70%	0.10%	0.00%	0.20%	0.50%	13.30%	8.20%	27.40%	20.30%
BBEH3	10.30%	14.30%	4.40%	0.20%	0.00%	2.00%	67.70%	1.00%	0.00%

Table A4-11: Port Adelaide Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	14.60%	2.00%	2.50%	34.80%	40.40%	2.00%	0.00%	0.50%	1.00%
THIN1	11.40%	3.80%	1.30%	43.00%	36.30%	0.80%	0.00%	0.40%	0.00%
DISP1	8.20%	13.30%	4.90%	39.80%	32.60%	0.00%	0.00%	0.00%	0.00%
APOS1	0.10%	0.20%	13.00%	28.10%	8.20%	19.60%	29.80%	0.00%	0.00%
BPOS1	0.10%	0.30%	4.60%	1.90%	41.90%	0.30%	0.00%	0.00%	0.30%
ABEH1	0.20%	0.00%	1.80%	2.50%	65.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.20%	0.10%	0.00%	0.00%	15.00%	0.10%
BUBO2	0.30%	0.00%	0.00%	0.70%	0.60%	0.00%	0.00%	10.30%	1.70%
THIN2	0.50%	0.10%	0.20%	0.80%	0.10%	0.00%	0.00%	11.50%	3.80%
DISP2	0.00%	0.00%	0.00%	0.20%	0.20%	0.00%	0.00%	7.40%	17.70%
APOS2	0.00%	0.00%	9.90%	7.80%	3.10%	1.00%	0.30%	0.10%	0.20%
BPOS2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.30%	0.10%	0.30%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.60%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%	0.00%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	27.20%	0.00%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.00%	1.50%	0.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.40%	1.30%	1.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.40%	0.40%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.60%	0.10%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	11.70%	4.80%	34.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	10.20%	4.90%	15.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	3.80%	43.70%	36.90%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%
BUBO2	2.90%	41.80%	40.00%	0.10%	0.00%	0.30%	0.30%	1.10%	0.00%
THIN2	3.30%	41.00%	37.50%	0.00%	0.00%	0.00%	0.20%	0.90%	0.00%
DISP2	3.70%	35.40%	35.00%	0.00%	0.00%	0.00%	0.20%	0.10%	0.00%
APOS2	15.40%	56.00%	6.00%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	15.50%	5.70%	55.30%	0.00%	0.00%	9.80%	3.80%	8.10%	1.00%
BUBO3	0.00%	2.90%	0.60%	11.00%	4.70%	4.10%	39.00%	33.70%	3.50%
THIN3	0.00%	1.20%	1.60%	9.80%	4.10%	0.80%	42.70%	39.00%	0.40%
DISP3	0.20%	0.50%	0.50%	8.10%	14.50%	4.60%	34.90%	36.70%	0.00%
APOS3	9.70%	32.40%	4.70%	0.00%	0.20%	4.50%	45.70%	2.00%	0.20%
BPOS3	0.80%	0.20%	0.00%	0.10%	0.20%	13.80%	9.30%	25.60%	22.60%
BBEH3	14.00%	17.40%	5.80%	0.50%	0.00%	1.70%	59.20%	1.40%	0.00%

Table A4-12: Richmond Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	12.00%	2.90%	1.70%	37.70%	41.70%	0.00%	0.00%	0.00%	1.10%
THIN1	12.90%	4.40%	2.20%	39.10%	38.20%	0.40%	0.00%	0.40%	0.40%
DISP1	7.80%	14.20%	3.10%	37.90%	35.80%	0.00%	0.00%	0.00%	0.00%
APOS1	0.20%	0.20%	14.20%	27.70%	8.10%	20.40%	27.60%	0.10%	0.00%
BPOS1	0.00%	0.20%	4.20%	1.20%	42.90%	0.20%	0.00%	0.10%	0.50%
ABEH1	0.20%	0.00%	2.40%	1.10%	64.50%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%	0.00%	15.40%	0.10%
BUBO2	0.40%	0.00%	0.00%	0.50%	0.90%	0.00%	0.00%	14.20%	2.30%
THIN2	0.40%	0.00%	0.10%	0.80%	0.50%	0.00%	0.00%	12.90%	2.90%
DISP2	0.00%	0.00%	0.00%	0.10%	0.20%	0.00%	0.00%	8.80%	16.60%
APOS2	0.00%	0.00%	11.00%	6.70%	3.80%	0.90%	0.20%	0.10%	0.20%
BPOS2	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.40%	0.10%	0.30%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.60%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.40%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	31.50%	0.00%	0.00%
BBEH3	0.00%	0.00%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.00%	0.00%	2.90%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	1.30%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.50%	0.50%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	1.00%	0.10%	0.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	10.20%	4.80%	35.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	13.50%	3.90%	14.60%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	2.30%	39.90%	42.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BUBO2	1.70%	40.80%	36.70%	0.10%	0.10%	0.00%	1.30%	0.90%	0.00%
THIN2	2.30%	44.60%	33.20%	0.40%	0.10%	0.10%	0.80%	0.70%	0.00%
DISP2	2.90%	36.00%	34.80%	0.00%	0.00%	0.20%	0.20%	0.10%	0.00%
APOS2	16.30%	54.00%	6.40%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	15.00%	5.50%	55.10%	0.00%	0.10%	10.00%	3.80%	8.70%	0.90%
BUBO3	0.00%	0.60%	0.60%	11.50%	4.20%	2.40%	40.00%	38.20%	1.80%
THIN3	0.00%	1.40%	1.40%	7.90%	3.30%	2.30%	52.80%	30.80%	0.00%
DISP3	0.20%	0.40%	0.10%	8.20%	12.80%	3.60%	36.60%	38.00%	0.00%
APOS3	10.00%	31.90%	5.20%	0.00%	0.20%	4.50%	45.60%	1.80%	0.30%
BPOS3	0.50%	0.10%	0.00%	0.10%	0.10%	14.90%	8.00%	24.00%	20.70%
BBEH3	11.40%	13.80%	4.30%	0.20%	0.00%	2.30%	66.30%	1.60%	0.00%

Table A4-13: St. Kilda Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	7.80%	1.60%	2.10%	35.90%	40.60%	2.10%	0.00%	1.00%	1.00%
THIN1	10.50%	4.40%	1.30%	36.70%	41.50%	0.90%	0.00%	0.90%	0.40%
DISP1	8.80%	12.50%	2.80%	39.10%	35.70%	0.00%	0.00%	0.00%	0.00%
APOS1	0.30%	0.30%	12.90%	25.00%	8.90%	19.50%	32.10%	0.00%	0.00%
BPOS1	0.10%	0.20%	4.40%	2.50%	44.10%	0.30%	0.00%	0.00%	0.40%
ABEH1	0.20%	0.20%	2.40%	2.40%	58.80%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	16.50%	0.00%
BUBO2	0.10%	0.10%	0.00%	0.30%	0.90%	0.00%	0.00%	15.40%	2.40%
THIN2	0.00%	0.10%	0.00%	0.80%	0.40%	0.00%	0.00%	14.30%	4.70%
DISP2	0.00%	0.00%	0.10%	0.20%	0.10%	0.00%	0.00%	10.30%	15.80%
APOS2	0.00%	0.00%	10.50%	8.20%	4.00%	0.70%	0.30%	0.20%	0.20%
BPOS2	0.00%	0.00%	0.20%	0.10%	0.00%	0.00%	0.30%	0.10%	0.40%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.60%	0.00%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.00%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.20%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	31.50%	0.10%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.50%	3.10%	4.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	1.30%	2.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.40%	0.20%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.40%	0.10%	0.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	12.70%	5.10%	30.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	15.20%	6.80%	14.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	2.20%	41.40%	39.70%	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%
BUBO2	1.80%	42.10%	35.40%	0.10%	0.00%	0.00%	1.00%	0.40%	0.00%
THIN2	1.90%	39.20%	36.70%	0.20%	0.10%	0.00%	0.60%	0.90%	0.00%
DISP2	2.90%	36.20%	34.00%	0.00%	0.00%	0.10%	0.20%	0.20%	0.00%
APOS2	15.90%	54.00%	5.80%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	18.60%	6.70%	51.90%	0.00%	0.00%	10.40%	3.70%	6.70%	0.90%
BUBO3	0.00%	4.20%	1.80%	13.20%	3.00%	0.60%	43.10%	33.50%	0.00%
THIN3	0.00%	2.50%	3.00%	7.50%	3.00%	0.00%	46.70%	36.20%	0.00%
DISP3	0.20%	0.10%	0.40%	8.70%	11.90%	3.90%	36.90%	37.90%	0.00%
APOS3	11.20%	33.70%	5.10%	0.00%	0.20%	4.20%	43.30%	1.90%	0.30%
BPOS3	0.90%	0.30%	0.10%	0.10%	0.40%	14.00%	7.70%	25.60%	19.50%
BBEH3	10.30%	18.10%	3.30%	0.00%	0.20%	1.10%	65.80%	1.10%	0.00%

Table A4-14: Western Bulldogs Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	6.30%	4.90%	4.20%	34.70%	41.00%	4.90%	0.00%	0.70%	0.00%
THIN1	8.90%	5.50%	2.10%	39.20%	38.40%	0.80%	0.00%	0.00%	1.70%
DISP1	7.30%	14.30%	4.60%	37.40%	35.60%	0.00%	0.00%	0.00%	0.00%
APOS1	0.10%	0.10%	14.50%	25.70%	8.50%	21.70%	28.20%	0.10%	0.00%
BPOS1	0.10%	0.30%	4.20%	1.40%	43.00%	0.30%	0.00%	0.00%	0.40%
ABEH1	0.20%	0.00%	1.50%	1.10%	61.30%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.00%	10.80%	0.00%
BUBO2	0.20%	0.20%	0.00%	0.90%	0.70%	0.00%	0.00%	11.60%	2.10%
THIN2	0.40%	0.10%	0.10%	1.40%	0.30%	0.00%	0.00%	10.60%	3.60%
DISP2	0.00%	0.00%	0.00%	0.20%	0.20%	0.00%	0.00%	6.90%	16.70%
APOS2	0.00%	0.00%	9.70%	8.20%	4.00%	0.90%	0.60%	0.10%	0.30%
BPOS2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.30%	0.10%	0.20%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.00%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.90%	0.00%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	29.00%	0.00%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	DISP2	APOS2	BPOS2	BUBO3	THIN3	DISP3	APOS3	BPOS3	BBEH3
BUBO1	0.00%	0.70%	2.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	1.30%	2.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.40%	0.00%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.60%	0.10%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	11.10%	4.40%	34.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	12.40%	5.50%	18.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	2.80%	42.20%	44.00%	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%
BUBO2	1.70%	38.50%	42.10%	0.50%	0.00%	0.00%	0.50%	1.00%	0.00%
THIN2	2.10%	39.40%	39.50%	0.40%	0.00%	0.10%	0.60%	1.40%	0.00%
DISP2	2.90%	35.30%	37.40%	0.00%	0.00%	0.10%	0.30%	0.10%	0.00%
APOS2	14.10%	56.30%	5.60%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	15.70%	5.40%	54.20%	0.00%	0.00%	10.50%	3.60%	8.80%	0.90%
BUBO3	0.00%	1.30%	1.30%	6.70%	2.70%	2.70%	43.30%	36.70%	3.30%
THIN3	0.00%	1.70%	0.40%	9.60%	3.50%	0.90%	45.40%	36.70%	0.90%
DISP3	0.10%	0.30%	0.30%	7.30%	14.10%	2.80%	33.60%	41.60%	0.00%
APOS3	9.90%	33.20%	4.80%	0.10%	0.20%	4.40%	44.30%	2.20%	0.40%
BPOS3	0.60%	0.30%	0.00%	0.10%	0.20%	12.70%	7.40%	28.00%	21.70%
BBEH3	8.70%	13.90%	2.80%	0.20%	0.00%	1.30%	71.60%	1.50%	0.00%

Table A4-15: West Coast Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	9.80%	4.30%	1.60%	33.70%	42.90%	3.80%	0.00%	0.50%	0.00%
THIN1	7.60%	2.90%	2.20%	37.70%	44.20%	0.00%	0.00%	0.00%	1.40%
DISP1	7.50%	14.90%	3.30%	38.60%	35.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.30%	0.20%	14.90%	24.50%	9.00%	22.70%	27.30%	0.10%	0.00%
BPOS1	0.20%	0.40%	4.80%	2.20%	43.50%	0.20%	0.00%	0.00%	0.40%
ABEH1	0.80%	0.00%	2.00%	3.00%	65.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	9.80%	0.00%
BUBO2	0.00%	0.00%	0.00%	0.50%	0.50%	0.00%	0.00%	8.00%	3.30%
THIN2	0.40%	0.10%	0.10%	1.00%	0.40%	0.00%	0.00%	11.40%	3.60%
DISP2	0.00%	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	6.60%	18.10%
APOS2	0.00%	0.00%	10.10%	7.40%	3.40%	0.90%	0.40%	0.10%	0.30%
BPOS2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.30%	0.10%	0.40%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.30%	1.30%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.00%	0.60%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.20%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	28.70%	0.00%	0.10%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.00%	0.00%	3.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	1.40%	2.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.20%	0.20%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.60%	0.00%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	12.70%	4.90%	30.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	12.30%	4.00%	12.90%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	3.30%	42.70%	44.10%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%
BUBO2	2.50%	43.30%	40.20%	0.20%	0.00%	0.00%	0.80%	0.80%	0.00%
THIN2	2.30%	42.50%	36.20%	0.10%	0.10%	0.30%	1.10%	0.40%	0.00%
DISP2	3.80%	37.20%	33.40%	0.00%	0.00%	0.20%	0.20%	0.20%	0.00%
APOS2	17.20%	54.40%	5.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS2	18.60%	6.60%	51.50%	0.00%	0.00%	10.30%	3.90%	7.30%	0.80%
BUBO3	0.00%	2.60%	2.60%	9.90%	2.60%	1.30%	37.10%	36.40%	4.60%
THIN3	0.00%	1.60%	1.60%	9.30%	6.10%	2.60%	39.00%	38.30%	0.00%
DISP3	0.20%	0.50%	0.10%	6.20%	18.10%	5.30%	35.50%	34.10%	0.00%
APOS3	12.20%	33.40%	4.70%	0.10%	0.30%	4.80%	42.00%	2.00%	0.30%
BPOS3	1.20%	0.40%	0.10%	0.30%	0.30%	15.00%	8.20%	25.90%	19.90%
BBEH3	12.90%	16.80%	7.50%	0.40%	0.20%	2.30%	58.60%	1.20%	0.00%

Table A4-16: Sydney Football Club

State	BUBO1	THIN1	DISP1	APOS1	BPOS1	ABEH1	CEBO2	BUBO2	THIN2
BUBO1	14.60%	5.00%	2.00%	32.20%	37.20%	2.00%	0.00%	2.00%	0.00%
THIN1	10.30%	3.00%	2.70%	40.90%	39.50%	0.00%	0.00%	0.70%	0.70%
DISP1	7.70%	17.30%	4.40%	36.80%	32.50%	0.00%	0.00%	0.00%	0.00%
APOS1	0.50%	0.50%	14.10%	29.70%	7.70%	19.40%	27.10%	0.00%	0.00%
BPOS1	0.20%	0.30%	5.00%	1.90%	40.80%	0.40%	0.00%	0.00%	0.40%
ABEH1	0.20%	0.00%	2.30%	1.70%	67.70%	0.00%	0.00%	0.00%	0.00%
CEBO2	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	18.50%	0.00%
BUBO2	0.30%	0.00%	0.00%	0.90%	0.70%	0.00%	0.00%	16.80%	3.70%
THIN2	0.20%	0.00%	0.20%	1.00%	1.10%	0.00%	0.00%	13.30%	5.80%
DISP2	0.00%	0.00%	0.20%	0.20%	0.30%	0.00%	0.00%	8.90%	20.40%
APOS2	0.00%	0.00%	10.30%	7.60%	3.70%	0.80%	0.30%	0.20%	0.50%
BPOS2	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.30%	0.20%	0.40%
BUBO3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.10%	0.70%
THIN3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.70%	1.60%
DISP3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
BPOS3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	27.60%	0.10%	0.00%
BBEH3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
State	**DISP2**	**APOS2**	**BPOS2**	**BUBO3**	**THIN3**	**DISP3**	**APOS3**	**BPOS3**	**BBEH3**
BUBO1	0.50%	1.50%	3.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
THIN1	0.00%	1.00%	1.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DISP1	0.50%	0.40%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
APOS1	0.70%	0.00%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BPOS1	15.60%	5.10%	30.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ABEH1	9.70%	3.30%	15.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBO2	4.40%	38.30%	38.60%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%
BUBO2	2.40%	38.20%	34.90%	0.30%	0.10%	0.10%	0.70%	0.80%	0.00%
THIN2	3.10%	37.60%	35.40%	0.20%	0.20%	0.10%	0.60%	1.50%	0.00%
DISP2	4.40%	31.50%	33.60%	0.00%	0.00%	0.10%	0.20%	0.10%	0.00%
APOS2	19.10%	51.00%	6.30%	0.00%	0.00%	0.10%	0.00%	0.10%	0.00%
BPOS2	21.10%	6.40%	48.20%	0.00%	0.00%	12.30%	3.30%	6.70%	0.90%
BUBO3	0.00%	2.80%	1.80%	10.90%	6.00%	1.80%	39.40%	33.50%	2.10%
THIN3	0.00%	1.00%	1.30%	18.00%	5.20%	3.30%	36.10%	31.80%	1.00%
DISP3	0.40%	0.40%	0.50%	11.10%	15.30%	4.40%	32.50%	35.50%	0.00%
APOS3	13.60%	31.20%	4.70%	0.10%	0.20%	5.00%	42.30%	2.10%	0.30%
BPOS3	0.80%	0.30%	0.10%	0.50%	0.60%	15.60%	8.80%	23.60%	22.10%
BBEH3	7.20%	12.00%	3.50%	0.20%	0.20%	1.40%	74.80%	0.80%	0.00%