Predicting sporting outcomes: a statistical approach

Michael J. Bailey

MSc(Statistics), BSc(Hons)

A thesis submitted for

the degree of

Doctor of Philosophy

Faculty of Life and Social Sciences

Swinburne University of Technology

March 2005

Acknowledgments

Prior to the commencement of this doctorate, I was fortunate enough to spend two years as a sports statistician for the Swinburne Sports Statistics department. It was during this period of time that much of the inspiration for this thesis was derived and I would like to thank both the Swinburne Department of Mathematical Sciences and more specifically the Swinburne Sports Statistics department for this opportunity.

I would like to thank Swinburne for offering a Swinburne University Post Graduate Research Award (SUPRA), and the Department of Education, Science and Training for ultimately providing an Australian Postgraduate Award (APA) to fund this body of work.

To my beautiful wife and daughter, Rochelle and Isabella, thank you greatly for your love, patience and understanding.

I would like to thank my father Bernard Bailey for his proof reading skill and astute comments.

I would like to thank my supervisors Julie Pallant and Stephen Clarke. In particular, it has been Steve's knowledge, humour, insight, persistence and tolerance that have made this dissertation possible.

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.

Michael J. Bailey

Preface

Three of the six analysis conducted in this dissertation have resulted in refereed publications. Of the four publications that have arisen from this work, one was published solely by the author; the other three as a result of the authors work, under the knowledgeable guidance of my supervisor Stephen Clarke.

The first analysis conducted, predicting the outcome of AFL matches, has resulted in two publications. The first of these publications was written solely by the author in the year 2000 and provides an early insight into development of the modeling process. (Bailey, M. (2000). Identifying arbitrage opportunities in AFL betting markets through mathematical modeling. *Proceedings of the Fifth Australian conference on Mathematics and Computers in Sport.* G. Cohen and T. Langtry. Sydney, University of Technology Sydney**:** 37-42).

The second publication relating to the prediction of AFL matches was written four years later under the supervision of Stephen Clarke and forms the nucleus of the analysis presented in chapter four of this dissertation (Bailey, M., J, and Clarke, S. R. (2004). Deriving a profit from Australian Rules football: A statistical approach. *Proceedings of the Seventh Australian conference on Mathematics and Computers in Sport*. Palmerston North, Massey University, R Hugh Morton & S Ganesalingam**:** 48-56). The conference presentation of this analysis resulted in an award for best student presentation.

In chapter seven of this thesis, the prediction of batsmen performance in limited overs cricket is presented. Under the guidance of Stephen Clarke, this analysis was published as a refereed book chapter in 2004. (Bailey, M. J. and S. R. Clarke (2004). Market Inefficiencies in player head to head betting on the 2003 cricket world cup. *Economics, Management and Optimization in sport*. S. Butenko, J. Gil-Lafuente and P. Pardalos. Heidelberg, Springer-Verlag: 185-201.)

In chapter eight, the prediction of the Brownlow medal winner in AFL football is discussed. Several media and academic articles have resulted from this body of work, the most notable of which being a refereed conference paper that was once again written under the guidance of my supervisor Stephen Clarke. This analysis has since been superceded, but provides much impetus for the results presented in this dissertation. (Bailey, M. J. & Clarke, S. R. (2002). Predicting the Brownlow medal winner. *Proceedings of the sixth Australian conference on mathematics and computers in sport*. G. Cohen and T. Langtry. Bond University, University of Technology, Sydney: 56-62.)

Table of contents

Acknowledgments .... **ii**

Candidate's statement .... **iii**

Preface .... **iv**

Table of contents .... **vi**

List of Figures .... **xi**

List of Tables .... **xiv**

Table of abbreviations .... **xv**

Abstract .... **xvi**

1 Introduction **1**

2 Background and literature review **4**

2.1 Australian Rules football *4*

2.2 Cricket *7*

2.3 Predicting sporting outcomes *10*

2.4 Home advantage *11*

2.5 Market efficiency *12*

3 Methodology **17**

3.1 Background *17*

3.2 Data *18*

3.3 Underlying distribution *20*

3.4 Identifying significant predictors *22*

3.5 Holdout sample *23*

3.6 Multivariate model building *24*

3.7 Accuracy of prediction *25*

3.8 Bookmaker prices *26*

3.9 Converting predicted outcomes into probabilities *27*

3.10 Wagering strategy *28*

3.11 Statistical analysis *29*

4 Predicting the match result in AFL football **30**

4.1 Introduction *30*

4.2 Background *30*

4.3 Predictors of MOV for AFL football *32*

4.3.1 HA in AFL football 32
4.3.2 Travel fatigue 33
4.3.3 Ground familiarisation 33
4.3.4 Measures of performance 34

4.4 Bookmaker prices *38*

4.5 Method *39*

4.6 Results *41*

4.6.1 Trends over time 43

4.7 Discussion *45*

5 Predicting the match result in ODI cricket **47**

5.1 Introduction *47*

5.2 Background *47*

5.3 Predictors of MOV for ODI cricket *52*

5.3.1 Home country advantage 52
5.3.2 Class structure 52
5.3.3 Distance travelled 53
5.3.4 Familiarization in host country 54
5.3.5 Familiarization at the venue 57
5.3.6 Experience 59
5.3.7 Experience against the opposition 60
5.3.8 Measures of performance 60
5.3.9 Specific measures of performance 62

5.4 Multivariate analysis *63*

5.5 Results *66*

5.6 Discussion *67*

6 Predicting individual player performance in AFL football **70**

6.1 Introduction *70*

6.2 Background *70*

6.3 Database *72*

6.4 Predictors of disposals *73*

6.4.1 Overview 73
6.4.2 Age 74
6.4.3 Experience 75

6.4.4 Physical attributes ... 76
6.4.5 Position ... 79
6.4.6 Fitness ... 82
6.4.7 Days rest ... 84
6.4.8 Recent form ... 85
6.4.9 Measures of quality ... 86
6.4.10 Within player variability ... 89
6.4.11 Key players missing ... 89
6.4.12 Within team ranking ... 90
6.4.13 Team ... 91
6.4.14 Opposition ... 92
6.4.15 Team and opposition experience ... 93
6.4.16 Night games ... 94
6.4.17 Weather conditions ... 94
6.4.18 Venue ... 95
6.4.19 Ground familiarisation ... 96
6.4.20 HA and interstate travel ... 97
6.4.21 Match result ... 98
6.4.22 Round ... 100
6.4.23 Year ... 101

6.5 Annual differences ... 102
6.5.1 Style of play ... 102
6.5.2 Rule changes ... 103
6.5.3 Data collection processes ... 103
6.5.4 Definitions ... 103

6.6 Training dataset ... 104

6.7 Linear regression ... 105

6.8 Converting predictions to probabilities ... 107
6.8.1 Independence ... 107

6.9 Market efficiency ... 109

6.10 Goodness of fit ... 111

6.11 Multiple linear regression ... 114

6.12 Results ... 115

6.13 Discussion ... 117

7 Predicting individual player performance in ODI cricket ... 119

7.1 Introduction ... 119

7.2 Background ... 120

7.3 Factors affecting batting performance ... 122
7.3.1 Batting position ... 122
7.3.2 Experience ... 124

7.3.3 Home country advantage 125
7.3.4 Match time and innings sequence 126
7.3.5 Opposition and team effects.. 126
7.3.6 Performance and form.. 129

7.4 Multivariate models *131*

7.5 Goodness of fit for predicted score.. *134*

7.6 Converting predicted scores into probabilities *135*
7.6.1 Comparison of conversion approaches 137

7.7 HtH prices for the 2003 World Cup *138*

7.8 Results.. *138*

7.9 Discussion... *142*

8 Predicting the Brownlow medal winner 145

8.1 Introduction *145*

8.2 Background.. *145*

8.3 Database.. *146*

8.4 Predictors for Brownlow voting *147*
8.4.1 Disposals 147
8.4.2 Won the game / MOV 147
8.4.3 Hit Outs 148
8.4.4 Standout 148
8.4.5 Best players 148
8.4.6 Quality of disposals.. 149
8.4.7 Goals 150
8.4.8 Marks 150
8.4.9 Rebounds / Frees for / Tackles / Inside 50... 151
8.4.10 Position 152
8.4.11 Distinct appearance 152
8.4.12 Team scoring shots 153
8.4.13 Captain 153
8.4.14 Past votes polled 153

8.5 Multivariate model *154*

8.6 Goodness of fit within the game *156*
8.6.1 Log likelihood... 156
8.6.2 Average Rank... 156
8.6.3 Top three for the game 157

8.7 Goodness of fit for season *158*
8.7.1 Average absolute error.. 158
8.7.2 Top 10 for season... 160

8.8 Measuring the bias of over-fitting the data *162*

8.9 Results ... *166*

8.10 Discussion ... *168*

9 Predicting the number of runs scored per over in ODI cricket matches ... 171

9.1 Introduction ... *171*

9.2 Background ... *171*

9.3 Predictive factors for runs scored per over ... *174*

9.3.1 Overview ... 174
9.3.2 Batting Sequence ... 174
9.3.3 Overs ... 175
9.3.4 Wickets ... 176
9.3.5 Wicket last over ... 178
9.3.6 Partnerships ... 178
9.3.7 Runs previously scored in the match ... 179
9.3.8 Best Bowler ... 182
9.3.9 Host Country ... 183
9.3.10 Team ... 184
9.3.11 Team by stage ... 184
9.3.12 Opposition by Stage ... 185

9.4 Multivariate linear models ... *186*

9.5 Alternative distributions ... *189*

9.5.1 Hold-out Sample ... 191

9.6 Results ... *191*

9.7 Discussion ... *193*

10 Conclusion ... 194

10.1 Summation ... *194*

10.2 Statistically driven prediction models ... *197*

10.3 Market efficiency ... *199*

11 References ... 203

12 Appendices ... 208

Appendix 1 Published multivariate analysis conducted by candidate ... *208*

List of Figures

Figure 4-1 Histogram of margin of victory (Home team score minus Away team score) 31
Figure 4-2 Home advantage and total game score stratified in five-year periods 32
Figure 4-3 Comparison of performance predictors for all AFL matches 35
Figure 4-4 Comparison of predictors stratified by venue and opposition 36
Figure 4-5 Average absolute error for exponentially smoothed predictor (x=.1)............ 38
Figure 4-6 AAE for the three models from 1997 to 2003 43
Figure 4-7 Winning Percentage from 1997 to 2003 44
Figure 4-8 Annual Return on Investment from 1997 to 2003 44
Figure 5-1 Histogram of MOV referenced against the home team 51
Figure 5-2 Team and individual predictors for quality using the AAE 61
Figure 5-3 AAE and percentage of winners for predictors of quality and form.............. 62
Figure 5-4 Prediction approaches used over different strata 63
Figure 6-1 Distribution of player disposals 71
Figure 6-2 Average disposals by age 75
Figure 6-3 Average disposals by experience 76
Figure 6-4 Average disposals by height.. 77
Figure 6-5 Average disposals by weight... 78
Figure 6-6 Average disposals by BMI 78
Figure 6-8 Average height & weight for each position 81
Figure 6-9 Average disposals for each position.. 82
Figure 6-10 Average disposals for weeks since last game 83
Figure 6-11 Average disposals for percentage of games played for the year.................. 84
Figure 6-12 Average disposals for days rest.. 85
Figure 6-13 Model R-square for predictors of form 86
Figure 6-14 Comparison for measures of performance 88
Figure 6-15 Relationship between average disposals and standard deviation................. 89
Figure 6-16 Relationship between key players missing and average disposals............... 90
Figure 6-17 Average disposal by ranking within the team 91
Figure 6-18 Average disposals by team.. 92
Figure 6-19 Average disposals by opposition... 92
Figure 6-20 Team and opposition experience... 93
Figure 6-21 Difference in experience 94
Figure 6-22 Average disposals by host venue 96
Figure 6-23 Previous games experience at each venue 97
Figure 6-24 Average disposals for distance travelled to venues...................................... 98
Figure 6-25 Disposals by match result.. 98
Figure 6-26 Disposals by predicted result 99
Figure 6-27 Average disposals for reciprocal of bookmaker prices for team outcome.. 100
Figure 6-28 Average possessions per round 101
Figure 6-29 Average disposals by year... 102
Figure 6-30 AAE in training and holdout data for 34 predictors of disposals................ 106
Figure 6-31 Relationship between winning percentage and AAE in training data......... 111

Figure 6-32 Relationship between AAE and goodness of fit (LogProb) 112
Figure 6-33 Relationship between goodness of fit (LogProb) and ROI 113
Figure 6-34 Relationship between AAE and ROI 114
Figure 6-35 Model R-square and ROI for 15 stages of model development 116
Figure 7-1 Distribution of runs scored by batsmen in ODI cricket 120
Figure 7-2 Distribution of log(runs+1) for batsmen in ODI cricket 121
Figure 7-3 Runs scored and balls faced for each position in the batting line up 123
Figure 7-4 Relationship between experience and runs scored 124
Figure 7-5 Relationship between experience and runs at start of career 125
Figure 7-6 Average runs scored per batsmen against each team since 1999 127
Figure 7-7 Average runs scored per batsmen for each team since 1999 128
Figure 7-8 Average absolute errors for predictors of form 130
Figure 7-9 Comparison for different predictors of quality 131
Figure 7-10 AAE for differing prediction models 135
Figure 7-11 Distribution of differences in scores for matched batsmen 137
Figure 7-12 Relationship between AAE and goodness of fit (LogProb) 140
Figure 7-13 Relationship between goodness of fit (LogProbs) and ROI 141
Figure 7-14 Relationship between ROI and AAE 142
Figure 8-1 Average number of votes polled for the best named players in order 149
Figure 8-2 Relationship between polling votes & kicking goals 150
Figure 8-3 Relationship between polling votes and taking marks 151
Figure 8-4 Relationship between polling votes and player position 152
Figure 8-5 -2 log likelihood for each stage of model development 156
Figure 8-6 Average rank for vote getters as variables are added to the model 157
Figure 8-7 Within game goodness of fit measures for stages of model development 158
Figure 8-8 AAE for seasonal vote totals for 25 stages of model development 159
Figure 8-9 Predicted top 10 who finish in top 10 and AAE <1, 2 & 3 160
Figure 8-10 Average rank with differing training and holdout sample sizes 162
Figure 8-11 Within game measures of goodness of fit 163
Figure 8-12 Seasonal AAE for all players and leading 20 players 164
Figure 8-13 Season measures of goodness of fit for differing holdout samples 165
Figure 9-1 Histogram of runs scored per over in ODI cricket 173
Figure 9-2 Average runs scored per over (both innings combined) 175
Figure 9-3 RPO for wickets fallen for the three stages of the first inning 177
Figure 9-4 RPO for wickets fallen for the three stages of the second inning 177
Figure 9-5 Relationship between partnership and runs per over 178
Figure 9-6 AAE for predictors created from previous overs bowled 180
Figure 9-7 RPO and run-rate for the previous five overs 181
Figure 9-8 RPO and run-rate for the last three overs from the same end 181
Figure 9-9 Average runs scored in the first fifteen overs 182
Figure 9-10 Run rate for host country 183
Figure 9-11 RPO and overs faced for the batting team 184
Figure 9-12 Average RPO for each country's batting performances stratified by stage 185
Figure 9-13 RPO for opposition stratified by stage 186
Figure 9-14 AAE and log likelihood for stages of development (1st innings) 188
Figure 9-15 AAE and log likelihood for stages of development (2nd innings) 188

Figure 9-16 R-square statistics showing model development for both innings............ 189
Figure 9-17 Comparison of AAE for the five distributions... 190
Figure 9-18 Comparison of average LogProb between models for first innings............ 192
Figure 9-19 Comparison of average LogProb between models for the second innings. 192

List of Tables

Table 4-1 Advantage of familiarisation at a given venue 34
Table 4-2 Average absolute error and percentage of winners predicted 37
Table 4-3 Parameter estimates for multivariate model 39
Table 4-4 Model comparison for AAE, percentage of winners predicted and ROI 41
Table 4-5 Statistical comparisons between models 42
Table 5-1 Percentage of resources available for overs remaining and wickets lost 48
Table 5-2 Comparison of MOV using only Duckworth and Lewis 49
Table 5-3 Comparison of MOV between innings using de Silva's adjustment 50
Table 5-4 Average MOV between classes for non-neutral matches 53
Table 5-5 Average MOV for distance travelled 54
Table 5-6 Number of matches played and hosted by each country 55
Table 5-7 Familiarization within a given country 56
Table 5-8 Familiarization within a given country at an individual player level 57
Table 5-9 Advantage of familiarization at a given ground 58
Table 5-10 Ground familiarization determined at an individual player level 58
Table 5-11 Comparison of continuous predictors for familiarization 59
Table 5-12 Difference in experience derived at both a team and individual level 60
Table 5-13 Multivariate models constructed at a team and individual level 64
Table 5-14 Model comparison for AAE, percentage of winners and ROI 67
Table 6-1 Factors that may affect player performance 74
Table 6-2 Average disposals by bookmaker rankings 107
Table 6-3 Results for the 34 univariate models, sorted by ROI 110
Table 6-4 Parameter estimates and model R-square for final 15-variable model 115
Table 6-5 Return on investment for each stage of the multivariate model 117
Table 7-1 Average runs for batting position 123
Table 7-2 Home country advantage for leading batsmen in ODI matches 125
Table 7-3 Interaction between match time and innings sequence 126
Table 7-4 Average runs scored and runs conceded for each team from 1999 - 2003 128
Table 7-5 Parameter estimates for runs scored by batsmen in ODI matches 132
Table 7-6 Results from 17 prediction models sorted by return on investment 139
Table 8-1 Example of individual match statistics available on the Internet 147
Table 8-2 Goodness of fit for the 25 stages of the multivariate model 155
Table 8-3 Seasonal goodness of fit measures for each stage of model development 161
Table 8-4 Actual and predicted Top 10 Brownlow results 2000-2003 167
Table 9-1 Example of data collected from the second innings of an ODI match 172
Table 9-2 Goodness of fit for stages of model development for both innings 187

Table of abbreviations

Abbreviation	Definition
AAE	Average Absolute Error
AFL	Australian Football League
B	Behinds
BMI	Body Mass Index
BREM	Balls remaining
CL	Clearances
D-L	Duckworth-Lewis
DI	Disposals
FA	Free Against
FF	Frees For
G	Goals
HA	Home ground Advantage
HO	Hit Outs
HtH	Head to Head
IN50	Inside forward 50
MA	Marks
MCG	Melbourne Cricket Ground
MLL	Maximum log likelihood
MOV	Margin Of Victory
NBA	National Basketball Association
ODI	One Day Internationals
RE	Rebounds
RO	Runs off the over
ROI	Return On Investment
RPO	Runs Per Over
RR	Run-rate for the innings
RR5o	Run-rate from the last 5 overs
RREQ	Runs required to win
RRREQ	Run-rate required to win the match
SD	Standard Deviation
se	Standard Error
TH	Total Handballs
TK	Total Kicks
TK1	Tackles
TO	Turn Overs
UAE	United Arab Eremites
VFL	Victorian Football League
Wr	Wickets remaining
LogProb	Log of probability
Ave.	Average

Abstract

The aim of this thesis is to establish a consistent statistical approach to aid in the prediction of sporting outcomes. This process has been achieved by considering six differing outcomes relating to both AFL football and cricket that cover aspects of team performance, individual performance and other associated match outcomes. The advent of the Internet has created a wealth of electronic data that has facilitated the use of particularly large datasets to identify historical features that could independently explain statistically significant portions of variation associated with an outcome. Multivariate modelling was used to weight the contributing effects of each significant variable and produce prediction equations that, when applied to holdout samples, could determine the predictability of the outcome and the predictive capacity of the chosen approach. Predicted outcomes were converted into probabilities and where betting markets were present, the efficiency of the corresponding market was determined. Where wagering markets were not in existence, this process can be seen to establish benchmarks for future reference. While undisputable evidence exists to the inefficiency of price setting techniques currently employed by Australian bookmakers, the process of using past data to aid in the prediction of sporting events can be seen to have benefits for both punters and bookmakers alike. While punters can use such techniques to exploit market efficiencies in the short term, the objectivity and speed provided by a computer driven process will undoubtedly replace the more traditional methods of price setting currently incorporated by bookmakers.

1 Introduction

The city of Melbourne in the state of Victoria, Australia is often referred to, (mostly by its own inhabitants) as the sporting capital of the world. Nestled on the bank of the Yarra River, in the heart of Melbourne is one of the most illustrious of sporting venues, the Melbourne Cricket Ground (MCG). Each September, 100,000 fans of Australian Rules football converge on the MCG to watch two teams compete in the ultimate match of the season - the Grand Final. Within months, the MCG reverts to a cricket ground and plays host to the most famous match of Test cricket on the world sporting calendar – the Boxing Day Test. During the festive period between Christmas and New Year, it is not unusual to see in excess of 200,000 fans journey to the MCG to watch the Australians compete against another of the world's leading cricketing nations.

Growing up as a sports mad Victorian, my childhood aspirations were always dictated by the season. On cold winter nights, a leather football was my night time companion as I dreamt of playing on the MCG. As the seasons turned, the warm summer evenings would bring countless hours of backyard cricket and subsequent aspirations of wearing the baggy green – the ultimate accolade for an aspiring Australian cricketer. As time went on, the reality that I would neither play football nor cricket at an elite level slowly sunk in, but the passion for these sports never waned. As a consequence, this dissertation is dedicated to the prediction of outcomes associated with both AFL football and cricket.

Two and half thousand years ago the Chinese philosopher Confucius was quoted, 'Study the past, if you would divine the future'. The approach of this thesis is to use historical data to aid in the prediction of future events. Whilst the process of using past data to predict football and cricket matches has been explored in the literature, the history is essentially contained to predictions of the match outcome. This dissertation looks to expand upon the current literature by establishing a consistent statistical approach that allows the user to predict not only the winner of matches with a greater accuracy than previously shown, but also to predict any other related sporting outcome for which sufficient past data is available. In particular, there is an emphasis on predicting micro events such as the individual performance of players from within the two competing teams. Because football and cricket are both team sports, how the team performs is dependent upon the performance of the individuals within the team. Modelling performance at an individual player level not only provides for potential improvement in the quality of team predictions, but further allows for the individual performances to be compared between players.

The primary aim of this thesis is to establish a consistent statistical approach to enable the prediction of sporting events. In order to do this, a range of different sporting outcomes has been used. Although Australian Rules football and cricket are the only two sports considered, the diversity that exists between these two sports opens the door for comparison with other team sports.

A practical legacy of a statistically derived prediction approach is that it provides a platform from which market efficiency can be assessed. It is a secondary aim of this thesis to establish a process that will readily enable the determination of market efficiency in sports betting markets should they currently exist or come into existence in the future.

Chapter 2 of this thesis will provide a background and a review of the literary contribution relevant to this body of work. Specifically, the five topics integral to this thesis are AFL football, cricket, prediction models, home ground advantage (HA) and market efficiency. A review of the past literature establishes the framework from which to build and assess prediction models.

To establish a consistent approach to predicting a sporting outcome and determining the efficiency of a corresponding wagering market requires a methodology that can readily be repeated. The approach used in this dissertation can be outlined by the following eight steps:

1. Collect and collate past data.
2. Determine the underlying distribution.

3. Identify significant predictors.
4. Partition data into training and holdout samples.
5. Construct a multivariate prediction model.
6. Establish a link between the quality of the prediction model in the training data and the goodness of fit of predictions in the holdout sample.
7. Convert predicted outcomes into probabilities that correspond with bookmaker prices.
8. Use an appropriate waging strategy to determine market efficiency.

Chapter 3 of this thesis discusses each of these eight steps in detail, effectively setting the platform for the forthcoming analysis. Commencing with chapter 4, the analysis for this dissertation will progress in three stages, with each stage comprising of two chapters; the first relating to Australian Rules football and the second to cricket.

In the first stage (chapters 4 & 5), models are constructed to predict the outcome of matches, while in the second stage (chapters 6 & 7) the aim is to predict individual player performances. In the third stage (chapters 8 & 9), predicted outcomes are related to individual performance, but occur either during a match or as a consequence of the match. At each stage of the analysis, outcomes relating to both sports are compared, providing insight into the wider application of this modelling approach. In stages one and two, match and individual outcomes have been chosen because betting markets already exist, allowing for market efficiency to be assessed. In stage three, the outcomes analysed were chosen in the hope that they would create new betting markets for the future.

Following these three stages of analysis, chapter 10 will seek to summarize and conclude these findings.

2 Background and literature review

2.1 Australian Rules football

Australian Rules football is a team sport that originated in and is primarily played in Australia. Commencing in 1897 with eight teams from Victoria only, the Victorian Football League (VFL) has grown into a national competition under the Australian Football League (AFL), and now boasts 16 teams from five Australian states. Each match is played between two teams of 22 players with only 18 players on the field at any given time. The ball is moved from one end of an oval shaped ground to another by either hand or foot with the team objective being to score goals. Six point goals are scored when the ball is passed by foot through the middle set of four upright posts. Should the ball pass through the outer posts or pass through the posts without being kicked, a single point is awarded. Matches are played over four 20 minute quarters, with extra time added for stoppages, ensuring the duration of most games is approximately two hours.

Despite being in existence for over a century, the literary contributions relating to the prediction of Australian Rules football are reserved to only a few authors, with the bulk of work attributed to Stephen Clarke.

Whilst originally following the lead of Ray Stefani, Clarke first published on Australian Rules football in 1981. Based loosely on the Elo (1978) system used to rate chess players, Clarke used an exponential smoothing process to develop team ratings and ground advantage. Using only the names of the competing teams and the venue, this process was used to predicted winners, margins and chance of winning. The program was originally optimised in 1980 using only one year’s past data, and re-optimised in 1986 using 6 years of data. Even so, the program has proved it can match it with the media experts in picking winners and margins as shown by Yelas (2003).

One of the early pioneers in the field of mathematically predicting sport in general was Ray Stefani. Stefani developed a least squares system in 1970 and first published it in 1977. This approach, in which ratings were created by averaging opponent's ratings with margin of victory (MOV), was applied to American football and basketball at both the college and pro levels. Stefani further added two major improvements to this approach with a compensation for overestimation based on Stein's Paradox as reported by Efron and Morris (1975), suggesting all individual predictions regress towards the mean and with an additional adjustment for HA as seen in Stefani (1980). Not satisfied with modelling American sports, Stefani and Clarke (1992) teamed up to further provide substantial insight into predicting the outcomes of AFL matches. Using both a least squares approach proposed by Stefani and an exponential approach favoured by Clarke this analysis was one of the first to statistically compare the benefits of averages and smoothing techniques to develop team rankings. Both these approaches report around 67% accuracy in predicting winners.

Whilst this dissertation draws heavily from the work of Stefani and Clarke, it also seeks to expand on their work in several ways. Where Clarke and Stefani provided analysis on 16 seasons worth of data (2361 matches), this thesis uses 107 seasons (12462 matches) of data. This five-fold increase in data enables statistical significance to be determined with an increased accuracy, and provides the first comprehensive analysis of the complete history of Australian Rules football.

An added benefit of analysing all past data is that it enables a greater understanding of the phenomena of HA. Whilst the existence of HA cannot be questioned, the explanations for HA are still in debate. Schwartz and Barsky (1977) proposed three explanations for HA, namely travel, familiarization and crowd support. While Stefani and Clarke (1992) confirmed the existence of HA in Australian Rules football, no attempt was made to determine the specific reason. While the subjectivity of crowd support will always be difficult to measure, by considering distances travelled in conjunction with the amount of experience gained at the specific venue, this dissertation will seek to statistically tease out the separate HA effects of travel and familiarization.

Finally, whilst Stefani and Clarke use exponential smoothing and averages to produce team ratings, no attempt has been made to separate long term measures of quality from short term measures of current form. Using multivariate modelling this thesis disseminates and numerically weights the contribution effects of travel, familiarization, overall quality and current form to produce superior predictions of match outcomes to those currently in existence.

Whilst Clarke provides an excellent reference point for the development of match prediction, when considering individual performance in AFL football, no literary benchmarks are currently available. When clear reference points are no longer provided by the literature, an alternative point of comparison can be found by considering bookmaker prices. In recent years, the increase in competition amongst Internet bookmakers coupled with the Australian desire to gamble on virtually anything has spawned many new bet types. Of particular interest for this thesis is when two individual players from competing teams are matched together with punters betting on who will have a superior performance. This bet type is commonly referred to as player Head to Head (HtH) betting.

Although player performance in AFL football could be measured in various ways, for the purpose of HtH betting, the outcome is determined solely by the number of disposals that each player gathers throughout the course of the match. When a player gathers possession of the football, to avoid penalty he must correctly dispose of the ball by hand or foot. The resulting handball or a kick is referred to as a disposal, with official results recorded on the AFL official website. The collection of specific match information not only aids in the prediction of disposals but also aids in the third area of AFL football covered in this thesis – the prediction of Brownlow votes.

The Brownlow medal is the highest individual honour that can be achieved by AFL footballers. In each of the 176 home and away matches for a season, votes are assigned to the three best players (3 – first, 2 – second, 1- third) by the umpires that preside over the game. With the use of an ordinal logistic regression model retrospectively applied to past data, Bailey and Clarke (2002) were able to identify specific player performance statistics from each match that aided in the prediction of who will poll votes. With the addition of extra data, this dissertation will not only seek to develop additional variables to aid in the prediction process, but will also address the issue of whether statistical significance equates to clinical or practical significance. A further benefit of this process will be to determine the optimal amount of data required to successfully model Brownlow medal voting.

2.2 Cricket

The first official international match of test cricket was played between Australia and England at the MCG in 1877. Test cricket is played between two teams of 11 players with each test lasting up to five days. For close to 100 years, test match cricket was the only form of international cricket that was played. In 1971 the first official ODI match was played, once again between Australia and England at the MCG. While ODI cricket is still played between two teams of 11 players, as the name suggests, the duration of the match is for one day only and is sometimes referred to as limited overs cricket.

Since the inception of ODI cricket, there have been various rule changes, although general principles have remained the same. Both sides bat once for a limited time (maximum 50 overs) with the aim in the first innings to score as many runs as possible, and in the second innings to score more than the target set in the first innings. Because an ODI match is comprised of two different stages (batting & fielding) teams are chosen in order to maximize performance in both areas. Generally a team will consist of specialist batsmen and specialist bowlers, with better batsmen batting higher up the order. Several constraints are imposed upon the fielding team, with no player being allowed to bowl more than 10 overs, ensuring that at least five different bowlers are used to bowl the required 50 overs. In addition, during the first 15 overs, no more than two players are allowed to field outside a specific ring placed about 30 metres from the pitch, and two players must field in what are deemed to be catching rather than run saving positions. Upon completion of the first 15 overs, restraints are relaxed with up to four fielders then allowed to field outside the inner circle, and catching fielders no longer made compulsory.

Whilst cricketing purists will always argue that test match cricket is a better form of the game than ODI cricket, there can be no denying the increased popularity for the shorter version. In the 34 years since its inception, in excess of 2200 ODI matches have been played. In comparison, only 1732 official test matches have been played in 128 years.

As an international sport it is of little surprise that cricket has attracted more attention in the literature than football. One of the first published bodies of work on cricket was presented by Elderton and Elderton (1909). Coincidently, this discussion of the distribution of batsmen scores was also one of the first published bodies of work on statistical methods. The distribution of batsman scores in test cricket was further looked at by Wood (1941 & 1945) and Elderton (1945), who investigated if the scores were geometrically distributed. More recently, Reep, Pollard & Benjamin (1971 & 1977) have explored whether a negative binomial distribution may in fact be a more appropriate fit to the distribution of batsmen scores. Kimber and Hansford (1993) also investigated the hazard function of top batsmen. An excellent summary of the work done on the distribution of scores in cricket can be found in Clarke (1998). In general these papers found that while the geometric distribution is a reasonable approximation, there are slightly more ducks and scores less than five, and slightly more very large scores than suggested by the geometric distribution. More recently, Allsopp and Clarke (2004) have used multiple linear regression to rate the relative batting and bowling strengths of cricket teams.

Clarke (1991) was one of the first to identify differences in the distribution of batsmen scores in ODI cricket compared to test cricket, with his findings that the observed frequency of very small and very large scores were actually less in ODI cricket than expected by a geometric distribution. This result was further confirmed by Bailey and Clarke (2004a) who found that the fielding restrictions and time constraints of ODI matches combined to ensure that the distribution of batsmen scores in ODI cricket could be reasonably well approximated by a log-Normal distribution, whilst comparisons made between batsmen could still be achieved using a geometric approach.

Despite the author harbouring an underlying preference for test cricket, for the purpose of this dissertation, only ODI cricket will be considered. There are several reasons why this is the case. Firstly, in order to predict a match outcome using a multivariate approach, there is an underlying requirement that the distribution of the outcome is parametric. In ODI cricket, the MOV can be readily quantified by converting remaining resources into runs, producing an outcome that is approximately normal. Because test match cricket is always played over two innings spanning a five days period, the MOV is more difficult to quantify. Deteriorating pitch conditions, time constraints, personal milestones or overall series scores can all contribute to the way in which a match is played. Because resources are sometimes sacrificed in order to win the game, a MOV reflected by runs or resources could often be misleading.

Secondly, as highlighted by Clarke (1991), the distribution of batsmen scores in test cricket is more difficult to quantify than that of ODI cricket. If the distribution of the outcome of interest cannot be well characterised, the validity of a multivariate modelling approach will be brought into question. Whilst time constraints and fielding restrictions ensure that batsmen scores in ODI can be reasonably well approximated using a log-Normal distribution, this is not the case for test match cricket as there are greater occurrences of very small and very large scores.

An additional reason why ODI cricket is preferred for mathematical modelling over test match cricket is the increased availability of data. With more than twice as many ODI played annually in comparison to test matches, this not only provides a greater wealth of data, but creates greater opportunity to explore market efficiency. In addition, more information is available for ODI matches than for test matches. In particular, the number of runs scored per over (RPO) is easily accessed for ODI matches, but not for test matches.

2.3 Predicting sporting outcomes

When predicting sporting outcomes, fundamental differences occur between the prediction of individual and team performance. Because individual player performances in football and cricket can be directly measured by the number of possessions gathered or runs scored, construction of multivariate models can readily proceed by identifying and weighting factors that contribute to performance. The prediction of match outcomes however, is dependent upon the performance of the two competing teams. Whilst the MOV referenced against the home team provides a normally distributed outcome for both AFL football and ODI cricket, valid predictors of the MOV must then take into account the relative performances of both teams. This is usually achieved by rating both teams and then using the difference in ratings as a predictor of MOV.

Mathematical rating systems for sporting events have been in existence long before the introduction of computer based techniques. Stefani (1998) provides a comprehensive review of 83 different sports classified into one of three categories, combat sports, object sports and independent sports, and reviews rating systems for each type of sport. For object team sports such as AFL football and ODI Cricket, Stefani compared a range of different rating approaches. Firstly he considered a least squares approach (Stefani 1977) with and without an adjustment for regression to the mean by James and Stein (1981). In addition, a probabilistic approach proposed by Harville (1980) and an exponentially smoothed approach favoured by Clarke (1993) were also considered. Stefani concluded that very little difference could be found between the four differing prediction approaches and to achieve additional accuracy would require more of an adjustment than for HA alone. This dissertation aims not only to quantify HA more accurately, but dissect team ratings into the separate components of overall quality and current form.

2.4 Home advantage

In addition to predictors of quality and form, the role of HA has been shown to play an integral role in any analysis of sporting events. The concept of HA has long been recognised as a known phenomenon in sport and has been basis for much research. Schwartz and Barsky (1977) provided one of the first detailed studies examining HA in four leading American sports namely baseball, gridiron, ice hockey and college basketball. Subsequent studies have shown HA to exist to some degree in many other professional sports. Pollard (1986), Barnett and Hilditch (1993) and Clarke and Norman (1995) confirmed the existence of HA in professional soccer, whilst Holder and Nevill (1997) have explored the extent of HA in individual sports such as tennis and golf. Courneya and Carron (1992) provide a helpful taxonomy listing a range of studies covering five American sports namely soccer, hockey, baseball, basketball and gridiron. HA is not limited to sporting events in the northern hemisphere. Lee (1999) has confirmed HA in Australian Rugby League, whilst Stefani and Clarke (1992) has explored HA in Australian Rules football. Whilst only considering about 12% of all matches played, Stefani & Clarke found HA in AFL football to be worth between one and two goals. This result was further confirmed by Bailey and Clarke (2004b).

Looking at pairs of matches from the 1980s, Stefani and Clarke (1992) found evidence for individual HA. Clarke (2005) has a detailed analysis of HA in the AFL for the period 1980-1998. He found significant evidence for an interstate advantage. Although the draw is not balanced for opponent or ground, Clarke also showed there is little difference in the HA produced by a regression model allowing for team ability and the simpler process of taking the average winning margin of the nominal home side.

Whilst many different approaches have been used to quantify HA, the underlying reason why HA exists has been reduced to three basic principles; travel, familiarization and crowd support. Although studies pertaining to HA have endeavoured to attribute the phenomena to one or more of these three components, little has been done in order to numerically quantify the independent effects of these three features in AFL football or

ODI cricket. By considering distance travelled in conjunction with the amount of experience gained at the specific venue, it is an aim of this dissertation to statistically tease out the separate effects of travel and familiarization.

HA has also been explored with regards to cricket. Davis (2000) has conducted HA analysis on all tests played between 1877 and 2000, finding that sub-continental teams enjoyed the greatest advantage. Likewise, de Silva, Pond & Swartz (1991) have examined the HA in ODI cricket matches and conclude that a side playing on their home soil have an increased probability of winning. These findings are supported by Bailey and Clarke (2004a) who found that the HA advantage for individual batsmen in ODI cricket was worth approximately two runs.

2.5 Market efficiency

Perhaps the single greatest reason to use mathematical models to predict sporting outcomes has been for financial gain. Whilst racing horses for sport can be traced back to the Crusades in the 12th century, it was not until in the early 1700s that horse racing became a professional sport with spectators betting on the results. Although punters have long sought to profit financially, the bulk of work published on race track betting has only developed in the later part of the 20th century.

A long list of authors has endeavoured to show market inefficiency in horse racing through the use of profitable betting strategies. The most comprehensive collection of work to date, detailing efficiency in betting markets has been compiled by Hausch, Lo & Ziemba (1994). This volume includes a collection of 61 papers covering statistical, mathematical, financial, economic and psychological aspects of inefficiencies in betting markets. Whilst most papers from this collection provide insight into aspects of horse racing, one particular paper, written by Benter (1994), has been a catalyst for this thesis.

Following on the work of Brecher (1980) and Chapman (1994), Benter used multinomial logit modelling to assign winning probabilities to horses running in Hong Kong horse races. Using complete past information on each runner in the race, a multivariate model was constructed to provide an unbiased probability of each horse running first, second or third. A minimum of 1000 races was used to develop a model including measures of current condition, past performance, specific preferences and current race information. Although horse racing differs in many ways from both football and cricket, the underlying principle of using past data to statistically develop a predictive model remains the same. Compared to a horse race, predicting the outcome of a two competitor event can often be simpler as the lower the number of competitors in an event, the less complicated the analysis becomes. From a statistical perspective, a binomial outcome is more readily understood, analysed and interpreted than a multinomial outcome. In addition, a higher degree of statistical power can be obtained when a binomial outcome can be defined retrospectively by applying a cut-off to a continuously normally distributed outcome. Because the MOV for both AFL football and ODI cricket can be shown to be approximately normally distributed, the probability that either team will win can be more accurately obtained by modelling the MOV rather than the actual winner.

The use of mathematical models to examine the efficiency of betting markets is by no means limited to the sport of racing. In recent times, betting has seen an expansion into sports where betting on the outcome has not been the primary reason for the sport to survive. With over 1000 online sports bookmakers available internationally, it is now possible to bet on virtually all sports of some recognition. The evolution of sports betting has bought with it a plethora of academics and economists reporting on the efficiency of betting markets. Two factors would appear to contribute to the depth of work in each sport, namely the popularity of the sport in question and the duration in which betting has been available on that sport. As Las Vegas bookmakers have been betting on sports since the 1960s it is of little surprise that extensive work has been written on the efficiency of American sports. As a consequence of the constant attention and improvement, most American sports betting markets are now considered efficient. The primary reason for this efficiency has been the bookmakers' ability to adjust to the market and improve over time.

Betting on the outcome of Australian Rules football and cricket has only been available since the late eighties. Brailsford, Gray, Easton & Gray (1995) were one of the first to explore the efficiency of betting markets for both the Australian Rugby League and Australian Rules football. Using probit and ordered probit models, various betting strategies were explored with the authors reluctant to suggest that conclusive evidence of market inefficiency was in existence. At the time that this study was published (1995), sports betting in Australia was based on a pari-mutuel system[1] with bookmakers guaranteed a fixed profit. Since then, increased competition has forced bookmakers to offer more competitive fixed price sports betting. While bookmakers do have a distinct edge provided by the margin[2], this only guarantees profit if the book is completely balanced. In reality, the increase in competition and the brief period of time that sports books are open for, ensure that most books are seldom balanced, further increasing the risk to the bookmaker.

Although Clarke (1993) has been using mathematical models to accurately predict AFL results since 1981, it wasn't until Yelas (2003) that some attempt was made to use these predictions to determine market inefficiency. Bailey and Clarke (2004b) further established the inefficiency of the AFL betting market using both an existing benchmark provided by Clarke and two multivariate models, derived at both an individual and team level. It is this publication that forms the nucleus of chapter 4 of this dissertation.

Whilst several attempts have been made to model both test match and one-day cricket, no formal attempts have been made to explore the efficiency of match outcome in ODI cricket. This process will be discussed extensively in chapter 5.

---

[1] Bookmakers would take a fixed percentage of all money that came in and then pay out the remainder to the winners, ensuring that final payout prices were unknown until the betting pool was closed.
[2] Bookmaker margin is the amount by which the summed probabilities for participants exceed one and represents the profit the bookmaker stands to win should he balance his books perfectly

Whilst initial books on sporting outcomes were set on the winner of matches, the growth of Internet betting and increased competition between bookmakers has resulted in many exotic bets being created, with bet types only limited by the imagination of bookmakers and the demand of the public. One recently introduced bet type that has found a steady increase in popularity, is wagering on the relative performance of two individuals competing in a team. Examples include whether one footballer will get more possessions than another, or if one batsman will score more runs than another. Because the performance of one player is generally independent of the other, comparisons between players can readily be made, and each player's probability of outperforming the other determined. Because betting of this type has only been available to punters in the last few years, these markets have a higher probability of inefficiency than more established markets. Since data collection, analysis and publication can take years from commencement to completion it is of no surprise that very little formal analysis reporting on the efficiency of player head to head betting has reached the literature. Perhaps the only publications to touch on this area Bailey and Clarke (2004a) will be presented in chapter 7 of this dissertation.

When exploring market efficiency, a variety of approaches have been used to identify and then classify the potential "inefficiency" of a given market. Fama (1970) defined an efficient market as one in which the price fully reflects all available information. Based on this definition, Fama further defined three levels of testing to quantify the level of information content as being 'weak' when only involving past prices, 'semi-strong' when involving all publicly available information or 'strong' when all possible information is included. This concept aligns well with a statistical view point in that the level of inefficiency in a betting market is directly proportional to the quality of the prediction model. Given that the general public will be unlikely to have complete information available on the fitness of competing players, the tests for efficiency used in this dissertation will equate to being 'semi-strong'. From a practical, statistical and economic point of view, a clear cut definition of inefficiency will be used in this thesis: statistical evidence that a wagering strategy can produce a profit that is significantly greater than zero. It is here the use of large data sets further aids in identification of a statistically inefficient market.

3 Methodology

3.1 Background

The statistical methods used in this thesis have been drawn from ten years experience as a statistician for a large teaching and research hospital. Over this period of time, it has been my privilege to conduct statistical analysis for a remarkably broad array of research, each making a unique contribution to the field of medicine[3]. Regardless of being dust mites, cancer patients, knee cartilages or surfboard riders, each statistical analysis was performed in a similar fashion using univariate and multivariate analysis.

Univariate analysis is the first stage in any statistical analysis and is used to explore the direct relationship between individual variables and an outcome. Although univariate tests give a good indication to the strength and nature of the relationships of interest, they are by no means definitive. To further strengthen and validate results, the use of a multivariate model represents the final step in any analysis. Multivariate models in medical research primarily involve the use of multiple linear or multiple logistic regression and are used to ensure that predicted results are not attributable to confounding sources. Multivariate modelling is the optimal way in which to maximise information derived from the available data, and represents a standard approach adopted by the medical community for acceptance into published medical journals.

In the field of medicine, multivariate analysis is generally used to describe an analysis in which several variables are used simultaneously to predict an outcome of interest. In advanced statistics, the term multivariate regression is usually reserved for situations in which there is more than one outcome. In this thesis I will often refer to the use of multivariate analysis but it is important to note that when stating this I will be using the

[3] See Appendix 1 for a complete list of previously published multivariate analysis

former definition, and will only ever be considering one outcome at a time, but with multiple predictors for each outcome.

In this dissertation, the standard statistical approach used in medical research has been applied to AFL football and ODI cricket. Whilst the statistical techniques used have changed little over the past 100 years, it is the ways in which these techniques have been applied that make this body of work unique. Two features differentiate this work. Firstly, there is a focus on individual player performance and how individual data can be used to aid in the prediction of team results. Although player performance has been addressed in individual sports such as tennis or golf, the specific performance of individuals within team sports such as football and cricket, remains undocumented. The second feature of this work that sees it differ from most others is the amount of data used to conduct the analysis. Rather than choose a sample size of convenience, every single official match played for both AFL football and ODI cricket have been used to analyse and develop the prediction models.

3.2 Data

To establish a modelling process firstly requires the collection of past data, and the development of a process to collect data for the future. The advent of the Internet has created the opportunity to download large amounts of data effectively and efficiently, rendering data entry by hand to be error prone and inefficient. While automated web collection programs can readily download pages of information from the Internet, extensive programming is then required to extract the appropriate data. Data collection processes must be suitably robust enough to ensure that websites with regularly changing structure and format do not present collection problems. Once the relevant data has been collected it must then be transferred into a format understood by a statistical package.

Although most statistical packages can readily determine statistical significance, few offer the data manipulation capacities of SAS[4]. Of particular benefit is the ability to write and store programs relating to the manipulation and analysis of the data. As data are often collected from differing sources, each new data source must be cleaned and formatted. In order to avoid bias of overfitting, all prediction variables must be created using only data relating to performances prior to the current event, creating additional programming requirements. Only when the data set has been properly prepared and collated, can analysis commence, with differing statistical analysis often requiring changes to the data structure.

As a statistician, I firmly believe that it is impossible to have too much data. Statistics are primarily used to determine how unusual any given relationship is, and the more data that are available, the more accurately we can define the result. Prior to the 2004 season, 12,462 AFL matches had been played over 107 seasons. The match venue, team names and scores were all collected from the internet. In addition, individual player data was collected on all AFL matches played between 1997 and 2003 resulting in a working database of over 60,000 individual player performances.

Prior to July 2004, 2136 ODI matches of cricket had been played in 33 years. Complete match information from each game was collected from the internet, resulting in a database of batsmen performance in excess of 33,000 data points.

The use of many thousands of data points to drive the statistical analyses enables highly statistically significant features to be identified – the greater the statistical significance, the more accurately it predicts an outcome. With this in mind, it has been the aim of this thesis to gather as much information as humanly possible, explore the data to determine the underlying distribution and what features have the strongest relationship with the outcome of interest, and the use these features to construct prediction models.

[4] SAS Institute Inc. Cary, NC, USA

3.3 Underlying distribution

There is an emphasis in this thesis towards constructing multivariate models. The ability to combine a range of prediction variables and numerically weight the individual contributions of each variable after adjusting for others is one of the most powerful and useful aspects of statistical analysis. In order to conduct multivariate analysis, there is an underlying assumption that the distribution of the outcome of interest must be readily identifiable. The validity of a multivariate analysis is heavily dependent upon the underlying distribution, and if the distribution can be well characterised, the corresponding analysis is referred to as being parametric. Exactly what type of statistical analysis is deemed to be appropriate is solely dependent upon the underlying distribution of the data. Each of the following six analyses is devoted to statistically analysing a different outcome relating to AFL football and ODI cricket.

In chapter 4 the outcome of AFL football matches is discussed. Because each game of AFL football is played between two teams, by referencing each result in terms of the home team, i.e. home team score minus the away team score, the resulting MOV can be shown to be well approximated by a Normal distribution. In chapter 5 a similar approach is used to predict the winner of ODI cricket matches. The MOV in ODI cricket matches is generally expressed in terms of runs or wickets, depending whether the winning team batted first or second. By converting into runs the remaining resources (wickets and balls remaining) for winning teams who batted second, and once again referencing each result in terms of the home team, the resulting MOV can be shown to be well approximated by a Normal distribution.

Chapter 6 addresses the number of possessions gathered by individual AFL players throughout the course of the match. An examination of the data shows the number of times leading players correctly disposed of the ball during the course of an AFL match can be well approximated by Normal distribution. Similarly, chapter 7 explores the number of runs scored by batsmen in ODI cricket matches. Whilst there have been claims that batsmen scores in test match cricket follow a geometric distribution, the fielding restrictions and time limitations of ODI cricket ensure that the distribution of runs scored by individual batsmen can be approximated by a log-Normal distribution. Thus, with the use of a logarithmic transformation, the resulting outcome can be treated as approximately normal.

For each of these analyses, an underlying assumption of normality allows for multivariate analysis to be performed using multiple linear regression. The term regression was first introduced in 1885 by Sir Francis Galton when describing the way the size of offspring tended towards the average, leading to the term “regression towards mediocrity”. Although the discovery of “least squares” was generally attributed to Carl Freidrich Gauss, Adrien Marie Legrendre was the first to publish in 1805. The combination of the two led to the term “least squares regression” which is still in use today. Where outcomes can be shown to be well approximated by a Normal distribution, least squares regression produces estimates that are the ‘best linear unbiased estimates’ under classical statistical assumptions (Gauss (1809) & Markov (1900)). While the first two stages of this analysis are developed on normally distributed outcomes, the third stage, seeks to apply a similar approach to differing distributions.

Chapter 8 examines the number of Brownlow votes polled by AFL footballers. The three leading AFL players for each match are awarded votes on a 3, 2, 1 basis by the presiding umpires. The number of votes polled per game is of an ordinal nature (0, 1, 2 or 3) and is analysed using ordinal logistic regression.

Finally, chapter 9 explores the number of runs scored per over in ODI cricket. Each over in ODI matches comprises of six legitimate deliveries. The number of runs scored per ball can range between 0 and 6, meaning that the number of runs scored per over can theoretically range between 0 and 36. The unique scoring structure of cricket ensures that the number of runs per over (RPO) does not clearly follow a readily identifiable distribution. In order to best predict RPO, a range of potential approaches have been used.

The common established theme throughout these six analyses is to use standard statistical techniques applied to very large bodies of past data in order to identify features that can be linked beyond doubt to the outcome of interest. The most appropriate statistical technique to use is dependent upon the underlying distribution of the data, with four of the analyses conducted using multiple linear regression, one with ordinal logistic regression and remaining analysis comparing a range of approaches.

3.4 Identifying significant predictors

Using univariate analysis, variables thought to affect the outcome of interest were tested for statistical significance. Prediction variables used were either categorical or continuous in nature. As continuous variables offer significantly more information than categorical variables, every attempt was made to construct predictors that displayed a linear relationship with the outcome of interest. For presentation purposes, univariate relationships for continuous variables are sometimes presented to the reader in categorical format to highlight the nature and extent of the existing relationship.

As discussed in section 2.3, predictors for match outcomes are more complicated than predictors for individual performances as they require information concerning both competing teams. As a consequence, it is easier to identify significant predictors for individual performances, resulting in the individual models having more significant variables in the multivariate analyses than do the models for match outcome.

A common theme throughout this dissertation is the use of past data to develop numerical measures relating to overall quality and current form. To do this, two processes have been used, rolling averages and exponential smoothing. Whilst averages give equal weight to past performances, exponential smoothing gives more weight to recent performances. Developed by Elo (1978) to rate chess players, a smoothing equation of the form

$$\textit{Smoothed predictor} = x(\textit{last result}) + (1 - x)\ \textit{previously smoothed predictor} \qquad (3.1)$$

is regularly use throughout this thesis. While it would be possible to specifically optimise the smoothing parameter x to produce an optimal fit to the data, this effectively combines a measure of overall quality with a measure of current form to produce a single prediction variable. One goal of this thesis is to separate measures of quality from measures of current form rather than seek to combine them together. For this reason three different smoothing parameters were regularly explored and unless specifically stated these parameters were (x=0.1, x=0.2 and x=0.3). By considering a range of exponential smoothers in conjunction a range of moving averages, the best possible predictors for each given outcome could be assessed.

3.5 Holdout sample

Having identified predictors of the outcomes of interest, there is a clear difference between being able to explain variation in past results and being able to predict future results. For example, in AFL football, inaccuracy could explain a significant proportion of variation in past results, but could not be used to predict future outcomes due to randomness. It is well established that predictions that are applied to the same data in which parameter estimates from the prediction model were determined, will be biased. This bias is often referred to as over-fitting the data. In order to avoid over-fitting, a separate holdout sample is required to accurately determine the predictive capacity of each model. In each analysis conducted, data sets are separated into two; a training dataset and a holdout sample. Multivariate models are developed on the training dataset to identify predictors. Statistical significance is ordinarily defined by a p-value of 0.05 (1 chance in 20), but the use of very large datasets allows for a lower level of significance to be used. A reduction in the accepted level of statistical significance not only decreases the chance of committing a type I error, but further increases the robustness of the prediction approach. The resulting regression models were then applied to data from the holdout sample to accurately determine the predictive capacity.

3.6 Multivariate model building

Multivariate models provide the user with a valuable insight into the relative importance of a group of prediction variables. Much has been written on the most appropriate way to construct a multivariate model with most concluding that there is no established 'best' way to build a multivariate model, as circumstances differ with sample size, variable numbers and types of data. One useful tool when building multivariate models is the use of significance driven selection techniques.

Stepwise forward selection starts by including the most significant univariate predictor in the model. The remaining variables are then examined to determine if they can contribute statistically to the outcome after adjusting for variables already in the model. This process continues until no further variable can be found to make a statistically significant contribution to the model.

Backward elimination works in reverse, by initially including all variables into the multivariate model. At each stage of selection, the least significant predictor is removed from the model, with this process continuing until all variables left in the model make a statistically significant contribution.

Whilst the disadvantages of constructing models based solely on selection techniques are well documented (Hurvich & Tsai (1990), Derksen & Keselman (1992) and Harrell, Lee & Mark (1996)), most problems relate to building multivariate models on small datasets and solely using selection techniques. When used appropriately on large data sets, selection techniques can be of great value.

Having performed many statistical analyses for medical research, I have developed my own technique for constructing multivariate models. By starting with the use of a forward stepwise model, I am able to identify and order the most significant predictors of the outcome. Because it is possible to have two independent variables that at a univariate level are not statistically significant, but when combined together become highly significant, it is possible for a forward stepwise procedure to overlook important variables. This problem is solved by also incorporating a backwards elimination technique. When both approaches produce the same model, this serves to validate the model building process. When models are found to differ, most often the differing variables will be in some way related to each other. The final stage of model building requires assessment of the clinical and biological plausibility of the direction and magnitude of the derived parameter estimates. This final stage further serves to identify any debilitating effects caused by collinearity between prediction variables. Only when these three stages have been addressed can a multivariate model be considered complete.

3.7 Accuracy of prediction

Various approaches are used to determine the predictive accuracy. The percentage of winners accurately identified and the average absolute error (AAE) between the predicted and actual results are used to determine the quality of the model.

One process designed to measure the predictive accuracy of an estimator is the Kullback-Leibler distance. Developed from statistics and information theory by Kullback and Leibler (1951) the Kullback-Leibler distance calculates the efficiency between a true probability distribution and an estimated probability distribution, with a smaller distance indicating greater efficiency. As shown by Dowe, Farr, Hurst & Lentin (1996), the Kullback-Leibler distance can readily be minimised by maximising the sum of the log of the predicted probabilities for the winning outcomes. Thus rather than considering the assigned probability for the winning outcome, the natural log of the probability (LogProb) for the winning outcome will be used. Because the natural log of any given probability will be below zero, the addition of a small constant is used. In a two outcome event, a constant of 0.5 allows predicted probabilities greater than 50% to be assigned positive values, winners with a predicted probability of 50% assigned zero, whilst winners with predicted probabilities below 50% will receive negative values. This process equates to a logarithmic score, which is a standard measure of the accuracy of probabilistic forecasts as given by Winkler and Murphy (1968).

By establishing a link between the goodness of fit of the prediction model developed on the training data and the goodness of fit of the resulting predictions that are applied to the holdout sample, the multivariate approach can be validated. In theory, the model that best explains variation in the holdout sample should in turn be the model that best predicts the result. If this can be readily established, then the addition of each significant variable into the multivariate model will improve the quality of the predicted outcome.

Where possible, comparisons are made with known benchmarks for each outcome, although predicting the outcome of AFL football matches is the only area in which a clear benchmark model has been established. An alternative approach to measuring the predictive capacity has been to compare results with bookmaker prices.

3.8 Bookmaker prices

Bookmakers seldom use mathematical modelling to set prices, rather relying on a combination of knowledge, experience and intuition. Because the economic livelihood of a bookmaker is dependent upon the accuracy in which he sets prices, to stay in business, a bookmaker must be good at what he does. In the absence of published benchmarks, bookmaker prices serve as a clear guide to the predictive capacity of chosen modelling approaches. Conversely, the use of mathematical models enables a practical evaluation of the efficiency of the chosen betting market. Greater insight into the economics of betting markets can be found in Sauer (1998) and Levitt (2003).

The development of the Internet has been accompanied by a growth in sports betting. Several government controlled and private bookmakers take bets over the Internet on a range of outcomes for many sports, including both football and cricket. While the operator’s percentage margin in traditional betting on horseracing is anywhere between

12 and 18%, it can be as low as 5% when only two outcomes are being wagered upon. Several Internet sites compare differing bookmaker prices, and by choosing the best available it is possible to operate in a market with a margin as low as 2 to 3%. This increases the chance that mathematical models can produce a long-term profit by performing better than the public in estimating probabilities. For the purposes of this thesis, bookmaker prices were primarily collected from Centrebet, 24 hours prior to the commencement of the sporting event.

Centrebet was established in 1992 and in 1996 became the first licensed bookmaker in the southern hemisphere to offer online sports betting. With a client base of 200,000 customers and up to 4000 betting events offered per week, Centrebet is one of the leading sports bookmakers in the world.

3.9 Converting predicted outcomes into probabilities

In order to assess market inefficiency, predicted outcomes must be converted into the probability that one combatant will defeat the other. In each of the four betting markets that are assessed for market efficiency, a parametric approach is used to predict outcomes with each outcome approximated by a Normal distribution. Based on the assumption of normality, predicted match outcomes can be readily converted into probabilities by dividing the predicted score by the predicted standard error, and comparing with a standard Normal curve. Similarly, where the distributions of two competing individuals are independently normally distributed, probabilities can be determined by dividing the difference in predicted scores by a prediction for the combined standard deviations of the two competitors and once again comparing with a standard Normal curve. In the case of batting performances in ODI cricket, an alternative conversion approach was also considered based on the assumption that batsmen scores followed a geometric distribution.

3.10 Wagering strategy

Exploring the efficiency of a betting market requires an appropriate wagering strategy. In order to maximise growth of wealth, a betting strategy must incorporate three specific features in determining the size of the wager, namely existing bank size, size of perceived advantage and the probability of winning.

In a fixed price market with only two outcomes, the market price will reflect supply and demand. By multiplying the predicted probability by the market price and subtracting the original unit bet, it is possible to gauge the size of any market imbalance and thus the perceived advantage. This is given by

$$A = (P * M)-1 \tag{3.2}$$

Where A is the perceived advantage, P equals the predicted probability and M represents market price.

Kelly (1956) developed a betting strategy designed to maximise the growth of wealth by maximising the expected log of wealth. This formula can be effectively simplified to

$$B = A / (M-1) \tag{3.3}$$

Where B equals the percentage of total wealth, A equates to the advantage given by (3.2) and M represents market price.

Whilst a strategy that produces maximum profits is desirable, the trade off between risk and return must be balanced. Further research into the Kelly wagering strategy by MacLean, Ziemba & Blazenko (1992) suggest that the Kelly model in its present form can often be too volatile in nature and a fractional Kelly criteria in which a fraction (eg ½) of the recommended bet is placed, offers a greater security. An underlying assumption for the Kelly process is that the punter holds an advantage. In order to ascertain if a punter does indeed hold an advantage requires statistical evidence that profits are significantly greater than zero. As a Kelly criteria is particularly responsive to trends, a clearer guide to market inefficiency can be determined if bet sizes are solely dependent upon the probability of winning and the perceived size of the advantage. Having established statistical evidence that a market is inefficient, additional benefit can then be gained by incorporating a fractional Kelly strategy.

Four of the six analyses covered in this thesis, explore the efficiency of the corresponding betting market. For each analysis, bet size was determined using equation (3.3) with a fixed bank size of $1000.

3.11 Statistical analysis

All analyses have been performed using SAS version 8.2. Univariate analysis was conducted using student t-tests, analysis of variance and Pearson correlation coefficients for normally distributed data, chi-square tests for comparison of proportions and Wilcoxon rank sum or Kruskal Wallis tests for non-parametric data. Average profit per bet was assessed for a significant difference from zero using Wilcoxon sign rank tests. Multivariate analysis was performed using multiple linear regression, multiple logistic regression, ordinal logistic regression and generalised linear modelling. A two-sided p-value of 0.05 was considered to be statistically significant. Continuous results are presented as mean ± standard error whilst variables that are well approximated by a log-normal distribution are presented as geometric means with a 95% confidence interval.

4 Predicting the match result in AFL football

4.1 Introduction

Using match information gathered from 100 seasons of Australian Rules football played prior to 1997, a multiple linear regression model was used to identify and weight numerical features that could independently explain statistically significant proportions of variation associated with the outcome of matches. Prediction models constructed at both a team and player level were applied to matches played between 1997 and 2003 with results compared against an existing benchmark for AFL prediction and bookmakers' prices.

4.2 Background

With the growth of the Internet has come a rapid increase in the amount of readily accessible data from which to explore sporting outcomes such as AFL football. Use of this data should improve modelling accuracy. Bailey (2000) uses team playing statistics such as turnovers and forward thrusts along with bookmaker prices to predict match outcomes. This chapter continues this approach by developing models using all previous match results, and investigates the additional benefits of incorporating individual player statistics in the prediction process.

Because each match of AFL football is played between two teams competing at a single venue, by convention, each game is assigned to have a home team and an away team. Data on home team, away team, venue and final scores for each of the 12,462 games played prior to 2004 were obtained from the Internet[5]. By considering the margin of games as being the home team score minus the away team score, the match result or

[5] http://stats.rleague.com/afl/afl_index.html

MOV is well approximated by a Normal distribution with a mean of eight points and a standard deviation of 40 points. (Figure 4-1)

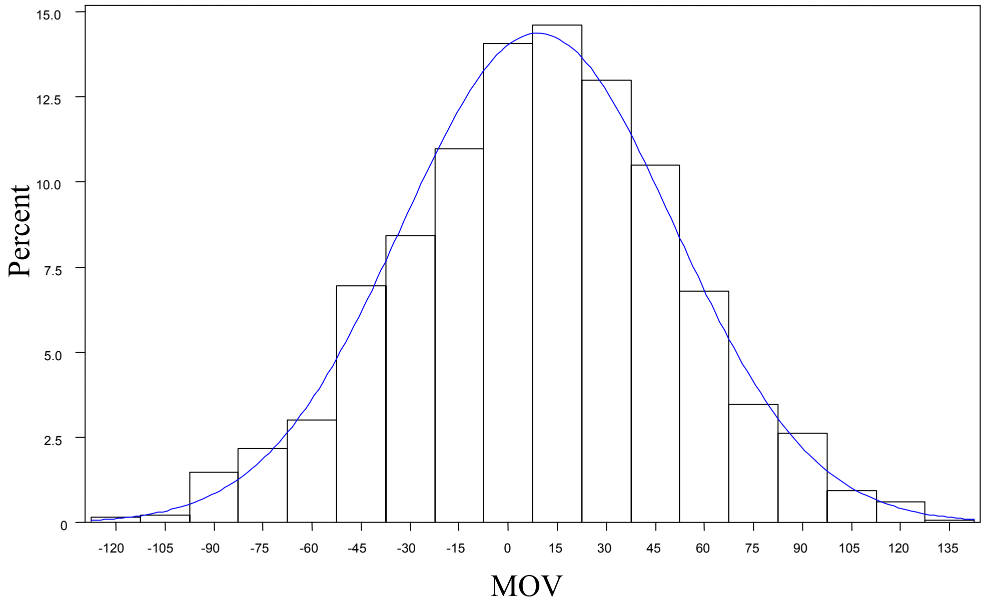

Figure 4-1 Histogram of margin of victory (Home team score minus Away team score)

Compliance with normality enables the use of multiple linear regression to weight the contributing effects of HA, team quality and current form to produce a prediction equation. Predicted margins can be divided by standard errors and compared with the standard Normal distribution to determine the winning probability of competing teams.

As a benchmark to compare the more complicated methods developed here, we use that of Clarke (1993). The predicted winners of this fully automated program have been published in various media outlets, including newspaper, radio, television and the Internet almost continually since 1981. For many years margins have also been published. Studies have shown this program to consistently predict as many winners as the best of the expert tipsters, and to outperform them in predicting margins. In recent years, when analysis of various betting strategies showed it might assist punters to exploit market inefficiencies, the complete output including estimated probabilities of winning, has been distributed via subscription. Records of predicted winners, margins and probabilities for several years past were available.

4.3 Predictors of MOV for AFL football

4.3.1 HA in AFL football

Figure 4-2 shows the total score and common HA (as measured by average winning margin of the nominal home side) in the AFL in five-year periods. Although the overall score for the matches has risen consistently, the HA has remained reasonably constant with a mean figure of eight points. In the past 20 years, it would appear that HA has increased slightly, although this could primarily be attributed to the increase in matches played between teams from differing states.

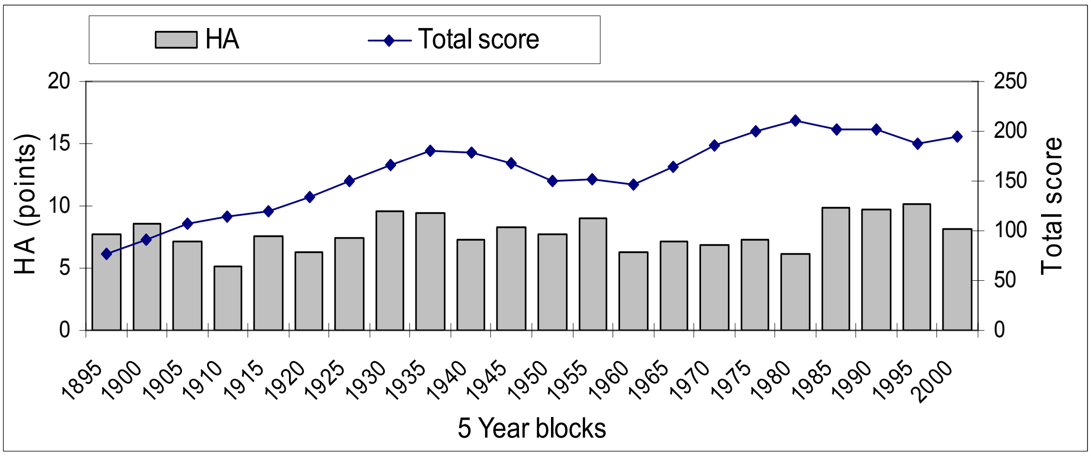

Figure 4-2 Home advantage and total game score stratified in five-year periods

Whilst the existence of HA in AFL football is beyond question, the specific reasons are unknown. Of the three hypothesised reasons for HA, namely crowd support, familiarisation and travel fatigue, two of these features can be numerically quantified. By measuring the distance that opposition teams must travel, it is possible to gauge the effects of fatigue. Similarly, by comparing the relative experience that the two teams have gained from playing at the chosen venue, it is also possible to numerically measure familiarisation. Unfortunately though, because information on crowd numbers and more specifically crowd passion is not readily available, it will always be difficult to quantify the effect due to crowd support.

4.3.2 Travel fatigue

By measuring the distance travelled by the opposing team, it is possible to tease out the negative effects due to travel. Prior to 1982, all AFL matches were played within Victoria. Since then teams have established home bases in New South Wales, Western Australia, Queensland and South Australia, creating the need for interstate travel. The most simplistic approach to quantifying the effects due to travel is to introduce a binomial variable to identify interstate travel. Overall, 13% of matches played have been between interstate opponents, although currently, approximately half of the matches played each season are between teams from differing states.

On average, the HA when opponents travel from interstate is almost double the advantage experienced when teams are from the same state (13.0±0.9 points vs. 6.8±0.4 points $p<0.0001$). The debilitating effects of interstate flights can be further quantified using a cut-off of 1500km (approximately two hours travel). Teams travelling for longer than two hours are disadvantaged by almost an additional goal, with their opponents enjoying a HA of 16.6±1.2 points compared with 11.2±1.3 points when travel is less than two hours ($p<0.0001$).

4.3.3 Ground familiarisation

The more often a side plays at a particular venue, the more familiar they become with the surroundings. In the 107 years of football, 35 different grounds have been used to host matches, although only 20 venues have been used on more than 100 occasions. By considering the difference in the number of times the two competing teams have played at the chosen venue, it is possible to numerically quantify the effects of familiarisation. Given that some teams have been using the same home venue for hundreds of matches, there is obviously a limit to just how familiar a side can become with a particular venue. To allow for this, the upper limit for experience gained at a particular venue was set at 100 matches. Whilst the difference in experience at the venue will be treated as a continuous variable, for presentation purposes the size of the differences has been categorised as small, medium and large. Table 4-1 illustrates a clear trend, with the MOV proportional to the difference in experience between the two teams.

Table 4-1 Advantage of familiarisation at a given venue

Difference in Experience	N	Average Result
Less than 10 matches difference	2811	3.6±0.8 points
Between 10 and 50 matches difference	3050	7.4±0.7 points
Greater than 50 matches difference	6601	10.1±0.5 points

Although by definition every AFL game played has a home team and an away team, 12.4% of all matches have been played on a shared or neutral venue. When this is the case, the HA is approximately halved (4.3±1.1 points vs. 8.5±0.4 points, $p<0.0001$). It could be hypothesised that when games are played at a neutral venue, the HA effects due to travel and familiarisation are removed. This suggests that the HA experienced at neutral venues (4.3±1.1 points) acts as a surrogate marker of crowd support. In reality, differences in ground familiarisation can still be shown to exist at neutral venues, indicating that ground familiarisation is a more specific measure of neutrality, thus alleviating the need for a separate variable to adjust for matches played at neutral venues.

4.3.4 Measures of performance

Various measures of team performance were considered. A simple way to gauge the performance of the competing teams is to take a moving average of past performances. An alternative approach is to give more weight to more recent performances by exponentially smoothing past results. Both moving averages and exponentially smoothing with different smoothing constants were used, with final predictions derived by subtracting the away team prediction from the home team prediction.

To measure the quality of each different predictor, two performance measurements are considered namely the percentage of winners predicted by each approach and the AAE between the predicted and actual margins.

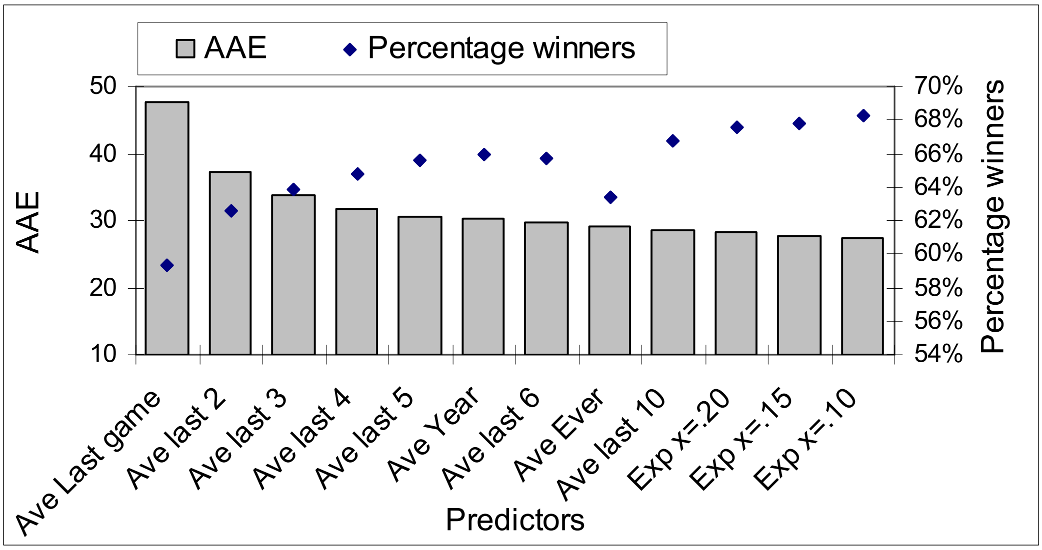

Figure 4-3 Comparison of performance predictors for all AFL matches

From Figure 4-3 it can seen that when considering all past matches, exponentially smoothed predictors produce a lower margin of error and a higher percentage of winners than do arithmetic averages. With the exception of Average Ever, a clear relationship exists between the models that produce the lowest AAE and the models that produce the highest number of winners. Despite producing a low AAE, the average of all past results was significantly worst at predicting the winner of matches. This may be because the average of all past results is slow to adjust to the considerable changes to team structure that occur on an annual basis.

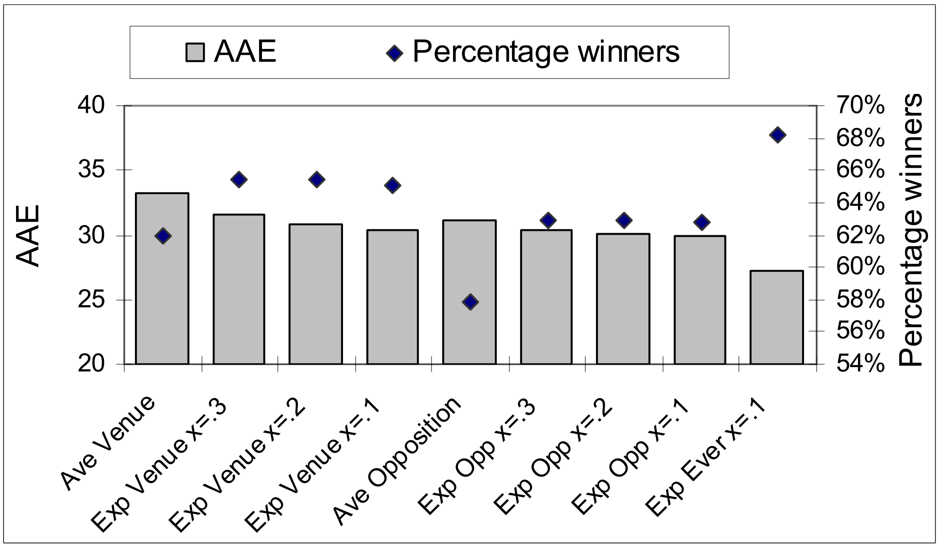

Figure 4-4 Comparison of predictors stratified by venue and opposition

By averaging and exponentially smoothing past performances at each venue and against each opposition, it is possible to derive predictors that are more specific to each game. Three different exponential smoothing parameter were considered (x=0.1, x=0.2, x=0.3). It can be seen from Figure 4-4 that the process of subsetting past data by venue or opposition is not as effective as using all past data to create a predictor. This is reflected by the fact that the exponentially smoothed predictor for all past performances predicted more winners and had a significantly lower AAE than all other approaches ($p<0.001$). With a data set of 12,462 points, a two percent difference in winners predicted is enough to show statistical significance with a p-value less than 0.001. In comparison, to show a statistically significant difference in AAE with a p-value less than 0.001 requires a difference of about one point. The corresponding AAE and percentage of winners predicted for Figure 4-3 and Figure 4-4 can be seen in Table 4.2.

Table 4-2 Average absolute error and percentage of winners predicted

Predictor	**AAE**	**Percentage of winners predicted**
Average for last game	47.7±0.3	59.3%
Average for last 2 games	37.3±0.3	62.6%
Average for last 3 games	33.7±0.3	63.9%
Average for Venue	33.2±0.2	61.9%
Average for last 4 games	31.6±0.2	64.8%
Exponential Venue x=0.3	31.5±0.2	65.4%
Average for Opposition	31.2±0.2	57.9%
Exponential Venue x=0.2	30.9±0.2	65.4%
Average for last 5	30.5±0.2	65.6%
Exponential Opp. x=0.3	30.4±0.2	62.8%
Exponential Venue x=0.1	30.3±0.2	65.1%
Average for Year	30.2±0.2	66.0%
Exponential Opp. x=0.2	30.1±0.2	62.8%
Exponential Opp. x=0.1	29.9±0.2	62.7%
Average for last 6	29.8±0.2	65.7%
Average for Ever	29.1±0.2	63.3%
Average for last 10	28.5±0.2	66.8%
Exponential Ever x=0.3	28.3±0.2	67.6%
Exponential Ever x=0.2	27.7±0.2	67.8%
Exponential Ever x=0.1	27.3±0.2	68.2%

As seen from Figure 4-2 the average score for each match of AFL has gradually risen over the 107 years of competition. Figure 4.5 shows that the AAE for an exponentially smoothed predictor (x=0.1) has also risen over this period. This decrease in predictability, specifically over the past 30 years, could well be attributed to the increased number of teams in the competition, salary cap and draft constraints as well as additional travel requirements.

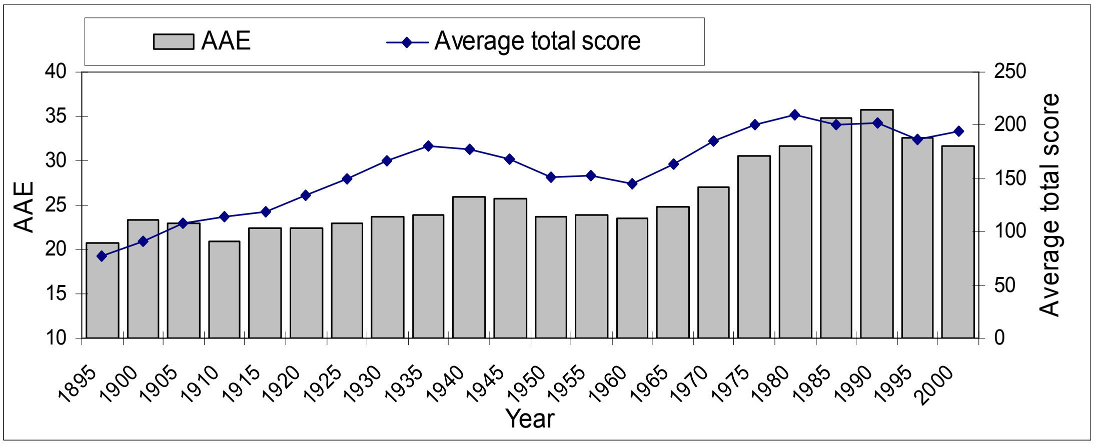

Figure 4-5 Average absolute error for exponentially smoothed predictor (x=.1)

4.4 Bookmaker prices

Bookmakers are quite efficient at predicting the winner of AFL football games, with the designated bookmaker favourite winning two thirds of matches. Because of the dynamic nature of fixed price betting markets, the bookmakers' greatest vulnerability occurs when initially setting prices. Bookmakers will traditionally post an opening market for each AFL match approximately 5-6 days prior to commencement of each game. In accordance with supply and demand, by the start of each match the bookmaker price will reflect the opinion of the general public, or more specifically, those in the general public who have placed the largest wagers on the game. Bookmaker prices for matches played between 1997 and 2003 were collected from Centrebet, on the Friday morning prior to the commencement of each round of matches.

4.5 Method

A multivariate model was constructed to predict the MOV between the home and away teams using data from all matches played prior to 1997 (11,167 games). Variables included in the multivariate model were HA, interstate travel, ground familiarisation, team quality and current form, with all variables being statistically significant with $p<0.0001$. The final model developed for MOV_T, the predicted difference between home and away teams derived at a team level was

$$MOV_T = 4.87 + 4.77I + 0.06(FT_h - FT_a) + 0.79(QT_h - QT_a) + 0.06(A2T_h - A2T_a), \qquad (4.1)$$

where the intercept (4.87), represents HA. I represents interstate travel (0, 1(<2hrs) or 2(>2hrs)). FT equals the number of matches each team has played at the venue (max=100). QT represents an exponentially smoothed predictor of team performance. $A2T$ equals the average team result for the last 2 games, with the subscripts h and a indicating the home and away teams. Details of the team model can be seen in Table 4-3.

Table 4-3 Parameter estimates for multivariate model

Variable	**Parameter Estimate**	**Partial R^2**	**P-value**
Intercept (HA)	4.87±0.56		<0.0001
Quality	0.79±0.02	28.3%	<0.0001
Interstate	4.77±0.72	0.3%	<0.0001
Form	0.06±0.01	0.3%	<0.0001
Familiarisation	0.06±0.01	0.3%	<0.0001

Although exponential smoothing produced a much better predictor of team quality (see Figure 4-3), it was interesting to note from the multivariate model that the average for the past two games also proved to be an independent predictor of outcome, thus providing a more accurate reflection of current form.

One flaw of a team performance model that is based solely on past scores is an inability to adjust for the loss of key players from within the team. Whether through injury or suspension, the loss of key players can severely reduce a team's probability of winning. Conversely, if quality players return to the team, the probability of success may increase.

One approach to compensate for individual players is to derive prediction variables at a player rather than team level. By assigning the relevant MOV to the competing players from both teams, each individual player then has a history relating only to matches in which they have played. By aggregating the predicted result for the 18 players who are named in the starting line-up for the team, it is possible to derive a team prediction that compensates for changes within the team. Starting line-ups for each team are initially named on the Thursday night prior to the weekend matches, and must be finalised 24 hours prior to the commencement of the game. Although last minute changes have been known to occur, the public is generally aware when key players are unlikely to play.

Because data at an individual player level were only available for matches played from 1997 onwards, it is impossible to derive separate parameter estimates for prediction variables using a sample of matches played prior to 1997. Since data post 1996 is used as a holdout sample for testing, an alternative approach was adopted. The form of the team model (3.1) was used; with the parameter estimates derived from the team model the same, whilst the values of the prediction variables were calculated using individual player statistics. Ground familiarisation, overall quality and current form are all calculated at an individual player level and then averaged to give a team rating. This means that the overall weighting of the five variables included in the multivariate model does not change, just that the predictor variables used become more representative of the actual players on the field. Thus we obtain the following model for MOV_I, the predicted difference between home and away teams derived at an individual player level.

$$MOV_I = 4.87 + 4.77I + 0.06(FI_h - FI_a) + 0.79(QI_h - QI_a) + 0.06(A2I_h - A2I_a.), \qquad (4.2)$$

where *FI* equals the average number of matches played at venue for the 18 starting players. *QI* equals the average exponentially smoothed predictor for the starting players, whilst *A2I* equates to the average result for the last two games for each of the starting 18 players.

Parameter estimates were applied to 1286 matches[6] played after 1996 with results compared against a benchmark of Clarke (1993) for the same period. Goodness of fit was assessed by three criteria, namely the AAE between the predicted and actual results, the percentage of winners successfully predicted and the potential ROI that could be derived from the fixed price bookmaker, Centrebet.

4.6 Results

Table 4-4 compares results from the three models. It can be seen that the model derived at an individual level produced the lowest AAE, the highest percentage of winners and the greatest ROI. Both the individual and team models produced average profits that were significantly greater than zero ($p<0.0001$).

Table 4-4 Model comparison for AAE, percentage of winners predicted and ROI

Model	**Benchmark**	**Team**	**Individual**
AAE	30.5±0.6	30.2±0.6	29.8±0.6
Percentage of winners	64.6%	65.8%	66.7%
Number of bets	923	1049	981
Total Outlay ($1000s)	157.0	203.3	159.1
Profit ($1000s)	2.0	20.4	24.0
Average Bet Size ($)	170±5	193±5	163±4
Ave. Profit per bet ($)	2±7	19±6*	25±6*
ROI	1.3%	10.1%	15.1%

[6] Individual player and bookmaker information was unavailable for the 1997 finals series

*Significantly greater than zero, p<0.0001

A statistical comparison between models can be seen from Table 4-5. With an AAE of 29.8 points per game, the individual model was significantly more accurate in predicting the MOV than both the team model (30.2, p=0.025) and the benchmark model (30.5 p=0.001). The difference between the team model and the benchmark model was bordering on significance (p=0.06). With a percentage of successfully predicted winners of 66.7%, the individual model was significantly better in predicting winners than the benchmark (64.6% p=0.02) but did not achieve statistical significance in comparison to the team model (65.8% p=0.21). Although also predicting more winners, the team model was not significantly better performed than the benchmark (65.8% vs. 64.6% p=0.20). When considering profit derived from betting on all situations in which there was perceived advantage, there was no significant difference between the team and individual models, although both models were significantly better performed than the benchmark model.

Table 4-5 Statistical comparisons between models

Outcome	**A**	**B**	**Difference (A – B)**	**P-value***
AAE	Individual	Team	0.38±0.16	0.025
AAE	Individual	Benchmark	0.72±0.25	0.001
AAE	Team	Benchmark	0.34±0.26	0.06
% winners	Individual	Team	0.9%	0.21#
% winners	Individual	Benchmark	2.1%	0.02#
% winners	Team	Benchmark	1.2%	0.20#
Ave. Profit	Individual	Team	$3±3	0.19
Ave. Profit	Individual	Benchmark	$17±5	0.0003
Ave. Profit	Team	Benchmark	$14±5	0.0007

*Wilcoxon sign rank test, #McNemar’s test for paired proportion

4.6.1 Trends over time

By exploring goodness of fit annually over the seven year period, trends can be explored. Figure 4-6 shows the annual AAE for the three prediction models from 1997 to 2003 and reflects the differences in predictability between seasons.

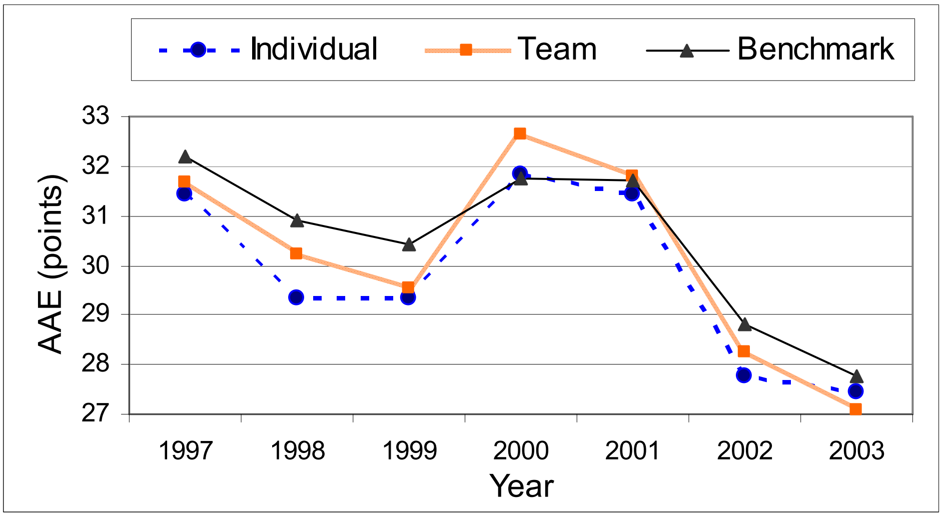

Figure 4-6 AAE for the three models from 1997 to 2003

Figure 4-7 Winning percentage from 1997 to 2003 shows the differences in the annual percentage of winners successfully predicted by each model. Once again, a high degree of variability can be seen to exist from year to year.

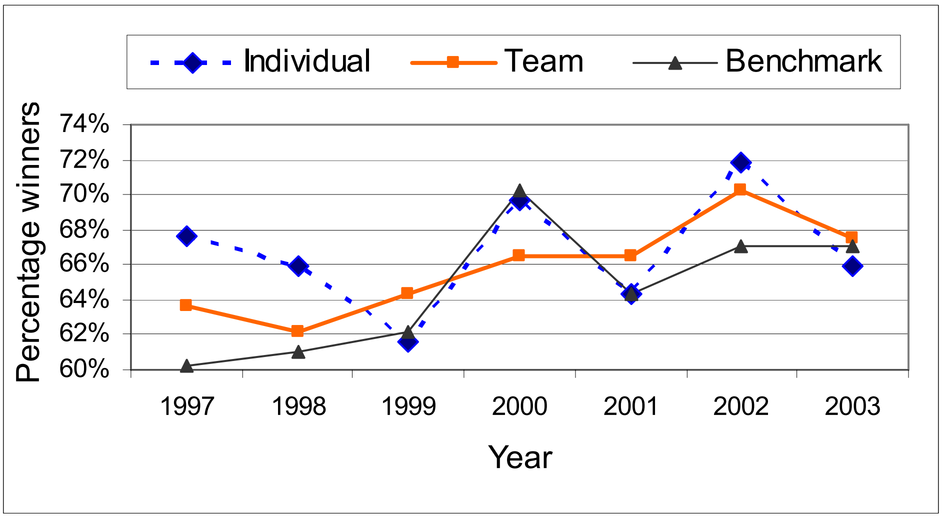

Figure 4-7 Winning Percentage from 1997 to 2003

While there is a suggestion that over the past few years AAE has gone down and the percentage of winners predicted has gone up, Figure 4-8 shows this has not equated to an increase in ROI. Despite the appearance of declining profits, both the individual and team models were able to produce a positive ROI for all seven years.

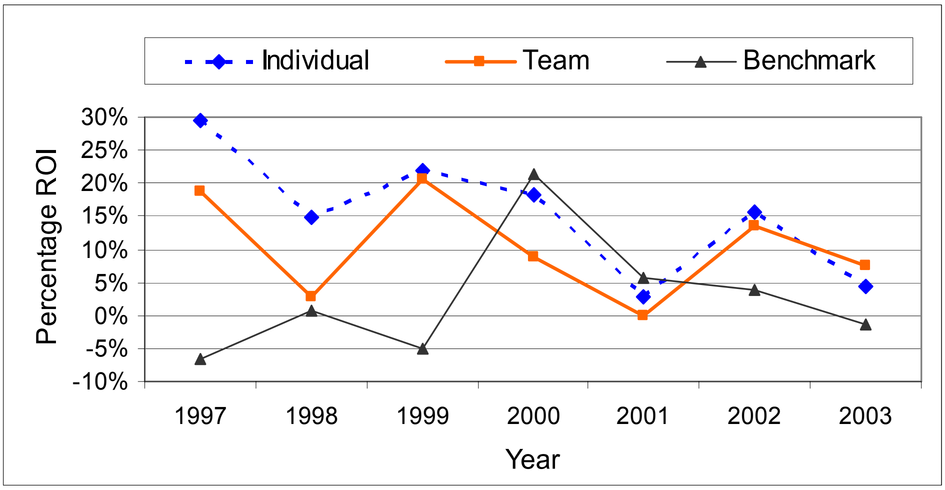

Figure 4-8 Annual Return on Investment from 1997 to 2003

4.7 Discussion

The use of multiple linear regression to identify and weight highly significant predictors of outcomes can clearly aid in the prediction of AFL matches. The use of data derived at an individual level can further benefit the prediction process. Although the bookmakers appeared to improve in their price setting processes over the past seven years, it is still possible to derive an annual profit, with statistically significant improvement coming through the use of data derived at an individual player level.

Although the individual model can been seen to produce the greatest profits, this model is dependent upon team selection and cannot be utilised until the Friday prior to the weekend's rounds of matches. Because both the team and benchmark models are based solely on past team scores, predictions for these models can be produced immediately after the last round has finished and can be available to use when markets are posted early in the week. As bookmakers are most vulnerable when prices are initially posted, it is realistic to assume that greater profit could be derived for both of these models by placing bets earlier in the week.

In addition there are many betting strategies that can be employed. Variations such as betting only when the advantage is at least some pre determined figure, betting only on some rounds in the season, betting only on favourites, are some of the strategies employed by punters. These all have the possibility of increasing returns or reducing risk. In addition, this chapter has only investigated head to head betting. Evidence suggests that mathematical models are relatively better than the media experts (and thus possibly the bookmakers and the public) at selecting more complicated outcomes such as margins and final ladder order. Inefficiencies in these markets may also be open for profit by using statistically derived prediction approaches.

Having established a feasible way to use past data to predict the outcome of matches in AFL football, the logical progression would be to ask if such a process could then be applied to predict the outcome of other sports. A sufficiently large amount of past data and an underlying assumption that the outcome of interest is normally distributed would appear to be the only two prerequisites required to facilitate this process.

5 Predicting the match result in ODI cricket

5.1 Introduction

While ODI cricket is a vastly different game from AFL football, there are enough similarities to suggest that the outcome of both sports can be statistically modelled in the same way. In AFL football the average score per team is about 100 points, while in ODI cricket, the average number of runs scored per team is in excess of 200. The high scoring structure for each sport is sensitive enough to ensure that when the winning margin is referenced with respect to the home team, the resulting MOV can be well approximated by a Normal distribution. This underlying assumption of normality facilitates the use of a multiple linear regression to predict MOV.

Using match information gathered from all 1800 ODI matches played prior to January 2002, a multiple linear regression model was used to identify and numerically weight features that could independently explain statistically significant proportions of variation associated with the outcome of ODI matches. Prediction models combining measures of experience, quality and HA were constructed at both a team and individual level with the resulting prediction model applied to the 336 matches played between January 2002 and July 2004. Predicted probabilities were compared with bookmaker prices and the efficiency of ODI betting markets was explored.

5.2 Background

Prior to July 2004, 17 countries had played 2136 completed ODI matches, although 82% of all matches have been played by eight main cricketing nations (Australia, England, India, Pakistan, West Indies, Sri Lanka, New Zealand, and South Africa). Just over 1200 cricketers have represented their country in ODI matches.

In ODI cricket the aim of the team batting first is to score as many runs as possible in the allotted time (usually 50 six ball overs). If the first team scores more runs than the second team, the MOV can readily be expressed in terms of runs difference between the two teams. The aim of the side batting second is to score more runs than the first team. Because the game is deemed to be finished if the team batting second achieves their target, the MOV is usually expressed in terms of resources (wickets and balls) remaining, rather than runs. In order to develop a predictive process for match outcomes, a consistent measure of the MOV is required. This can be achieved by following the work of Duckworth and Lewis to convert resources available into runs.

Frank Duckworth and Tony Lewis invented a now well-known system for resetting targets in ODI matches that were shortened due to rain. Although this system has undergone several refinements in recent years, the general way in which the Duckworth-Lewis (D-L) method is calculated has not changed, with wickets and balls remaining expressed as resources available and converted to runs. Table 5-1 shows an abbreviated version of the remained resources (R) for wickets lost and balls remaining. A complete tables and detailed account of the derivation of this table is given by Duckworth and Lewis (1998).

Table 5-1 Percentage of resources available for overs remaining and wickets lost

Overs remaining	**Wickets lost**									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
50	100.0	93.4	85.1	74.9	62.7	49.0	34.9	22.0	11.9	4.7
40	89.3	84.2	77.8	69.6	59.5	47.6	34.6	22.0	11.9	4.7
30	75.1	71.8	67.3	61.6	54.1	44.7	33.6	21.8	11.9	4.7
25	66.5	63.9	60.5	56.0	50	42.2	32.6	21.6	11.9	4.7
20	56.6	54.8	52.4	49.1	44.6	38.6	30.8	21.2	11.9	4.7
15	45.2	44.1	42.6	40.5	37.6	33.5	27.8	20.2	11.8	4.7
10	32.1	31.6	30.8	29.8	28.3	26.1	22.8	17.9	11.4	4.7
5	17.2	17.0	16.8	16.5	16.1	15.4	14.3	12.5	9.4	4.6
1	3.6	3.6	3.6	3.6	3.6	3.5	3.5	3.4	3.2	2.5

Whilst the D-L approach was specifically designed to improve 'fairness' in interrupted one-day matches, de Silva, Pond & Swartz (2001) found that when used to quantify the MOV, the D-L approach sometimes overestimated the available resources when the second team to bat won easily, and underestimated the available resources when the second team to bat only just won. Whilst de Silva used almost 800 ODI matches to derive this result, further confirmation of this bias can be achieved by considering all past ODI matches play prior to July 2004.

Of the 2064 ODI matches in which there was a decisive winner, the side batting first won the game 50.1% of the time. Because this result is not significantly different from 50%, we can assume there is little evidence to suggest that batting order plays a role in determining victory. Thus, when comparing MOV between the teams batting first and teams batting second, no systematic difference should be present.

When an ODI match is won by the team batting first, the MOV is readily determined by the difference in runs scored. When the match is won by the team batting second, the MOV can be found by multiplying the first innings run total by the corresponding percentage of resources remaining (see Table 5-1). When using the D-L approach, it can be seen from Table 5-2 that clear bias exists, with the average MOV for the team batting second almost 15 runs higher than that of the team batting first.

Table 5-2 Comparison of MOV using only Duckworth and Lewis

Batting Order	N	**Average MOV**	**Standard Deviation**	Median	**Wilcoxon Rank Sum**
First	1035	56.5±1.5	47.2	46	P< 0.0001
Second	1029	71.3±2.6	84.4	43	

By minimizing the Cramer-von Mises statistic for the differences between actual and predicted runs, de Silva derived a formula to reduce bias by modifying the remaining resources. This is given by

$$R_{mod} = (1.183 - 0.006R)R \tag{4.1}$$

where R_{mod} = modified resources and R=resources given using D-L (see Table 5-1).

By using modified resources to calculate MOV, it can be seen in Table 5-3 that the average MOV between the first and second inning was no longer statistically significant, with a difference of only 1.5 runs. This suggests that de Silva's adjustment of D-L resources provides an unbiased approach to determining MOV.

Table 5-3 Comparison of MOV between innings using de Silva's adjustment

Batting Order	N	Average MOV	**Standard Deviation**	Median	**Wilcoxon Rank Sum**
First	1035	56.5±1.5	47.2	46.0	P=0.46
Second	1029	58.0±1.4	44.7	46.0	

Using de Silva's adjustment to modify resources, the MOV for each ODI can then be referenced with respect to the home team. Based on the finding from chapter 4 that HA is partially attributed to familiarisation, for matches played in neutral countries, the home team was deemed to be team that had played the most matches in the host country. From Figure 5-1 it can be seen that the distribution of MOV when referenced against the home team, can be well approximated by a Normal distribution, thus facilitating the use of multiple linear regression to develop prediction models.

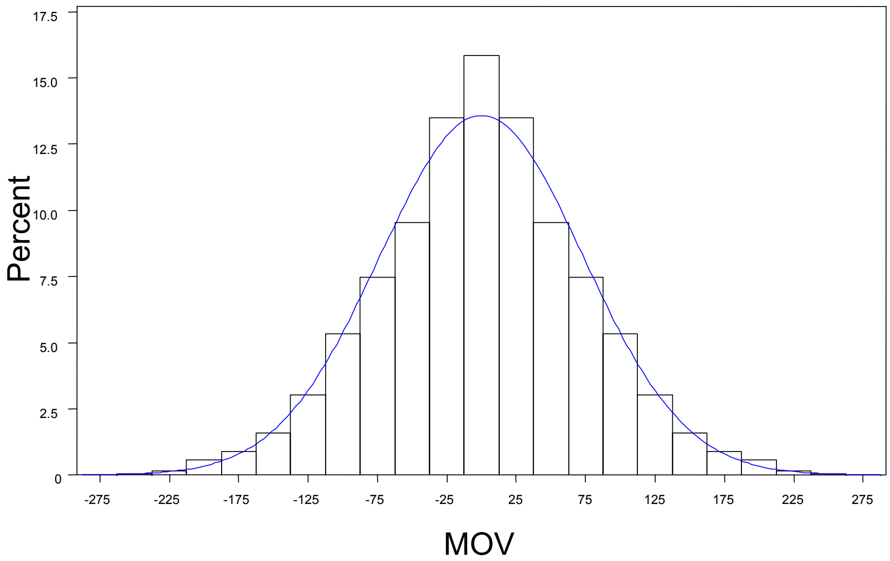

Figure 5-1 Histogram of MOV referenced against the home team

As individual player information was shown to improve prediction models for AFL football, individual player information was once again used for ODI cricket matches, although not in exactly the same fashion. Because individual player data for AFL football was not available for all matches played, parameter estimates had to be derived for a team level, with individual player information then used to improve the sensitivity of the predictors in the holdout sample. Because individual player data was available for all ODI matches played, parameter estimates for ODI cricket were derived separately for team and individual models.

Measures of experience, quality and form were created in the same fashion as for AFL football, by differencing the predictors for the two individual teams. HA was once again explored to ascertain if additional benefit could be gleaned by differentiating between distance travelled and familiarisation.

5.3 Predictors of MOV for ODI cricket

5.3.1 Home country advantage

Because ODI tournaments are often played in triangular or round robin formats, only 66% of the 2064 resulting ODIs prior to July 2004, were played on the home soil of one of the two competing countries, with the home team winning 58% of matches. Since ODI cricket matches are played between countries, the advantage that a team may have by playing within their own country effectively equates to HA.

By ignoring matches played at neutral venues, and averaging the MOV in terms of the home team, a quantitative measure of HA can be found that equates to 11.4±1.9 runs per game which is highly significantly different from zero ($p<0.0001$). Despite confirming the presence of HA, this measure fails to take in to account a class difference that existing in ODI cricket.

A greater insight into HA can be achieved by considering the leading eight established cricketing nations (Australia, West Indies, England, India, Pakistan, New Zealand, South Africa and Sri Lanka) separately from the developing cricketing nations (Zimbabwe, Bangladesh, Namibia, United Arab Emirates (UAE), Netherlands, Canada, Kenya and Scotland).

5.3.2 Class structure

By separating the home and away teams with regards to whether they come from an established or developing cricket nation, the full effects of a class divide become apparent. Table 5-4 shows that when both the home and away teams are from established cricketing nations the average MOV is equal to 14.5 runs. When an established nation plays host to a developing nation, the average MOV is equal to 63.1 runs, but when the opposite occurs – an established nation travels to a developing cricketing nation, the average MOV in favour of the home team is equal to minus 66 runs! Given the magnitude of the difference in class between established and developing cricketing nations, the relative strength of the two competing teams appears to be of much greater importance than HA.

Table 5-4 Average MOV between classes for non-neutral matches

Home Team	Away Team	N	Average MOV for home team
Established	Established	1126	14.5±2.0
Established	Developing	103	63.1±6.5
Developing	Established	111	-66.0±6.3
Developing	Developing	23	6.5±13.9

As previously discussed, it has been hypothesized that the effects of HA can be attributed to travel fatigue, ground familiarization and crowd support. Like AFL football, crowd support for ODI cricket is difficult to quantify, but it is possible to numerically quantify distance travelled and familiarization.

5.3.3 Distance travelled

By measuring the distance travelled by the away team to the host country it is possible to explore the potential effects due to fatigue. To avoid the bias associated with matches involving the developing nations, only established cricketing countries were considered. Distances were measured in thousands of kilometres, and as seen in Table 5-5 have been stratified into blocks of 5000.

Table 5-5 Average MOV for distance travelled

Distance Travelled	N	**Average MOV**
0 – 5000km	186	6.8±3.1
5000 – 10,000km	432	21.5±2.5
10,000 – 15,000km	289	15.0±3.6
>15,000km	213	6.1±4.3

From Table 5-5 it can be seen that while there do appear to be differences in the average MOV depending on the distance travelled, there is no apparent linear effect. This is of little surprise as ODI games are generally played in a series of matches, with the travelling team spending several weeks in the host country. Once a series of matches has commenced, the travel commitments of both teams become similar in that both must travel throughout the host country to compete in matches. So, whilst the 'away' team may well take a few days to acclimatise to the host country at the commencement of a tour, any debilitating effects due specifically to the distance travelled to get to the venue become negligent as they are similar for both teams.

5.3.4 Familiarization in host country

ODI matches have been played in 20 different countries. From Table 5-6 it can be seen that Australia is the leading host nation having hosted 424 ODI matches. Two competing cricketing countries have visited Australia for a triangular ODI competition every year since 1979. Given the travelling distance required to get to Australia, this reflects both the passion for cricket within the country and the organizations skills of the Australia Cricket Board. India and England have hosted the second and third most matches with 230 and 221 respectively. Although only actually competing in seven ODIs, because of its central location to teams on the sub-continent, the United Arab Eremites (UAE) has played host to 198 ODI matches, all held at the same venue.

Table 5-6 Number of matches played and hosted by each country

Division	**Country**	**Played**	**Hosted**	**Neutral matches**	**Venues used**	**Main venues**
Established	Pakistan	570	157	17.8%	16	6
Established	Australia	554	424	32.5%	16	6
Established	India	535	230	23.9%	31	9
Established	West Indies	482	121	1.7%	11	5
Established	New Zealand	437	177	5.6%	10	6
Established	Sri Lanka	423	134	23.9%	6	3
Established	England	379	221	38.9%	14	7
Established	South Africa	303	168	31.0%	11	7
Developing	Zimbabwe	256	83	3.6%	5	2
Developing	Bangladesh	86	55	45.5%	2	1
Developing	Kenya	62	42	50.0%	4	1
Developing	Netherlands	13	1	100.0%	1	1
Developing	Canada	9	21	100.0%	1	1
Developing	UAE	7	198	99.0%	1	1
Developing	Namibia	6	5	80.0%	1	1
Developing	Scotland	5	2	100.0%	1	1
Developing	East Africa	3	0	0.0%	0	0
Developing	Singapore	0	12	100.0%	2	1
Developing	Morocco	0	7	100.0%	1	1
Developing	Wales	0	5	100.0%	2	1
Developing	Ireland	0	1	100.0%	1	1

In cricketing terms, familiarization can be viewed at two levels – how often teams have played within a given country or how often teams have played at a given ground within each country.

Although ground conditions can change from one venue to the next within a county, it is generally felt that each country has general conditions and pitch preparation philosophies that are unique. Looking firstly at a within country level, there is clear evidence to suggest that the more often a side plays within a given country, the more familiar they become. It follows that the greater the difference in familiarity between the two teams, the greater the MOV. Table 5-7 shows that when the two competing teams have less than 10 matches difference in experience within the host country, the average MOV is less than zero. When the difference in experience is between 10 matches and 50 matches, the average MOV equates to about 14 runs, whilst if the host team has more than 50 matches experience in comparison to their opponents, the average MOV is worth about 20 runs.

Table 5-7 Familiarization within a given country

Difference in experience	**N**	**Average MOV**
Less than 10 matches	570	-3.0±3.3
Between 10 and 50 matches	836	13.9±2.4
50 or more matches	658	20.5±2.7

Because the nucleus of cricket teams will change over time, it makes intuitive sense when considering familiarization, to model data at an individual player rather than team level. By averaging the number of times that each of the 11 competing players has played in a country it is possible to derive a measure of the difference in experience that is potentially more sensitive.

Like most sports played at an elite level, there are limitations as to how long a cricketer can play for. With only 11 active positions available in the team and extensive competition for those positions, it is unusual for a cricketer's career to extend much beyond his mid thirties. As a consequence, when modelling experience at a player level the differences in experience/familiarization within each country are not as great as those observed at a team level.

Once again using a 10 match difference as the cut-off, it can be seen from Table 5-8 that if the home team averages more than ten games experience over their opposition, the MOV equates to 22.5 runs, whereas if the difference in experience between the two teams is less than ten games then the average MOV is only 1.9 runs. This would seem intuitive as information at a player level should be more sensitive to measuring experience.

Table 5-8 Familiarization within a given country at an individual player level

Difference in Experience	N	Average MOV
Less than 10 matches	1128	1.9±2.2
10 or more matches	936	22.5±2.3

5.3.5 Familiarization at the venue

ODI matches have been played at a total of 137 different venues with most established cricketing nations having at least five venues that are used on a regular basis (see Table 5-6). Only 20 venues world wide have hosted 30 or more ODI matches, with 55 venues hosting 10 or more matches. When considering familiarity specifically at the chosen venue, if neither team has played at the venue, or both teams have played at the venue an equal number of times then the HA is only 2.1 runs. From Table 5-9 it can be seen that although there is a clear advantage in having played at a venue more times than the opposition, this effect does not have a strong linear trend as the average MOV for less than 10 matches is equal to 12.4 runs, whilst the average MOV for more than 10 matches difference is only 13.9 runs.

Table 5-9 Advantage of familiarization at a given ground

Difference in Experience	N	Average MOV
No difference	323	2.1±4.3
Less than 10 matches	1043	12.4±2.2
10 or more matches	698	13.9±2.8

When considering familiarization at a given ground reduced to an individual player level (Table 5-10), an average difference in experience between the two teams of less than 10 games equates to an average MOV of 12.8 runs, whilst a difference in averages greater than 10 equates to an average MOV of 23.5 runs. It should be noted that a difference in experience greater than 10 has only occurred on 142 occasions (7%).

Table 5-10 Ground familiarization determined at an individual player level

Difference in Experience	N	Average MOV
No difference	362	0.0±3.9
Less than 10 matches	1560	12.8±1.8
10 or more matches	142	23.5±6.1

Although both familiarization within a country and at a specific venue has been stratified for the purposes of presentation, all predictors relating to experience can be treated as continuous variables.

Using an R-square statistic to reflect variation explained, from Table 5-7 it can be seen that the difference in experience within a country acts as a better predictor of familiarization than the difference in experience at the specific venue. This could be attributed to the large number of venues that have been used and the infrequency in which games are played on some of those venues. Similarly, experience measured at an individual player level can explain significantly more variation in MOV than experience measured at a team level.

Table 5-11 Comparison of continuous predictors for familiarization

Variable	**Parameter**	**R-square**	**P-value**
Country – Team	0.34±0.05	2.0%	<0.0001
Country – Individual	1.21±0.15	3.0%	<0.0001
Venue – Team	0.33±0.10	0.5%	0.001
Venue – Individual	0.90±0.39	0.3%	0.02

5.3.6 Experience

Experience can be directly measured by the number of previous matches played. It will be shown in subsequent chapters that experience is a strong predictor of individual performance in both AFL football and ODI cricket (chapters 6&7). When considering the difference in experience as a predictor of outcome, it can be seen from Table 5-6 that all established cricketing nations have played in excess of 300 ODI matches each, whilst most of the developing cricketing countries have played less than 50 ODI games. Zimbabwe, although playing over 250 matches has not been able to make the transition from a developing to an established nation possibly due to political instability which in turn has had a destabilizing effect on the national cricket team.

Although the difference in experience measured at a team level is a statistically significant predictor of MOV ($p<0.0001$), this significance may primarily be driven by the difference between established and developing cricketing countries. When experience is measured by averaging the 11 individual members of both teams, the resulting difference in experience is vastly better than the team constructed predictor (see Table 5-12). Over 8% of the variation associated with MOV can be explained by a difference in experience at an individual level, with each additional match difference equating to about half a run.

Table 5-12 Difference in experience derived at both a team and individual level

Variable	Parameter	R-square	P-value
Experience – Team	0.07±0.01	2.0%	<.0001
Experience – Individual	0.55±0.04	8.3%	<.0001

5.3.7 Experience against the opposition

Having considered overall experience, experience within a given country and experience at each venue, let us now consider experience against the opposition. At a team level, experience against the opposition is pointless as both teams will have met an equal number of times. At an individual level this variable is statistically significant. Although only explaining 0.6% of the variation associated with MOV, experience against the opposition derived at an individual level, is statistically significant (p=0.0005) with each additional games experience equating to an advantage of 1.6±0.5 runs.

5.3.8 Measures of performance

Performance can be considered at two levels: current form and overall quality. Moving averages with differing denominators were used to measure current form whilst past averages and exponential smoothing with three different smoothing parameters (x=0.1, x=0.2, x=0.3) were used to create each team's measure of overall quality. Final predictions of quality and form were created by subtracting the away team's predictions from the home team's predictions. To measure the worth of each prediction approach, two performance measurements are considered, namely the percentage of winners predicted by each approach and the AAE between predicted and actual margins. Prediction variables were created at both an individual level and at a team level.

From Figure 5-2 it can be seen that the individual model produced lower AAEs for the average of recent games. Whilst information derived at an individual level was more sensitive than the team model for measures of current form, there was no difference between approaches for long term measures of quality.

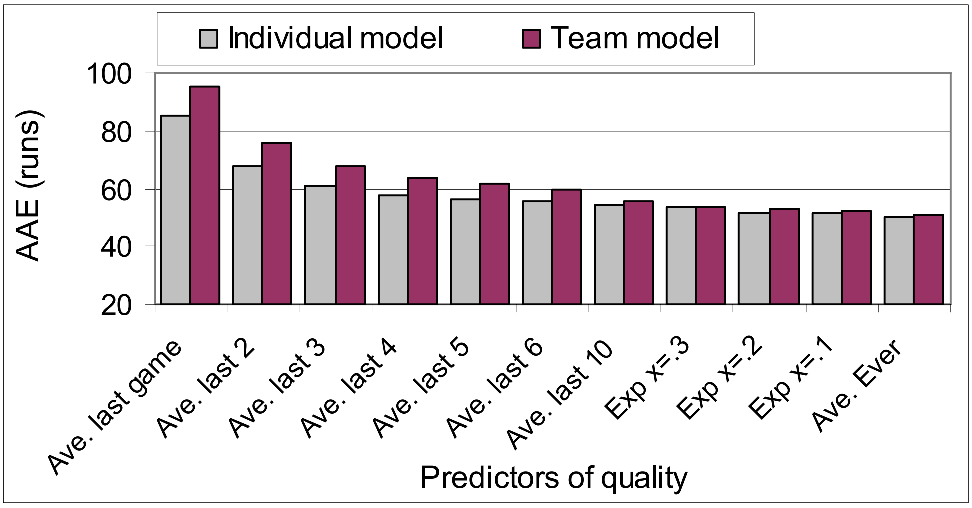

Figure 5-2 Team and individual predictors for quality using the AAE

Because AAE reflects goodness of fit, the model that produces the lowest AAE should also be the most accurate when predicting winners. From Figure 5-3 it can be seen that the greater the number of past matches used to create a prediction variable, the better the result. This is reflected by a reduction in the AAE and a corresponding increase in the proportion of winners correctly identified. Whilst this relationship is consistent for both AFL football and ODI cricket, unlike predicting the outcome of AFL football matches, for ODI cricket, the past average of all previous matches appears as a slightly more accurate prediction method than exponential smoothing.

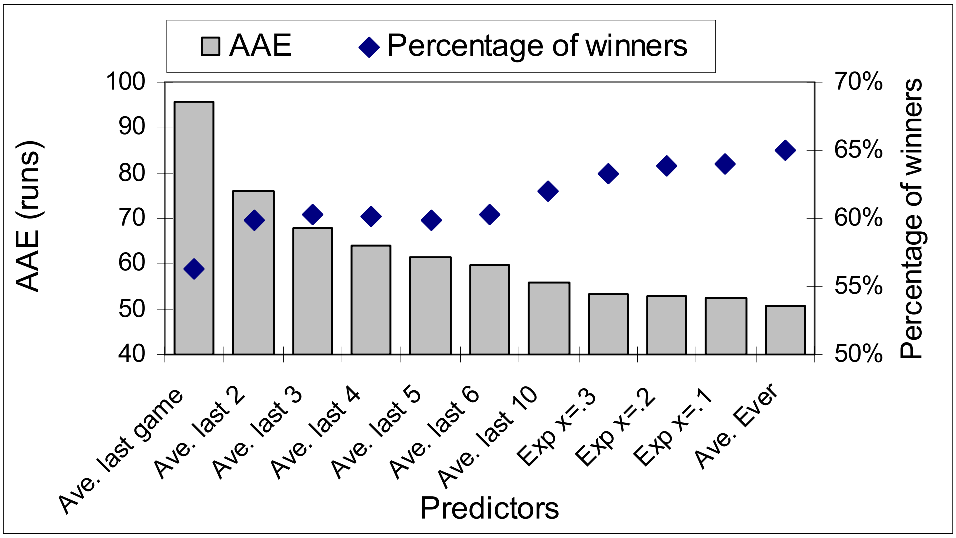

Figure 5-3 AAE and percentage of winners for predictors of quality and form

5.3.9 Specific measures of performance

Past average and exponential smoothing provide a process to measure quality and form. Considering past average and smoothing parameters for performances at specific countries, at specific venues and against specific oppositions, provides an alternative process that combines aspects of form and familiarly. By considering the AAE for variables derived at a team and individual level, comparisons can be made. From Figure 5-4, smaller AAEs suggest that variables relating to performance at a given venue or within a given country are more accurately defined at an individual level. Conversely, variables relating to the opposition are better determined at a team level. Regardless of the level of stratification, the prediction variable producing the lowest AAE was the average of all past results. With over 2000 matches of data available, a difference in AAE of about 4 runs is enough to show a statistically significant difference with a p-value less than 0.001. From Figure 5-4 it can be seen that the average of all past results is significantly better than most other approaches.

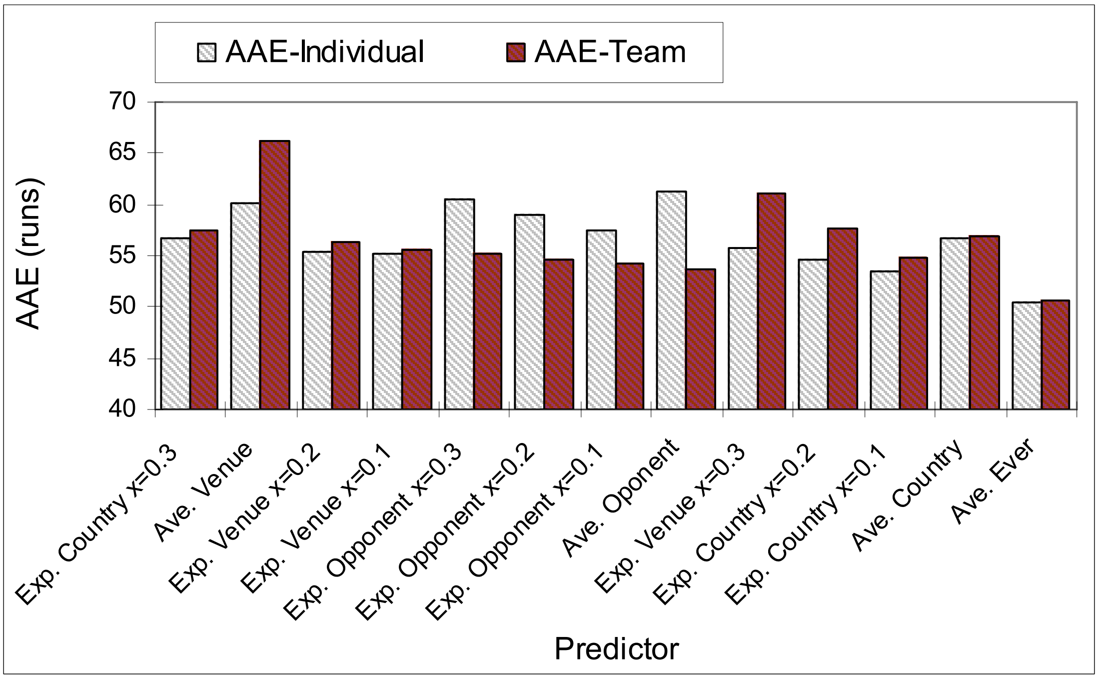

Figure 5-4 Prediction approaches used over different strata

5.4 Multivariate analysis

Two multivariate prediction models were constructed using match and player information from all 1800 ODIs played prior to Jan 2002. The first model was built with information gathered at a team level, whilst the second model was created with information that was gathered at an individual player level and then aggregated to a team level. Unlike AFL football, complete player information was available for all matches played ensuring that completely separate models could be created at both the team and individual levels.

Although selection techniques were used to identify and rank potential predictors, the multivariate models were constructed with a specific aim in mind: to combine measures of recent form, experience, overall quality and HA. Prediction variables of experience, quality and form were derived by developing separate measures for both teams and then subtracting the away team values from the home team values. This effectively references the final result in term of the home team. Indicator variables were created to identify matches played at a neutral venue and matches where the two competing teams were clearly from different class structures (established nation versus developing nation). From Table 5-13 the resulting multivariate models constructed at both a team and individual level can be seen.

Table 5-13 Multivariate models constructed at a team and individual level

Variable	Team model			Individual model		
	Estimate	**P-value**	**Partial R^2**	**Estimate**	**P-value**	**Partial R^2**
Intercept / HA	13.4±1.9	<0.0001		12.9±1.9	<0.0001	
Average Ever	0.6±0.1	<0.0001	17.3%	0.5±0.1	<0.0001	17.3%
Class*	-29.6±6.7	<0.0001	1.2%	-31.3±6.6	<0.0001	1.3%
Experience	0.2±0.1	0.002	0.4%	0.3±0.1	<0.0001	0.6%
Ave. last 10	0.1±0.04	0.003	0.4%	0.2±0.07	0.002	0.5%
Neutral Venue	-8.6±3.2	0.007	0.3%	-9.0±3.2	0.005	0.4%

* When a developing cricket nation played host to an established cricket nation

Because the MOV in the regression model is nominally structured in favour of the home team, the intercept term in the regression equation reflects HA. It can be seen from Table 5-13 that HA for both the team and individual model is equivalent to about 13 runs and is highly statistically significant ($p<0.0001$). Because one third of all ODI have been played at neutral venues, a binomial indicator variable was imperative to negate the HA for these games. If all matches played at neutral venues were devoid of HA then the binomial variable for a neutral venue would be the exact negative of the intercept term. This was not the quite the case, with the neutral variable equivalent to about nine runs in both models, suggesting a HA in neutral matches equivalent to about four runs.

When modelling the outcome of AFL football matches, it was found that although matches were supposedly played at neutral venues; if one side had played at the venue more often, they had increased familiarity and thus a slight advantage. To account for this in ODI matches, when teams were competing at a neutral venue, the side with the greatest amount of experience within the chosen country was assigned to be the home team. In this fashion it was hoped that the difference in familiarity between the two teams would act as continuous predictor that would supersede the binomial variable used to identify neutral venue, however this was not the case, as the binomial variable for a neutral venue proved to be more significant for both the team and individual models. The resulting HA for matches played at neutral venues (about four runs) could thus be thought of as a surrogate marker for the difference in familiarity between the competing teams.

Unlike the multivariate models constructed for AFL football, HA for ODIs could also not be statistically reflected by the distance travelled. As previously hypothesized, this is probably due to the fact that when travelling abroad, teams tend to play blocks of matches rather than individual games, thus reducing the debilitating effects due to travel.

Because of the vast difference in quality between established and developing countries, appropriate measures were required within the multivariate model to compensate for these mismatches in class. Of the four home/away case scenarios, (establish/establish, establish/develop, develop/establish & develop/develop) three dummy variables were created to allow for class differences. In the multivariate model, only one of these dummy variables was statistically significant - when a developing team was playing host to an already established team. In this scenario, the predicted MOV in favour of the home team was reduced by 29.6±6.6 runs for team model and 31.3±6.7 runs for the individual model. The primary reason as to why only one dummy variable was required was because other class imbalances were effectively accounted for by the inclusion of a difference in overall quality and experience.

The difference in quality, as measured by the difference in averages between the two teams for all past matches, was by far the strongest predictor, explaining 17.3% of the variation in both the individual and team models. The best measure of current form was the difference in averages for the past 10 matches, whilst the difference in overall experience (games played) between the home and away team was also statistically significant. To increase the robustness of the prediction model a reduced level of statistical significance was incorporated with all variables achieving a level of significance below p=0.01.

Overall, the team model could explain 19.6% of the variation associated with MOV, whereas the individual model could explain 20.1% of the variation explained. Whilst the difference between the two models does not appear to be great, this difference is statistically significant (p<.001). The greatest benefit of modelling data at an individual player level was the improvement in quantifying experience and current form.

5.5 Results

The resulting parameter estimates developed from the training data were applied to the 336 matched played between February 2002 and July 2004 to create predictions of match outcomes. By dividing the predicted margin by its standard error and comparing with a standard Normal distribution, each of the two competing teams could be assigned a probability of winning the match. Where predicted probabilities were found to exceed the inverse of the corresponding bookmaker prices, bets were placed. Bet sizes were determined in accordance with the perceived size of the advantage, the predicted probability of winning and a fixed bank size of $1000 as detailed in section 3.10. From Table 5-14 it can be seen that very little practical difference could be found between the model constructed at a team level and the model constructed at an individual player level. Both models found potential market inefficiencies in approximately two of every three matches. Although the team model was slightly better performed with regards to the AAE (54.6±0.9 vs. 54.8±0.9) and the percentage of winner successfully identified (69.6% vs. 69.0%), it was the individual model that produced the highest ROI. Although the average profit per bet for both models was in excess of $20, neither could be shown to be significantly greater than zero, with the individual model's profit per bet of $27±15 closing on statistical significance with a p-value of 0.07. This lack of definitive statistical evidence undoubtedly reflects the need for a large data set to establish that any given market imbalance is consistently prolonged enough to produce a ROI that is beyond chance. Whilst the benefit of a statistical model is clear, results of this nature provide a catch 22 situation. Obvious imbalances can be seen to exist in new betting markets, but in the three years that it would take to collect sufficient data to show that potential profit is beyond chance, the market has matured and the profits may no longer exist!

Table 5-14 Model comparison for AAE, percentage of winners and ROI

Model	**Team**	**Individual**
AAE	54.6±0.9	54.8±0.9
% winners	69.6 %	69.0 %
Number of bets	209	224
Total outlay	\$31,500	\$36,000
Profit	\$4260	\$6080
Average bet size	\$151±9	\$161±9
Average profit per bet	\$20±14	\$27±15
Probability that profit >0	p=0.14	p=0.07
ROI	13.5%	16.9%

5.6 Discussion

Despite the fundamental differences that exist between Australian Rules football and ODI cricket, the outcomes from both sports can be modelled in a similar fashion using multivariate analysis. When comparing the two, the underlying assumption of normality facilitates the use of an R-square statistic to measure predictability.

In ODI cricket, about 20% of the variation can be explained through the use of multivariate modelling. In comparison, 28% of the variation in AFL results could be explained. While it might appear that AFL football is a more predictable sport, when modelled over the last 30 years only, only 22% of the variation in AFL results could be explained using a multivariate model, suggesting little difference in predictability between ODI cricket and AFL football.

Several other parallels exist between the models used for AFL football and cricket. In both cases, an overall measure of team quality was the most significant predictor, HA was clearly present in both models, and a measure of current form could be shown to be highly significant for both. To find such commonalities in two sports that are so fundamentally different, suggests that a statistical prediction approach of this nature may well be applicable to other sports. Provided an outcome could be shown to be approximated by a normal distribution, and sufficient past data was available, a statistical approach would be beneficial.

Basketball is another sport where the high scoring nature of the game facilitates a normally distributed outcome. Using information from 5000 NBA basketball games played between October 2000 and July 2004, a multivariate prediction model was constructed. While parameters of HA, form and quality were all significant predictors for basketball ($p<0.0001$), an overall R-square figure of 15% clearly suggests basketball to be a less predicable sport than both AFL football and ODI cricket. In comparison, when bookmaker prices were used to predict NBA outcomes, they could explain 17.5%, suggesting Las Vegas bookmakers to be significantly more efficient than a simple mathematical model driven solely by team name and venue. While individual player data was unavailable, the fact that the model derived at a team level could not produce a positive ROI suggests a simple statistical model is unlikely to be find inefficiency in a well established betting market.

In AFL football two components of HA, namely travel and ground familiarisation, could be distinguished. While HA was still highly significant in ODI cricket, a travel component could not be recognised. Because ODI cricket is played between countries, the travel and scheduling requirements are not as consistent or as readily definable as those found for within country sports such as AFL football.

Perhaps surprisingly, ground familiarity could also not be recognised as a significant predictor for MOV in ODI cricket. This may reflect a greater degree of conformity in ODI pitches around the world or may simply be because overall experience is an independent variable in the ODI models and a high degree of correlation exists between overall experience and experience at a particular venue.

The added benefit obtained from modelling data at an individual player level was small but statistically significant, for both AFL football and cricket. Whilst this result offers encouragement it must be weighed against the added time constraints imposed by the individual models. The need to know the specific personal of the competing teams means that market assessments cannot be made until teams have been named. Because bookmakers will balance books in accordance with supply and demand, often the most glaring market imbalances will disappear by the time teams are named.

While the benefits of statistically modelling match outcomes can clearly be observed, a multivariate approach is better suited to modelling a more direct outcome. Because the MOV is dependent upon the scores from the two competing teams, a successful modelling process must incorporate features from both teams. This increase in complexity makes it more difficult to identify statistically significant predictors, resulting in only a handful of predictors being identified for MOV of both AFL football and ODI cricket. In comparison, individual performances can be modelled using features that impact directly upon the individual. As a result, a greater number of statistically significant predictors are identifiable when predicting independent outcomes. This will be borne out in the proceeding chapters.

6 Predicting individual player performance in AFL football

6.1 Introduction

Whilst only seven years of AFL data was available at an individual level, with 185 matches per year and 44 players per game, this equated to over 8000 individual data points per year. Data of this magnitude readily ensure that features thought to affect player performance can be identified.

Using individual player information gathered from AFL matches played between 1997 and 2000, 34 individual predictors for player performance were identified and compared using linear regression. In addition, a 15-parameter multiple linear regression model was used to identify and numerically weight features that could independently explain statistically significant proportions of variation associated with the number of possessions that each player would gather throughout the course of an AFL game. Using the predicted number of possessions, leading players from competing teams were matched to determine the probability that one player would gather more possessions than the other. Resulting probabilities were compared with 1597 bookmaker prices collected between 2001 and 2003 and the efficiency of the player HtH betting market was explored.

6.2 Background

Although Australian Rules football is essentially a team game, there has always been an interest in the performance of specific players. There are numerous ways in which to measure and grade the performance of specific players, with the ultimate award for individual performance given by the Brownlow medal. This will be discussed in greater detail in chapter 8.

One of the bet types introduced by sports bookmakers in recent years is to match leading players from competing teams with punters betting on who will have a superior performance (HtH betting). Although player performance can be measured in various ways, for the purpose of HtH betting the outcome is determined solely by the number of disposals that each player gathers throughout the course of the match. When a player gathers possession of the football, to avoid penalty he must correctly dispose of the ball by hand or foot. The resulting handball or kick is referred to as either a possession or disposal.

Like match outcomes for AFL football, the number of disposals gathered by leading players during the course of the game can be reasonably well approximated by a Normal distribution. This can be seen from a histogram of player disposals in Figure 6-1.

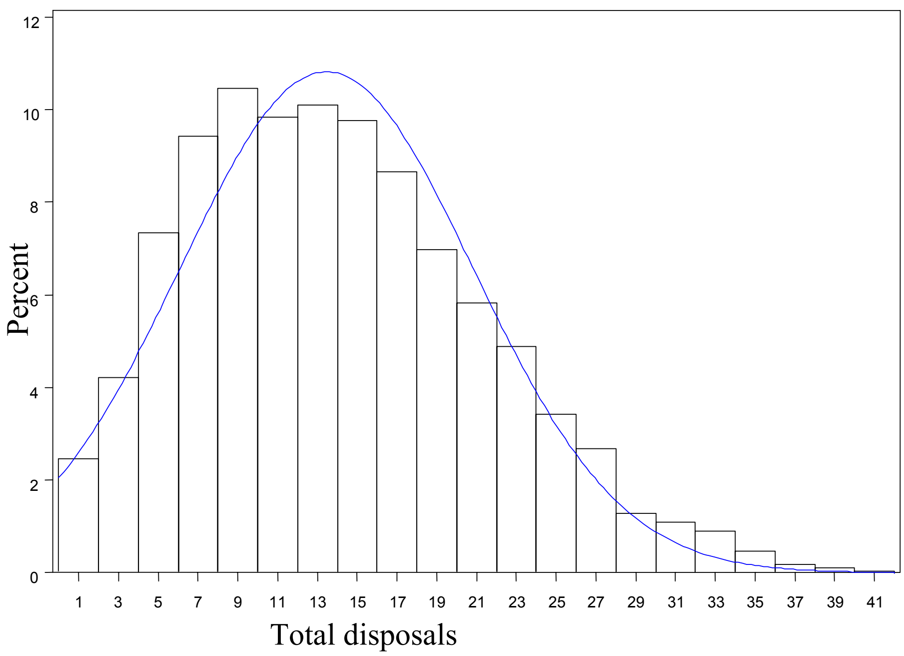

Figure 6-1 Distribution of player disposals

Because the performance of leading players from competing teams are essentially independent of each other, the task of comparing player performance is made easier than predicting the outcome of matches. Rather than having to model the difference in predicted disposals between players, valid comparisons can be made by modelling the predicted number of disposals for all leading players, and then comparing any two players through the use of standard statistical techniques for two independent normally distributed outcomes.

The aim of this chapter is to identify the best way in which to accurately assign a probability of success for the two competing players. The criteria for success will be measured in four different ways – the percentage of winners correctly chosen, the AAE between predicted and actual values, the LogProb for the winning outcomes and the return on investment that can be derived by betting against the bookmaker.

6.3 Database

In order to statistically quantify the effect of each factor, a database was progressively constructed containing match information on all games played between 1997 and 2003. Match statistics are reported by the media at the completion of each game and can readily be viewed on the AFL website[7]. Official match statistics, as endorsed by the AFL, are collected and published by Champion Data for each of the 185 matches played per year (176 home and away matches, 9 finals). Available statistics include the number of individual possessions gathered by each player along with team totals and match results. When collated, this created a working database with over 55,000 individual player performances.

Player demographics such as age, height, weight and games played are generally updated on an annual basis and can be accessed on the AFL website. Demographic

[7] www.afl.com.au

information on the 1099 AFL players that played between 1997 and 2003 was merged with match information data.

HtH betting was first introduced by bookmakers in 2001, and with approximately three different HtH bets offered per game, a data base of 1641 match-ups was collected from 555 matches played between 2001 and 2003. On 44 occasions (3.7%), one or both of the players in the HtH match-up did not take the field of play leaving a database of 1597 valid comparisons between players.

Of the 711 players that have played AFL football between 2001 and 2003, 199 players (28%) were used by bookmakers in player HtH match-ups. Players used in match-ups were more likely to play in the centre or on-ball positions and as a consequence, tended to figure more heavily in the play. This is reflected by the fact that players chosen for match-ups average over eight possessions more per game in comparison to those not chosen (20.6 ± 0.12 vs. 12.4 ± 0.04, $p<0.0001$).

6.4 Predictors of disposals

6.4.1 Overview

In order to identify factors that could affect performance, an exploratory analysis was conducted using all available data. Factors that may affect a player's performance can be categorised at a player, team and match level and can been seen in Table 6-1.

Table 6-1 Factors that may affect player performance

Player	Team	Match
Age	Key players missing	Venue
Experience	Within team ranking	Ground familiarity
Height	Quality of team	HA
Weight	Quality of opposition	Interstate travel
Body mass index	Team experience	Year
Position	Opposition experience	Round
Fitness	Difference in experience	Game time
Injury status		Ground condition
Current form		Predicted Result
Past form		
Variability in past performance		

6.4.2 Age

In recent years, the AFL has introduced a minimum age requirement of 17 years for all AFL footballers. Although there is no maximum age limit, the physical nature of AFL football ensures that few players play beyond their mid thirties. The mean age of footballers in 2003 was 23.5±0.15 years. From Figure 6-2, it can be seen that a clear relationship exists between age and performance. Not unlike most elite sports, players appear to undergo a training effect early in their career, with performance improving annually. This is reflected with the average number of disposals gathered per game increasing with age from 17 years through to 26 years (see Figure 6-2). This improvement in performance eventually plateaus out when the training effect is outweighed by the reduction in physical capacity and hand eye coordination that accompanies an increase in age. Although performance can be seen to drop slightly beyond 26 years of age, the decline is minimised as older or weaker players voluntarily or involuntarily finish their careers.

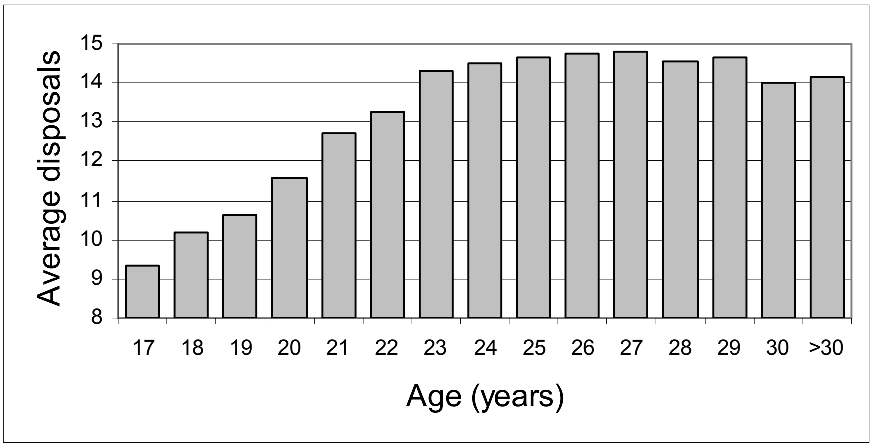

Figure 6-2 Average disposals by age

From Figure 6-2 it can be seen that the relationship between age and disposals is clearly not linear. One way to account for this departure from linearity is to set a maximum age for all players equivalent to the approximate age in which players reach their peak. By setting the maximum age to be 25 years, and using an R-square statistic as a measure of goodness of fit, it is possible to improve the amount of variation explained by age from a figure of 3.8% to 4.5%.

6.4.3 Experience

Not all players commence their AFL career at the age of 17. Because of a large physical difference between age restricted junior football (under 18) and senior football, most young players are acclimatised to open age football by playing matches at a senior level in a competition inferior to the AFL. The amount of time that a young player may spend in a lower competition can be dependent upon his personal rate of development or the specific requirements of his team. If for example an AFL team is playing poorly or has problems with injuries, they might be more inclined to play younger players, than a side that is playing well and suffering few injuries. Given the inconsistencies associated with using player age as a predictor for performance, an alternative way to numerically quantify the training effect is to use the number of matches played as a predictor of performance.

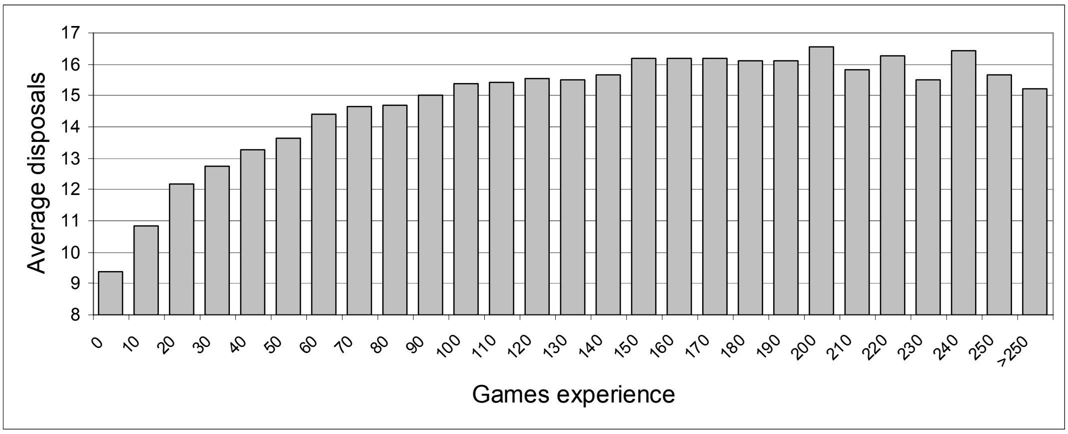

Figure 6-3 Average disposals by experience

By stratifying experience into blocks of 10 matches played, it can be seen from Figure 6-3 that AFL footballers appear to improve in performance over the first 160 games, with the greatest improvement coming over the first 50 matches. Regardless of whether the training effect is measured using age or experience, a common pattern emerges with the greatest improvement achievable in the early stages of each player's careers. Where age could explain a maximum of 4.5% of the variation associated with performance, the number of previous matches played was clearly a more sensitive measure, explaining 7.1% of the variation in possessions gathered. By setting an upper limit of 100 matches, the variation explained by experience could be increased to 9.4%.

6.4.4 Physical attributes

Body shape and size have long been linked to sporting prowess. Because of the differing physical requirements, some sports are better suited by specific body shapes. In AFL football, height, fitness and strength are all important physical attributes that can benefit a player.

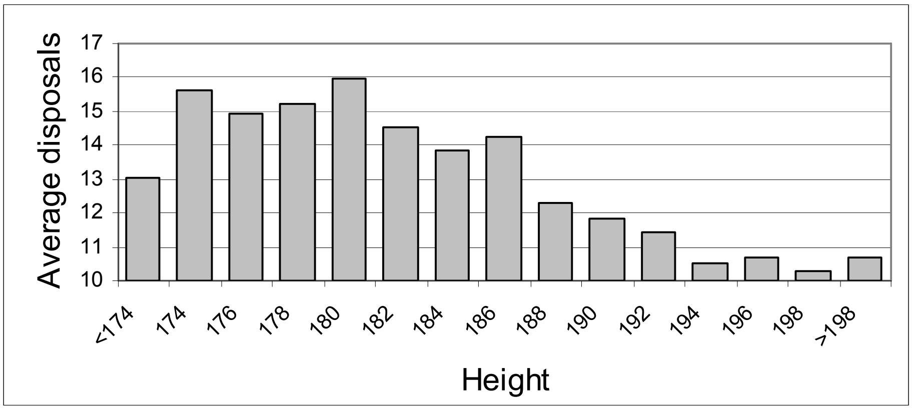

Figure 6-4 Average disposals by height

The average height of an AFL footballer is currently 185cm, but increasing at a rate of 0.1 centimetres per year. On average, shorter players tend to gather more disposals than their taller counterparts, with players below 185cm averaging 15.0±0.04 disposals per game, whilst players 185cm and above average only 12.0±0.04 disposals per game ($p<0.0001$). Although strictly not linear, Figure 6-4 clearly indicates the disadvantage that taller players have in comparison to their smaller team members.

The average weight of an AFL footballer is 85kg. Lighter players gather significantly more disposals, with those below 85kg averaging 14.2±0.04 disposals per game compared to those 85kg and above who average 13.7±0.04 disposals per game ($p<0.0001$).

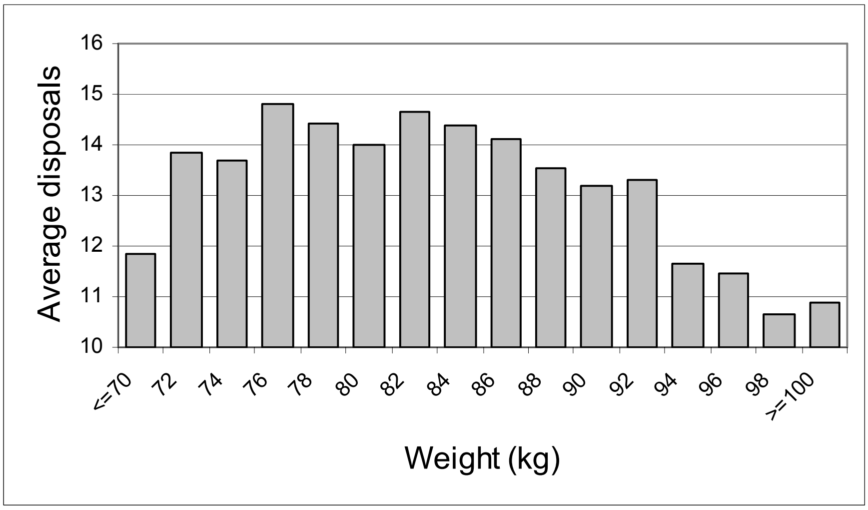

Figure 6-5 Average disposals by weight

From Figure 6-4 and Figure 6-5 it can be seen that while obviously in existence, the relationship between height, weight and disposals is far from linear. An alternative measure of physical size is body mass index (BMI), which is defined by weight divided by height squared, and reflects a player's relative body size for his height.

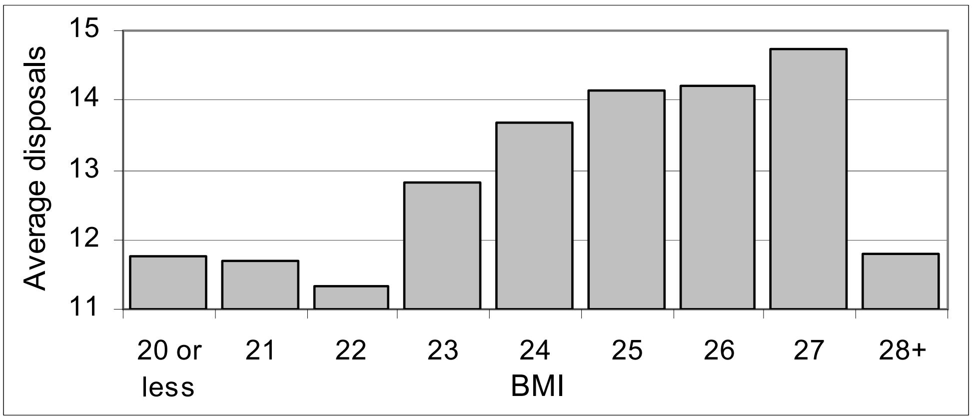

Figure 6-6 Average disposals by BMI

AFL football is an elite sport requiring high aerobic fitness. Because very few players are overweight, BMI acts as a good measurement of physical strength. From Figure 6-6 it can be seen that player performance is related to physical strength, with performance improving with BMI until a certain level (> 27), before diminishing. This suggests that players that are particularly light or overly heavy do not perform as well. BMI is also strongly linked with age (experience) as players tend to get physically stronger as their career progresses.

Although a player's physical attributes can be linked to performance, each player's specific size is more closely linked to the position on the field in which he is chosen to play, which in turn greatly affects the number of disposals he will gather.

6.4.5 Position

While 22 players are named, only 18 players from each AFL team can be on the field at any given time, with the remaining four players acting as substitutes. On-field playing positions can roughly be broken into three categories, forwards, backs and midfielders. Although player positions can vary throughout the course of the match, the starting line-up generally provides a good indication as to where a player would normally play. The 22 player positions per team are announced to the public on the Thursday evening prior to each weekend round of matches, but must be finalised 24 hours prior to the commencement of the match. A breakdown of specific player positions can be seen from Figure 6-7.

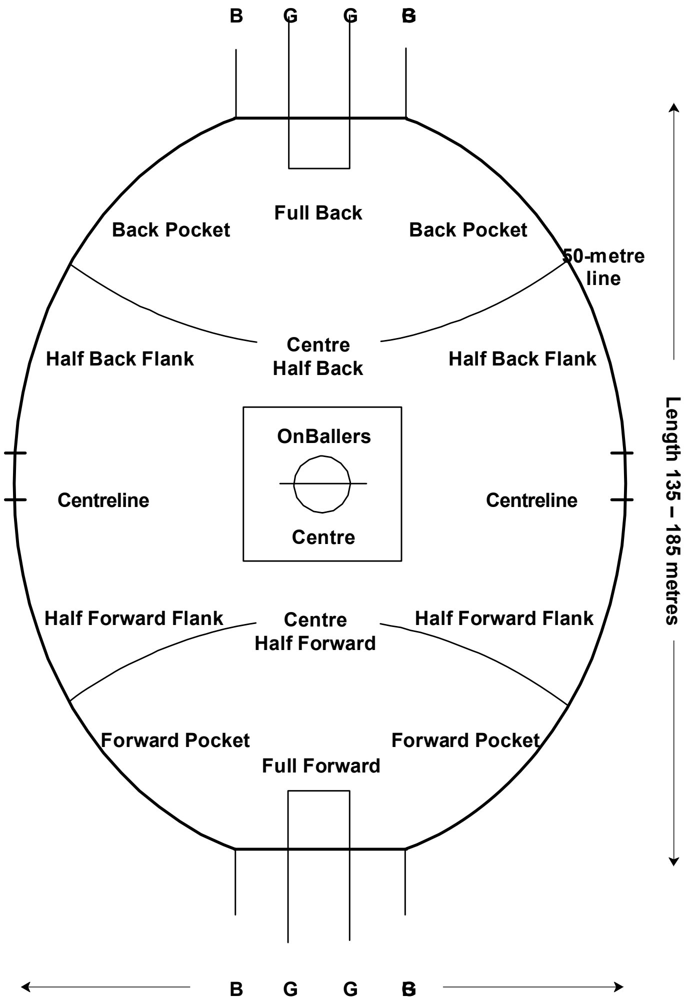

Figure 6-7 Player positions in AFL football

On average, about 30 goals are scored per game. At the commencement of each quarter and whenever a goal is scored, the ball is bounced in the centre of the ground, with ruckmen competing for first possession. As a consequence, ruckmen tend to be taller and heavier in stature. This is reflected in Figure 6-8 with ruckmen averaging 12 centimetres more in height and 10 kilograms more in weight than all other players. Little difference in height and weight can be seen to exist between backs and forwards although centre and on-ball players are significantly shorter and lighter.

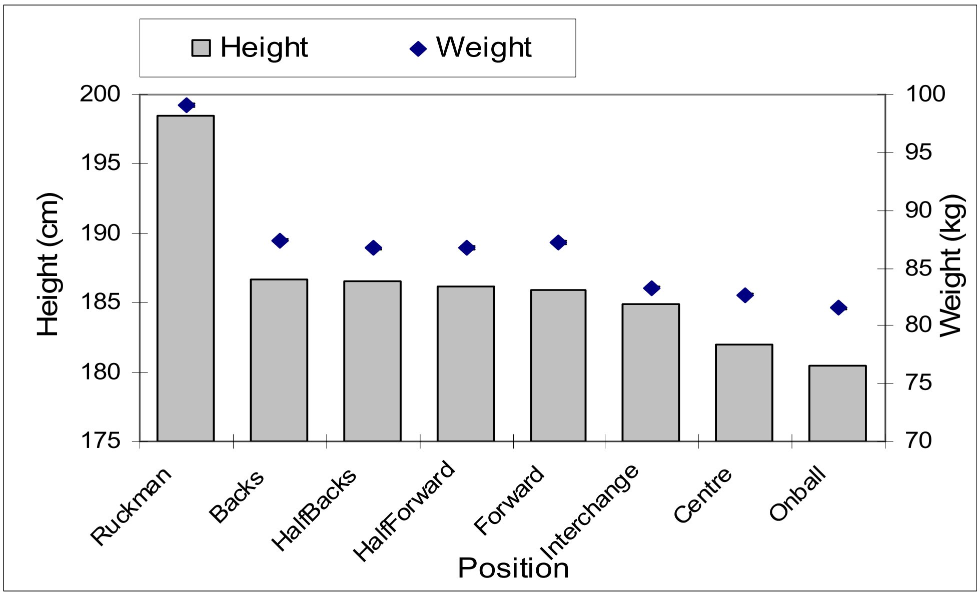

Figure 6-8 Average height & weight for each position

Because the ball is returned to the centre of the ground regularly, it is of little surprise that on-ball and centre-line players get more possessions than all other players (see Figure 6-9). There is little difference in disposals between players commencing on the half-back or half-forward lines, but players starting further from the centre such as back and forwards average less. Ruckmen traditionally average fewer disposals than others, whilst players commencing the game on the interchange bench average significantly less again.

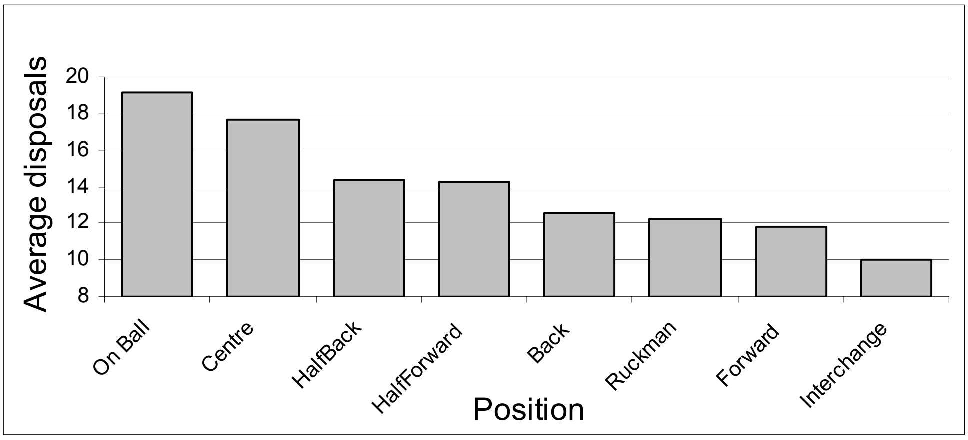

Figure 6-9 Average disposals for each position

6.4.6 Fitness

Although a player's given level of fitness is related to performance, from a statistical point of view, it is difficult to numerically quantify fitness. For tactical reasons, teams may be reluctant to completely divulge the true health status of leading players, ensuring that a player's true level of fitness is unknown. Similarly, the players themselves may be reluctant to divulge their true level of fitness to club officials. Given the physical nature of AFL football, it is widely claimed that towards the end of a season, a high percentage of players go into each match with some form of minor injury ensuring that they are not at peak physical fitness. A player's ability to perform whilst not at peak fitness is a characteristic sometimes used to differentiate between good and great players.

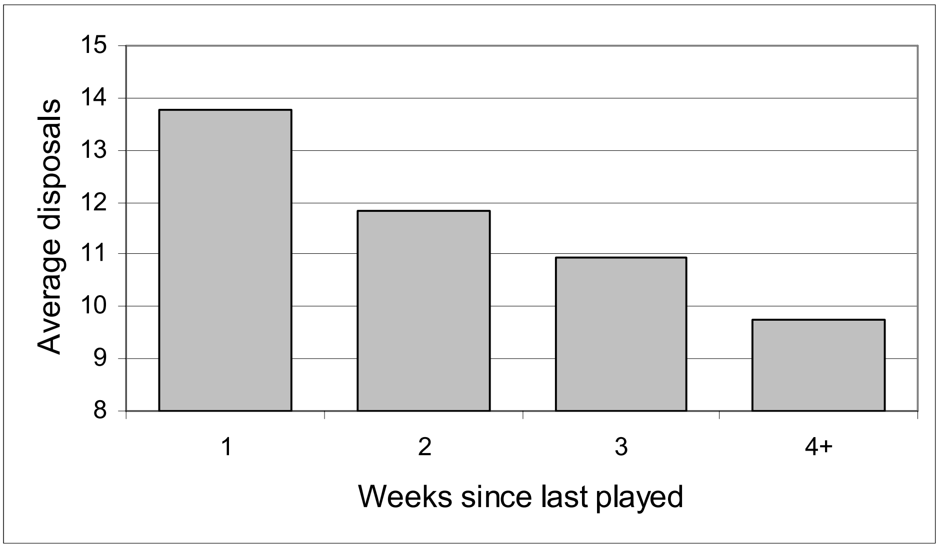

Figure 6-10 Average disposals for weeks since last game

One approach to measuring the effects due to injury is to measure the number of weeks since a player last played. From Figure 6-10 a clear negative effect can be seen with the average number of disposals going down proportional to the number of weeks since the last match. Because the severity of injuries is often measured in terms of weeks or matches missed, this relationship is of little surprise.

Another potential way to measure player fitness is to use the proportion of matches that each player has played for the year. From Figure 6-11 it can be seen that quite a strong relationship exists, with regular players averaging significantly more disposals than those who play infrequently. This may indeed be a surrogate marker for quality and experience as fringe players not only play fewer matches, but spend more time on the interchange bench and less time on the field in the matches that they do play.

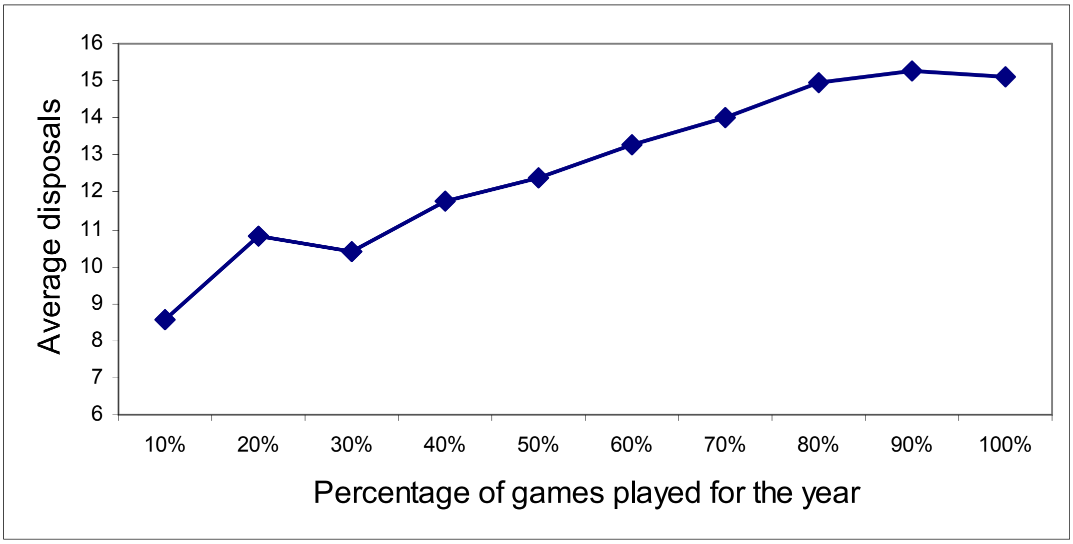

Figure 6-11 Average disposals for percentage of games played for the year

6.4.7 Days rest

From Figure 6-12 it can be seen that when players are only given five days rest between games, performance suffers accordingly. Fortunately, a five-day rest has only occurred on less that one percent of all occasions, and when it has, the AFL usually ensures that the two competing teams are equally matched for days rest. No significant difference in possessions can be observed between six, seven or eight days rest. 13.5% of all performances were played with more than eight days rest. On these occasions the average number of disposals gathered was two fewer. This could be attributed to return from injury or reflect fringe players that are not playing on a regular basis as more than eight days would equate to more than one week since the last game.

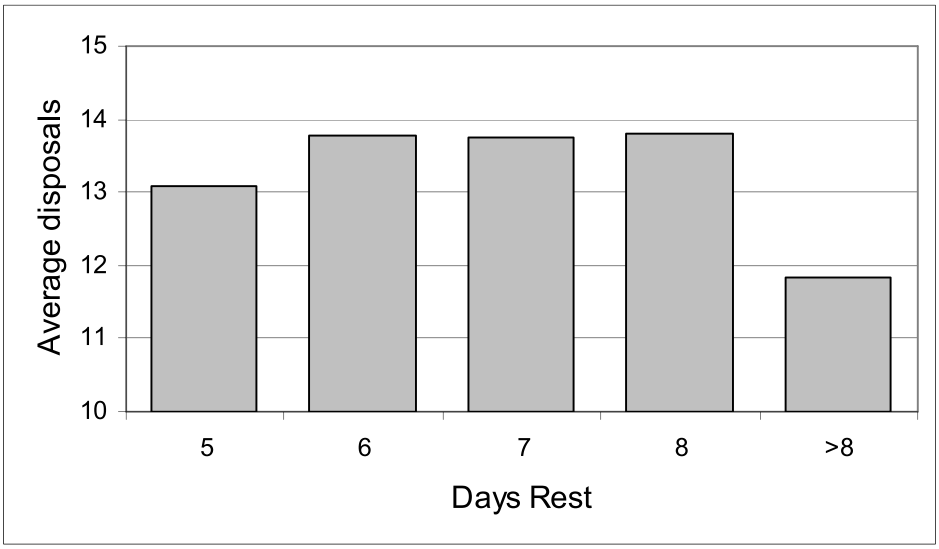

Figure 6-12 Average disposals for days rest

6.4.8 Recent form

If a player played well in his last match, he is likely to retain his position in the team because he is perceived to be in good form – does this necessarily mean he will play well the following week? Upon examination of the data, it is of little surprise to see that there is a high degree of auto-correlation between consecutive player performances. In fact, 23.4% of the variation associated with number of disposals gathered, can be directly explained by using each player's most recent performance as a linear predictor. By averaging each player's number of possessions for his last two matches, it is possible to obtain an improved predictor of form that can explain 29.4% of the variation associated with number of disposals gathered. This poses the question, what is the optimal predictor of recent form?

Using the R-square statistic as a guide to model fitting, it can be seen from Figure 6-13 that the best predictor for form can be obtained by averaging each player's previous performances for the given season. This can be primarily attributed to the fact that there is a six-month break between seasons in which the structure and nucleus of each team can change significantly. Coupled with the opportunity to fully recover from injuries and build on physical strength, it makes intuitive sense to examine player performance on a seasonal basis.

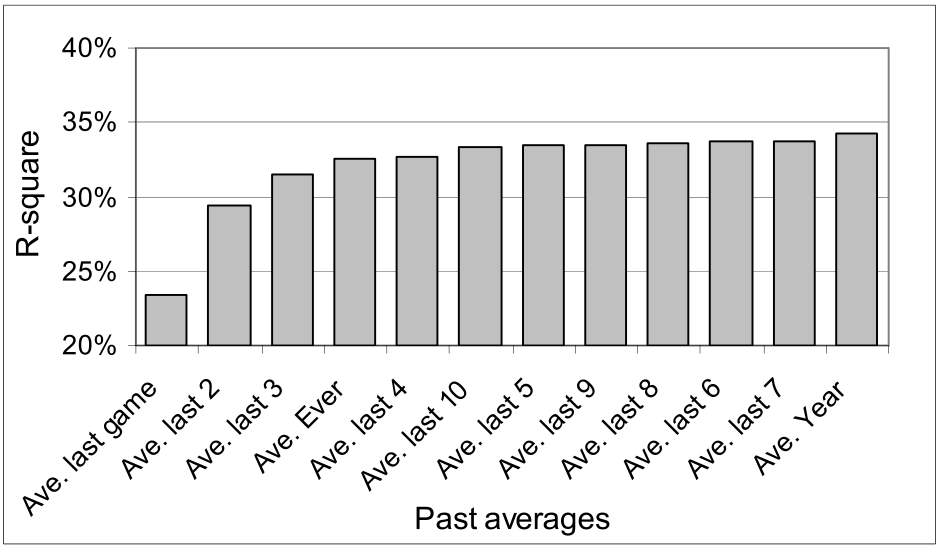

Figure 6-13 Model R-square for predictors of form

While seasonal measures would appear to be the strongest, there is actually very little practical difference between using a player's average for the season and using a moving average of anywhere from four to 10 weeks. This poses the question of when does a predictor go from measuring recent form to measuring the overall quality of a player?

6.4.9 Measures of quality

Although a high degree of correlation exists between a player career average and his most recent performance, it could be argued that the former is a measure of quality, whilst the later is a measure of current form. Clearly, any measure of overall quality should account for recent performances, but a career average gives equal weighting to all past performances. Intuitively, how a player performed five years ago should be less important when predicting performance than how he played last week. By using an exponentially smoothed predictor, (see section 3.4) it is possible to derive a prediction of quality that gives greater weight to more recent performances. When exploring the goodness of fit for exponential smoothers, three smoothing parameters were chosen ($x = 0.1$, $x = 0.2$ & $x = 0.3$). Although it would be possible to specifically optimise the value of x to further improve results, the small differences that were found to exist between the three chosen levels, suggest three values would be sufficient.

In order to ascertain if player performance is specific to a certain opposition or a chosen venue, additional smoothed predictors were created to smooth each player's performance against the given opposition as well as at the chosen venue. The same three values for x (0.1, 0.2 & 0.3) were once again chosen. Using the amount of variation in disposals that each predictor could explain as reflect by an R-square statistic, these nine exponentially smoothed predictors were compared with each player's career average against the opposition, at the given venue and overall.

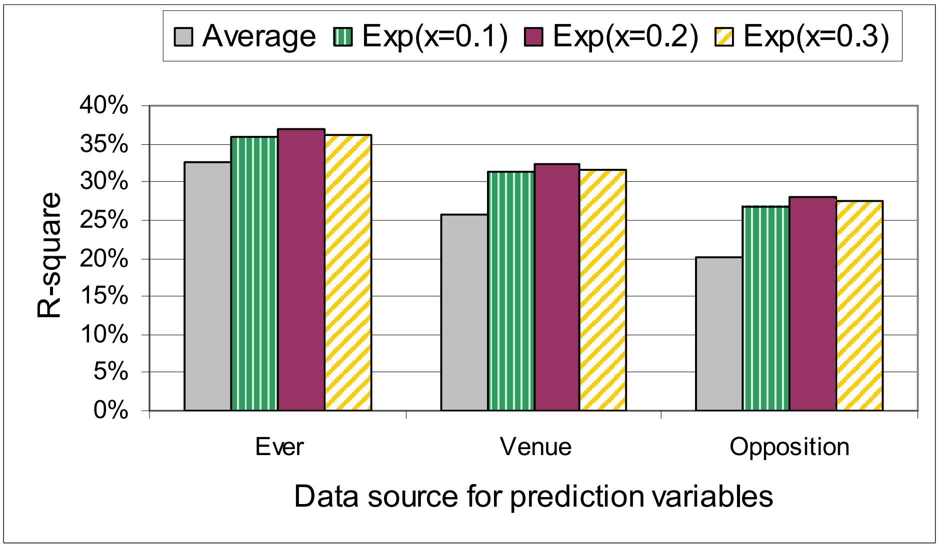

Figure 6-14 Comparison for measures of performance

From Figure 6-14, three important points can be observed.

1. The use of past performances can explain large amounts of variation associated with the number of disposals gathered throughout the course of the game. Of the 12 predictors considered, each could explain more than 20% of the variation, with the best approaching 37%.
2. Predictors using all past data can explain more information that predictors created at a venue or opposition level.
3. Exponential smoothers can explain more variation than past averages, and although a smoothing parameter of x=0.2 produced the best results; very little practical difference could be observed between the three chosen levels of x.

6.4.10 Within player variability

Not only can past results be linked to future performances, but the standard deviation of past results can also be linked to future performances. This can be seen from Figure 6-15 with players that have higher variability gathering more possessions. Intuitively, this may be of little surprise, as we would expect those with higher averages to have higher variability. But, the standard deviation of past performances can be viewed as an independent predictor for disposals, as it is still highly statistically significant ($p<0.0001$), even after adjusting for the effects of past average.

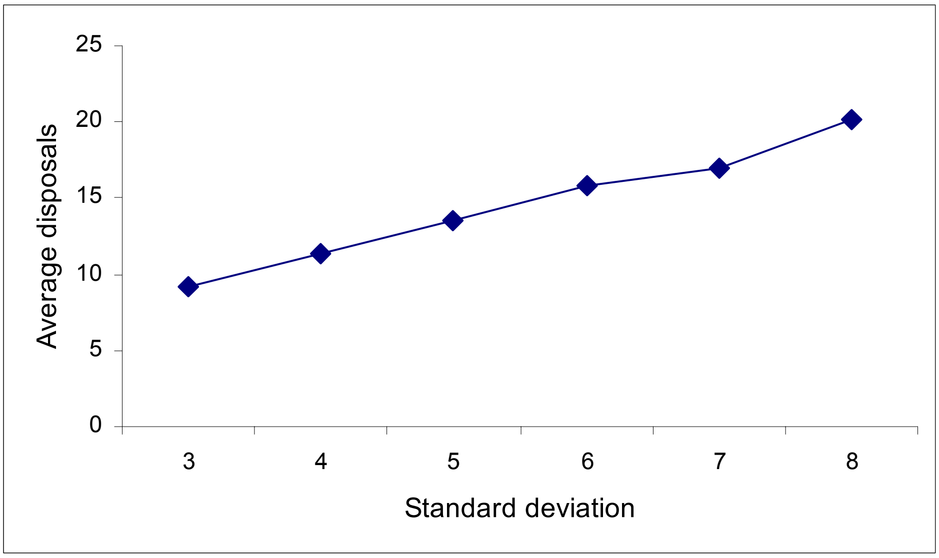

Figure 6-15 Relationship between average disposals and standard deviation

6.4.11 Key players missing

When good players are unavailable for a match, the resulting change to team structure can have a positive or negative effect on other players. By using player averages it was possible to identify the leading five possessions winners from the previous game. By then counting how many of these five players were unavailable for the current game, it was possible to numerically measure how many key players were missing. As can be seen from Figure 6-16 the more key players missing, the harder it becomes for other players to gather possessions.

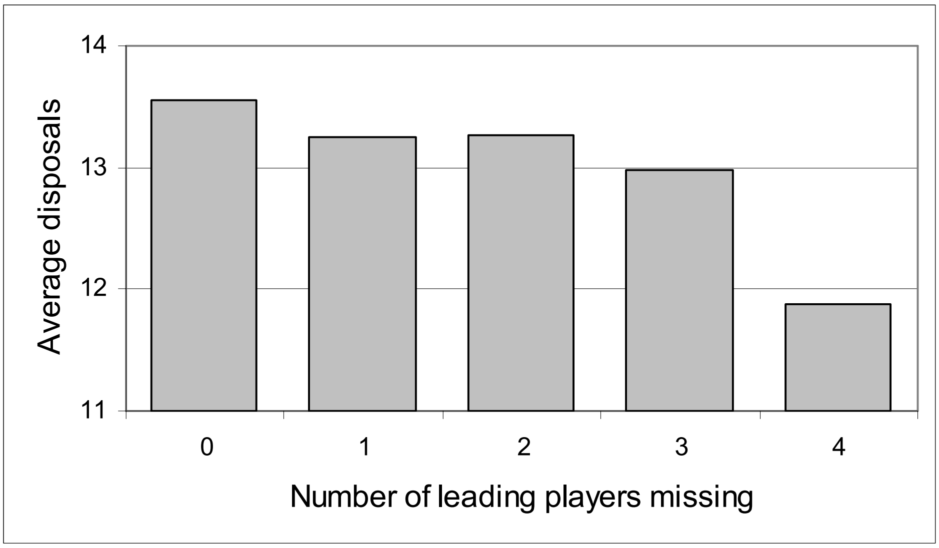

Figure 6-16 Relationship between key players missing and average disposals

6.4.12 Within team ranking

Although it would appear that losing key players has a negative effect on the number of disposals collected, it can also have a positive effect. Theoretically, having the leading possession winner out of the team could mean that the second leading possession winner could have more opportunity. In order to measure this, average disposals were used to create within team rankings. Thus, if the leading ranked player was missing for a week, all players would move up one position in the rankings. As team ranking are simply a non-parametric marker of past average, it is of little surprise that team rank, as seen in Figure 6-17 is a highly significant predictor of player disposals.

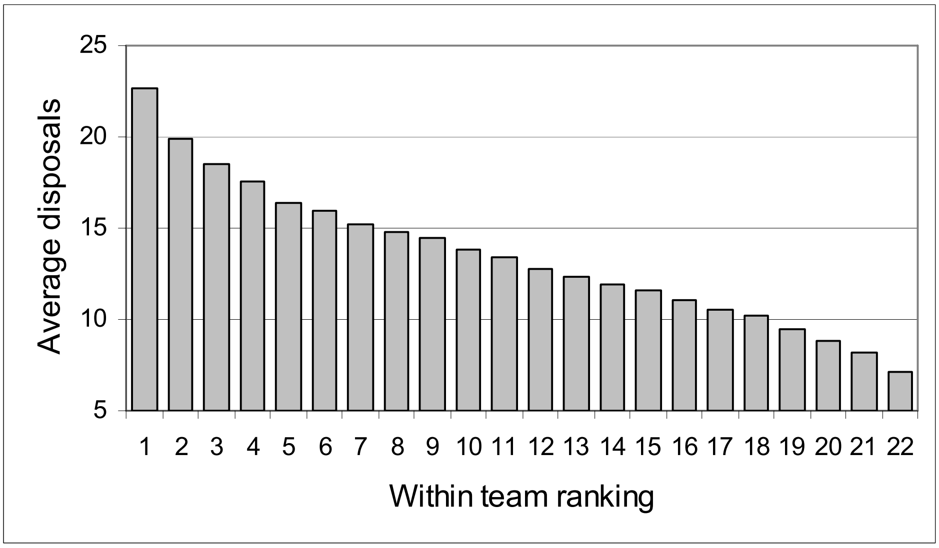

Figure 6-17 Average disposal by ranking within the team

6.4.13 Team

The team that a footballer plays for can have an impact on the number of possessions that they will gather for the game. This can be seen from Figure 6-18 and may be a reflection of the composition of the team, the home venue or even the style of coaching. Over the period 1997 to 2003, the Kangaroos have been the team with the lowest possession tally. Coincidently, Denis Pagan was the coach of the Kangaroos for the majority of this period, and would regularly adopt a 'kick the ball long' policy toward winning matches. Ironically, since the start of the 2003 season, Pagan has been coaching Carlton which has been the team with the highest possession count over the period of analysis.

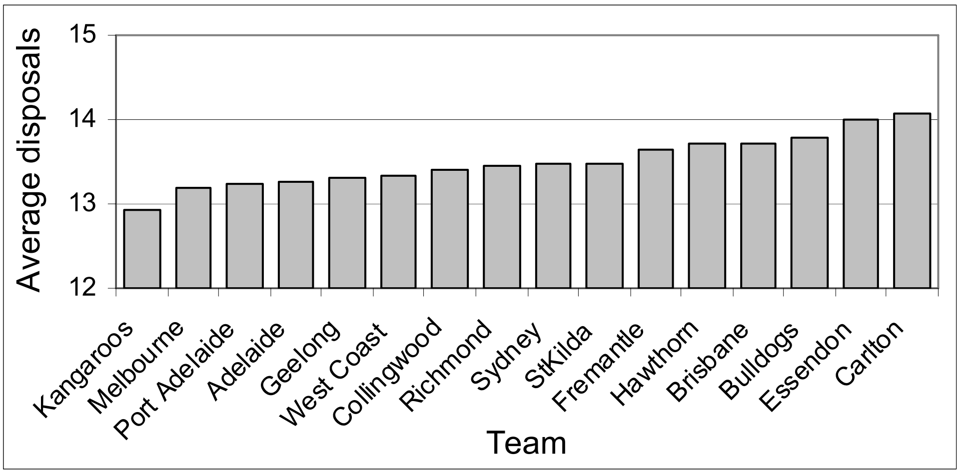

Figure 6-18 Average disposals by team

6.4.14 Opposition

The quality of the opposition is a statistically significant predictor for the number of possessions that a player will gather. From Figure 6-19 it can be seen that Sydney has been the team in the last seven years that it has been most difficult to gather possessions against. This may in part be a legacy of the fact that Sydney's home ground is the smallest of all venues, making it slightly harder for players to find free space, and subsequently easier possessions.

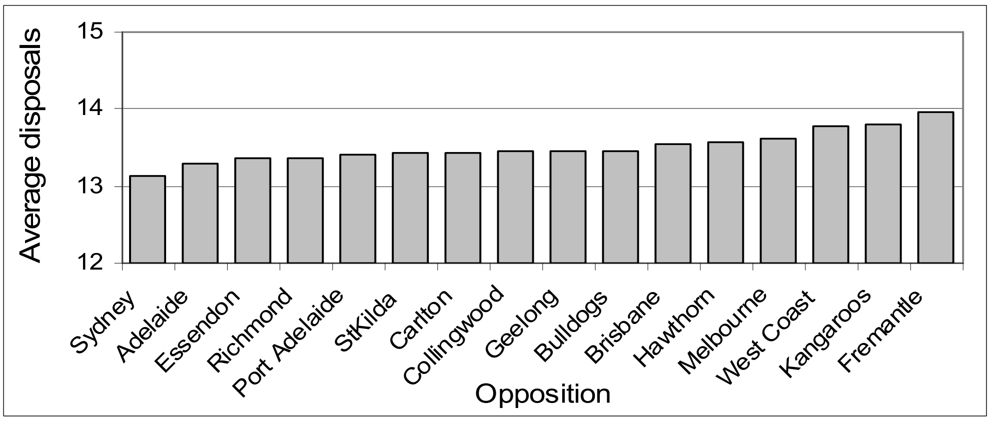

Figure 6-19 Average disposals by opposition

Because the coaching staff and team nucleus can change from year to year, it is unrealistic to expect team or opposition effects to remain as consistent predictors. An alternative way to quantify team and opposition effects is to consider the combined amount of experience that each team has for each match.

6.4.15 Team and opposition experience

By averaging the numbers of matches played by players from competing teams, it is possible to derive a numerical measure of team and opposition experience. Figure 6-20 suggests evidence of a team training effect, with players finding it harder to get disposals when playing with less experienced team members. When a team average exceeds 50 matches per player little additional improvement appears to be gained.

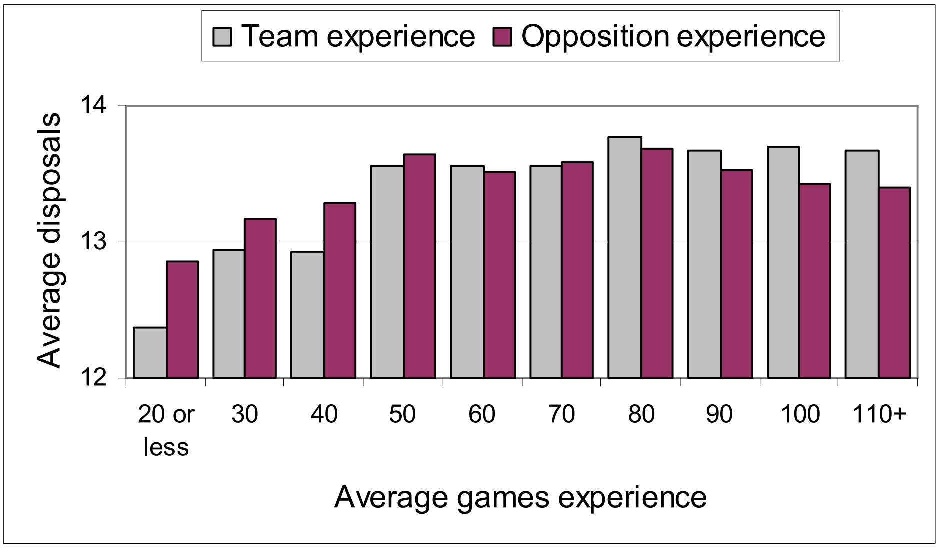

Figure 6-20 Team and opposition experience

Given that it appears harder to get possessions when you are playing with a relatively inexperienced team, one could reasonably expect that it might be easier to get possession when playing against inexperienced opposition. Surprisingly, the opposite is true. Figure

6-20 shows that the more inexperienced the opposition the harder it becomes to gather disposals!

By subtracting opposition experience from team experience it is possible to derive a quantitative measure for the difference in team experience. This has approximate linear properties with the greater the difference in experience between teams, the more likely a player is to gather possessions. (Figure 6-21)

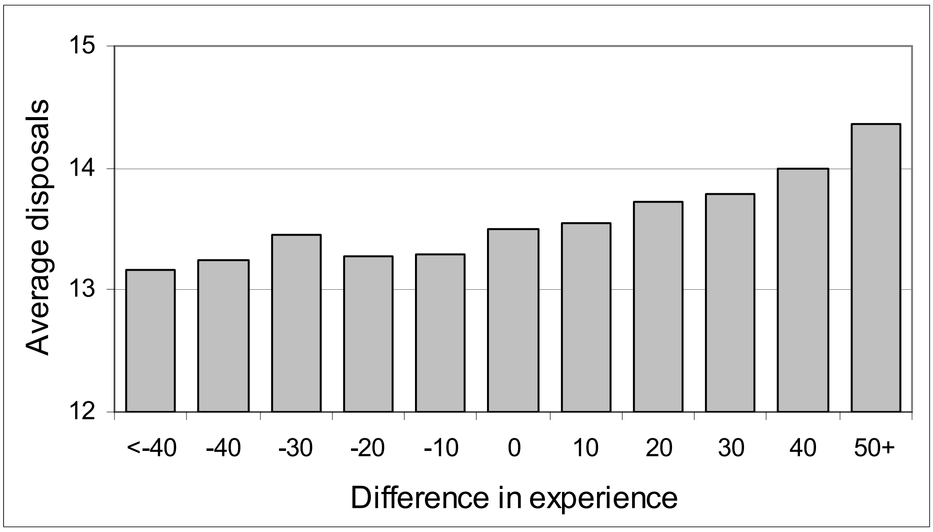

Figure 6-21 Difference in experience

6.4.16 Night games

There is no difference in the number of possessions gathered between matches played during the night and matches played during the day (13.50±0.04 vs. 13.50±0.04, p=0.99).

6.4.17 Weather conditions

Although AFL football is a winter sport with the majority of matches played outside, less than four percent of all matches played in the past seven years have been played in wet conditions. In comparison to dry conditions, players on average perform a little worse in the wet, although this difference is not statistically significant (13.5±0.03 vs. 13.3±0.13, $p=0.09$). It is realistic to expect that wind may also be a contributing factor towards player performance but unfortunately, this information is not readily available. Given the dramatic climate differences that are evident between the capital cities within Australia, it is not surprising that on average, players gather fewer possessions when playing in open venues in the colder southern states. This can be seen more clearly when considering the specific venues for each game.

6.4.18 Venue

Seventeen different venues have been used to host AFL matches in the last seven years although only 10 venues have hosted more than 20 matches. Of the 17 venues used, only one has the facilities to completely protect the players from inclement weather (Colonial Stadium). Depending on the size, shape, location and surrounding structures, each venue offers different forms of protection from the weather. Figure 6-22 shows clear differences in the average number of possession gathered at each venue with higher possessions gathered in the warmer states and the enclosed venue, whilst fewer possessions are gathered in venues that are more exposed to the effects of winter such as Skilled stadium in Geelong.

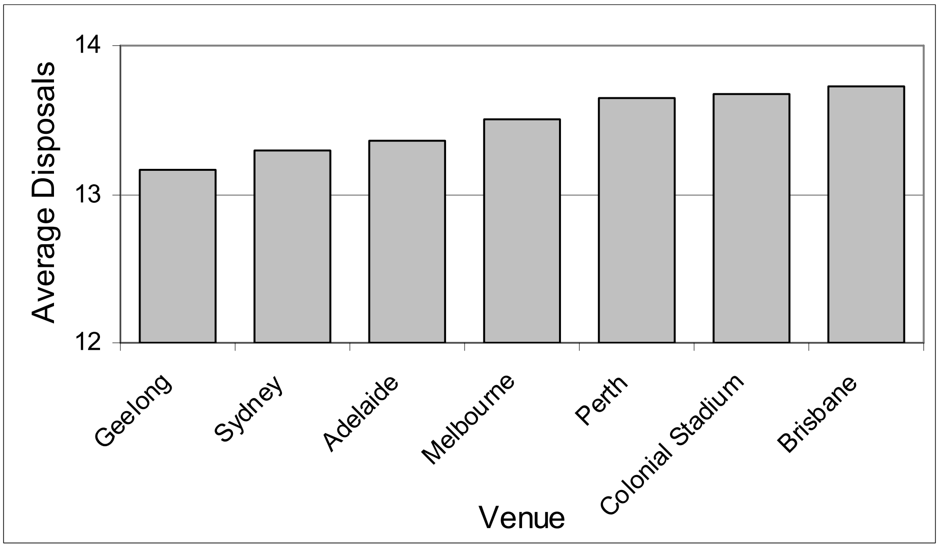

Figure 6-22 Average disposals by host venue

When matching players from competing teams, an overall effect due to venue will be of little consequence as both players will be affected equally by playing on the same venue. In order to determine if certain players play better at specific venues, each player's average at the venue along with exponentially smoothed predictors for performance at the specific venue were calculated. Whilst being highly significant predictors for disposals, predictors created at a venue level were not as good as predictors created using all past data, but may still be of some practical benefit when considering multivariate models.

6.4.19 Ground familiarisation

As previously discussed, it has been hypothesised that ground familiarly is a potential reason for the established phenomena of HA. This effect is clearly evident in Figure 6-23, with the number of disposals gathered increasing with the number of matches played at the venue. Despite being strongly correlated with overall experience, ground familiarisation is an independent predictor of player performance.

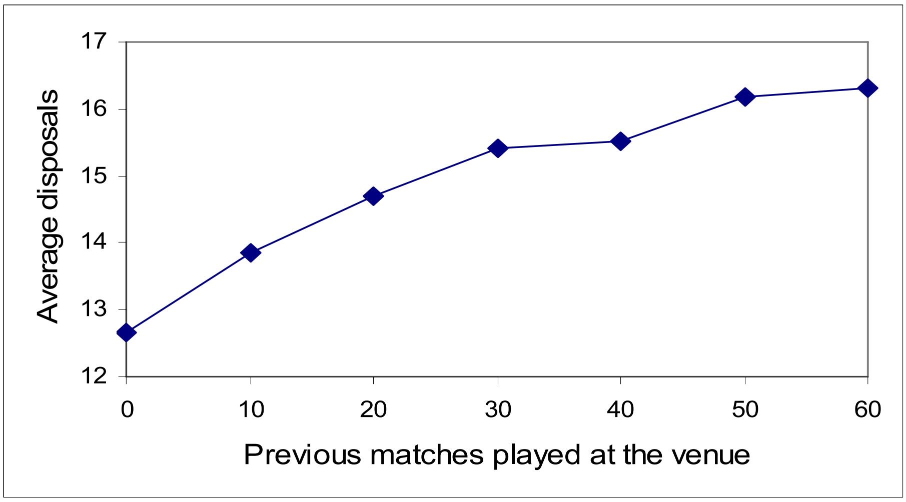

Figure 6-23 Previous games experience at each venue

6.4.20 HA and interstate travel

Players that play on their home venue on average gather a half a possession more than players playing away from home (13.8±0.04 vs. 13.3±0.04 $p<0.0001$). From Figure 6-24, no linear relationship could be seen to exist between distance travelled and possessions gained, although players who flew interstate on shorter flights (less than two hours) were significantly worse off than those travelling further or those playing at an away venue in their own home city. (13.7±0.07 vs. 13.3±0.07 $p=0.006$). This may be attributed to the fact that for shorter flights, teams often chose to travel and play on the same day whilst for longer journeys, travel will occur at least one day prior to the scheduled match time, giving players time to recover from any effects due to travel.

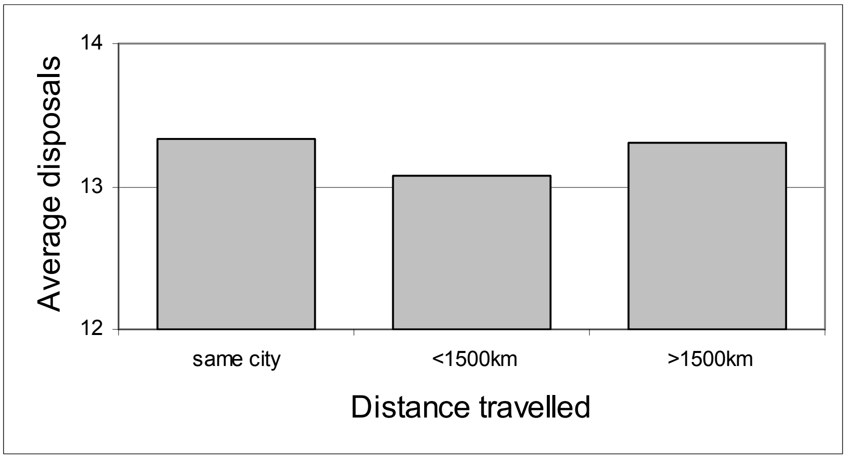

Figure 6-24 Average disposals for distance travelled to venues

6.4.21 Match result

On average, players from the winning side will accumulate one more possession per game than players from the losing team (14.0±0.04 vs. 13.0±0.04 $p<0.0001$). From Figure 6-25, this relationship can be further extended, with numbers of possessions gathered, clearly related to the MOV.

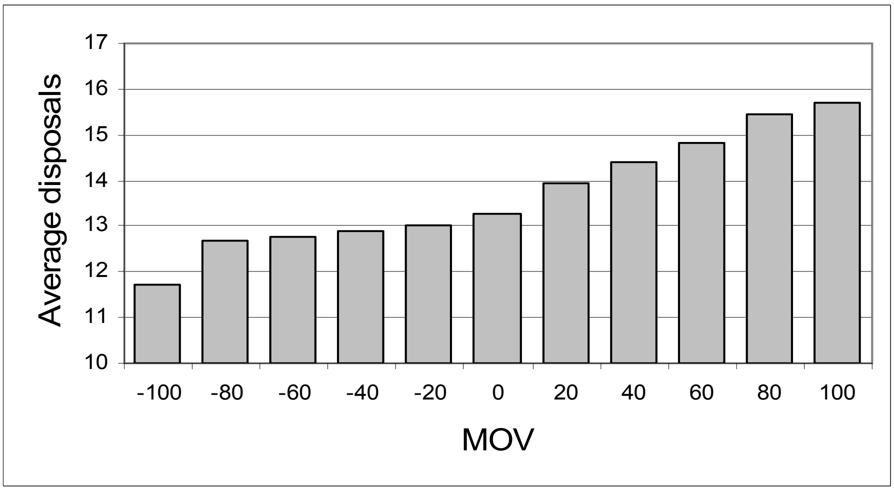

Figure 6-25 Disposals by match result

Unfortunately, the result of the match cannot be used to predict the number of possessions that a player may gather, as the result of the game is obviously not known prior to the commencement of the match. Alternatively, the predicted MOV was explored as a predictor of individual performance. Using information based on HA, interstate travel, ground familiarisation, team performance and form, a multivariate linear model was constructed to predict the outcome of matches. As shown in chapter 4, such a modelling approach could successfully be used as a predictor of match outcome. Unfortunately, no significant relationship can be seen to exist between the predicted match result and the individual performance of players within the game. (Figure 6-26)

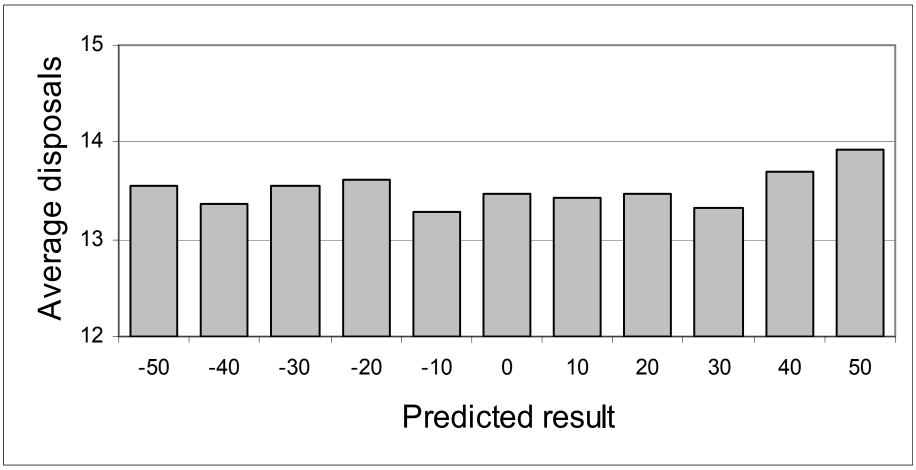

Figure 6-26 Disposals by predicted result

As an alternative to using the predicted result, the bookmaker price for each player's team can be used to predict individual player performance. Using the reciprocal of the team price offered by bookmakers it can be seen from Figure 6-27 that as the team probability of winning increases, so to do the average number of disposals for the players from that team.

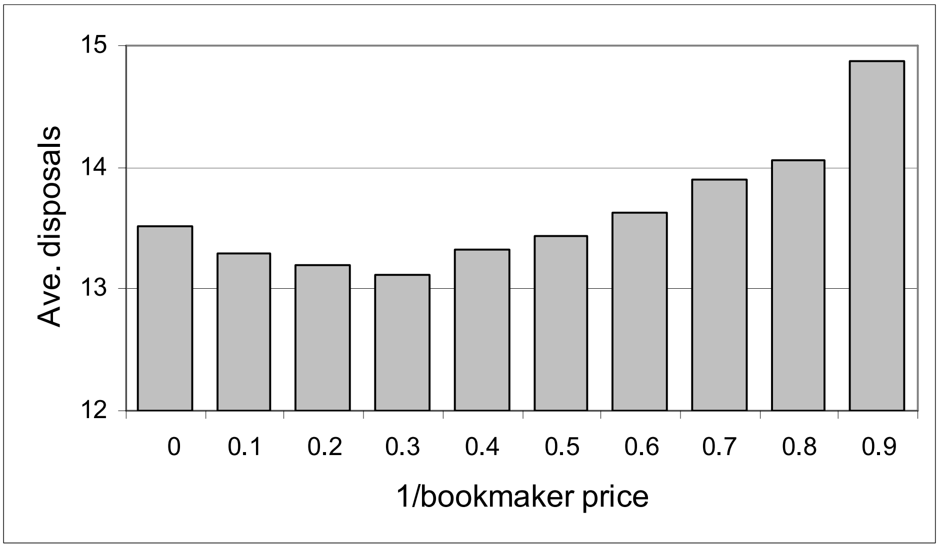

Figure 6-27 Average disposals for reciprocal of bookmaker prices for team outcome

6.4.22 Round

AFL football is played over a six-month period from April through until September. This comprises of 22 rounds of home and away matches and four rounds of finals. With complete data from only seven seasons of football it is not feasible to attempt to isolate effects due to specific rounds, but is perhaps feasible to explore trends that may occur throughout the course of the season. By dividing the data into blocks of about five rounds, Figure 6-28 enables us to explore potential trends. Although there does not appear to a strong relationship, there is evidence to suggest that players get progressively better approaching the finals. Once in the finals, it then becomes more difficult to gather possessions. Similarly, players gather fewer possessions during preseason matches, although this can be directly attributed to the shorter duration of preseason games.

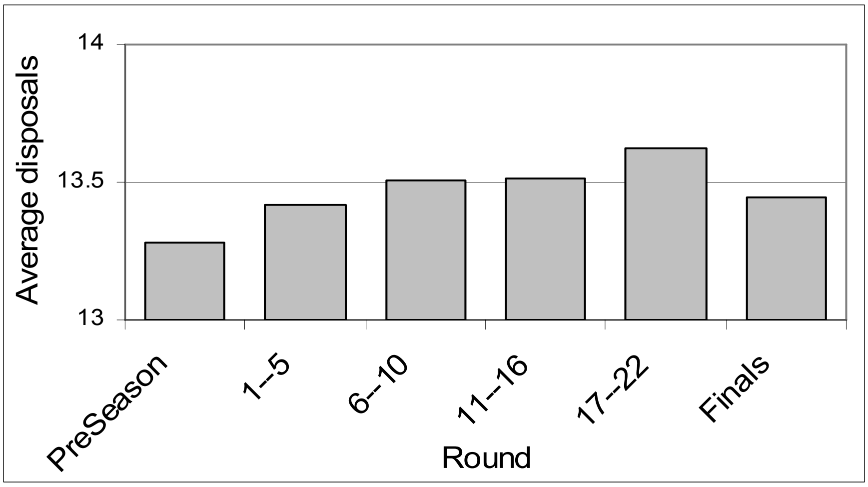

Figure 6-28 Average possessions per round

6.4.23 Year

In the seven years from which the database has been established, significant annual changes have occurred in the average number of possessions gathered. From Figure 6-29 it can be seen that the most notable of these differences occurred in 1997 when the average number of possessions was significantly lower than all other years and in the year 2000 when the average number of possessions gathered was significantly higher than all other years. Although difficult to pinpoint the exact reasons why such dramatic differences can be seen to occur, there are several possible explanations for these differences.

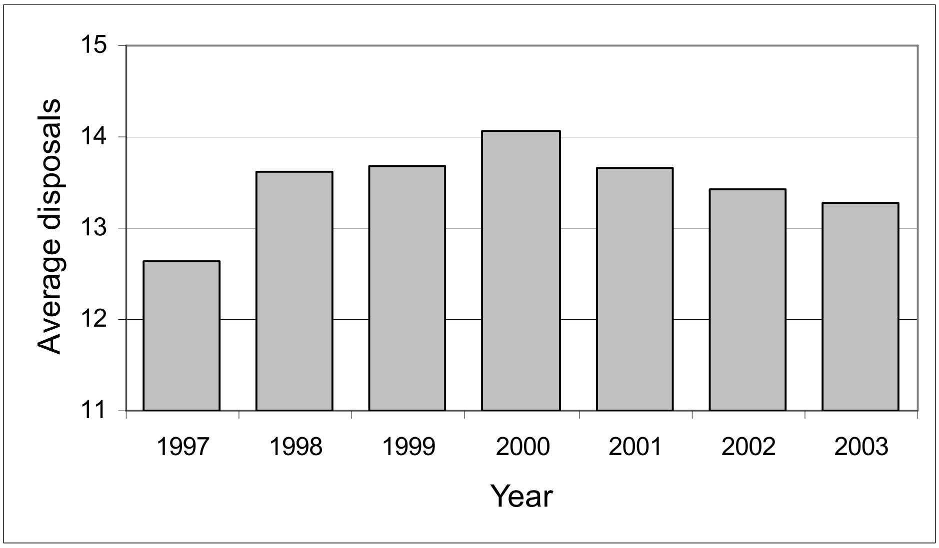

Figure 6-29 Average disposals by year

6.5 Annual differences

6.5.1 Style of play

Although AFL football has been played for over 100 years, it has only been in the last 20 years or so that the style of play has begun to change. It is of little surprise that this change in style has coincided with the increased professionalism now associated with playing football at an elite level. In the past it was the norm for players to engage in regular Monday to Friday employment, train after work and play football on the weekends. Today, the commitment required to play AFL at the elite level ensures that very few current players have other jobs, with clubs careful to ensure that should players have alternative commitments, it does not encroach on their playing performance. This increase in professionalism has resulted in fitter and stronger players which in turn has increased the pace at which football is played.

Footage of older matches suggest a style of play based on field position, with less emphasis on retaining possession, and more emphasis or moving the ball toward goals. Play was more stop-start in nature with players regularly kicking towards a contested possession. Today, teams are far more concerned with maintaining possession as much as possible, ensuring that there are far fewer contested possessions. This approach to maintaining possession at all costs probably reached its peak in the year 2000 when eventual premiers Essendon lost only one game for the season. Since that time, opposing coaches have taken counter measures such as flooding the defence to ensure that this approach is not quite so successful.

6.5.2 Rule changes

At the end of each season, a review is made of the rules with minor changes often made prior to the next season. Although changes are often only small, it is still possible to impact on the style of play and as a consequence, the number of possession gathered throughout the course of the game.

6.5.3 Data collection processes

Champion data have been the official collectors of AFL football statistics since 1998. Although player statistics were collected prior to this point in time, it is unlikely that an established and consistent process was used to define and collect data, which is probably why the average number of disposals in 1997 was significantly lower than all other years.

6.5.4 Definitions

During the course of a match, numerous occasions arise when it is questionable as to whether a player correctly disposed of the ball. This creates uncertainty as to whether an official possession should be awarded or not. Although there may always be a degree of subjectivity involved in the allocation of possessions, Champion Data have looked to standardise the definition of what constitute an official possession. By incorporating additional staff to validate results, Champion data have worked on developing and refining a consistent and reproducible data collection process. Not surprisingly, this evolution has seen subtle changes in definition occur, with most changes occurring at season's end. These changes of definition could well account for the differences that occur between years.

6.6 Training dataset

Although it is feasible to use all available data when investigating causes of variation in the number of possessions gathered, this is not the case when looking to identify potential predictors. In order to accurately determine the predictive capacity of a regression model, data is traditionally partitioned into a training data set from which parameter estimates are derived, and a hold-out sample in which the regression equations are applied. Such a process is used to avoid the bias associated with over-fitting.

Because an assessment of market efficiency requires the use of bookmaker data, the three years in which bookmaker data are available namely 2001, 2002 and 2003 have been chosen as the holdout sample with parameter estimates derived from the four years played prior to 2001. When comparing the average number of possessions per player per game gathered in the training data with that of the holdout data, there are no significant differences (13.5 ± 0.05 vs. 13.4 ± 0.05 $p=0.16$). But, when comparing the variances between the two data sets, the training dataset was found to have a significantly higher standard deviation (7.0 vs. 6.8 $p<0.0001$). When looking at the differences that occur from year to year, this is of little surprise as both the highest of years (2000) and the lowest of years (1997) are in the training dataset. Because both the training and holdout samples have in excess of 25,000 data points, this difference in standard deviation should have very little negative impact on the prediction process.

6.7 Linear regression

Perhaps one of the simplest prediction processes comes in the form of a linear regression. Linear regression involves the fitting of a straight line to a set of data in order to predict the value of one variable with the use of another.

By using the method of least squares on the training dataset it is possible to derive a regression equation that minimises the error between predicted and actual possessions. By then applying the prediction equation to data in the holdout sample, it is possible to predict the number of possessions that each player will gather. It is important to note that each variable is treated as though having a linear relationship with disposals. From our exploratory analysis it was apparent that for several variables such as age and experience, a linear relationship was not the best possible option, although, all 34 variables considered did have a highly statistically significant linear relationship with disposals ($p<0.0001$).

From a statistical viewpoint, these results were highly significant, but from a practical point of view, with a training data set in excess of 33,000 data points, an R-square value of 0.1% was sufficient to produce a p-value of this magnitude. Whilst it may be possible to improve goodness of fit for each individual variable though the use of transformations or constraints, at this stage, each variable in the multivariate regression will assume a linear relationship.

From Figure 6-30 it is possible to see that the predictor that produces the lowest standard error is an exponentially smoothed predictor for all past performances. Curiously, almost all predictors produced smaller AAEs in the holdout rather than training dataset. After careful cross validation of these surprising results, the author can only attribute this paradox to the large quantity of data and the decreased quality in the training data that was reflected by an increased variance.

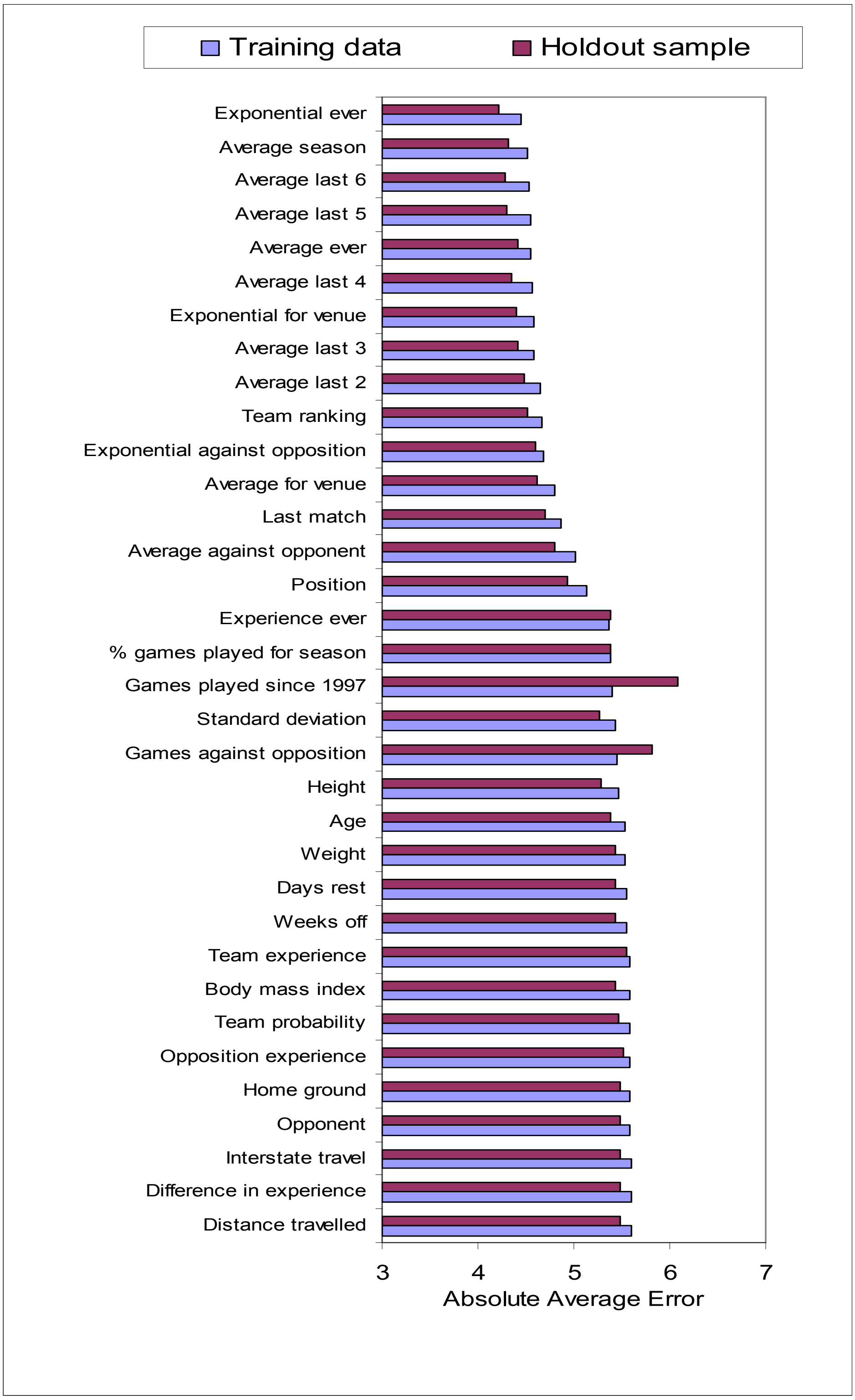

Figure 6-30 AAE in training and holdout data for 34 predictors of disposals

In order to explore market inefficiency in HtH betting, individual player predictions must be compared and the probability that one player will outperform the other determined. Should the performance of the two competing players be independent of one another, then this process can readily be achieved.

6.8 Converting predictions to probabilities

6.8.1 Independence

By randomly matching disposals gathered by players from opposing teams it was found that no significant correlation could be identified between players. When specifically looking at the 1597 bookmaker created match-ups, a weak but statistically significant relationship could be observed between players from opposing teams (R-square=1.3% p<.0001). As seen in Table 6-2, the first match-up presented by the bookmaker for each game is generally the leading possession winner from each team, with first ranked players averaging 23.5 possessions per game. The second and third match-ups for each game average 20.4 and 18.8 possessions per game respectively. When considered as a whole, the natural ordering that exists in the three match-ups for each game ensured that competing players appear correlated. After accounting for natural ordering, the correlation between matched players for both sides is reduced to an R-square figure of 0.5%. Although still statistically significant (p=0.05) a correlation of this magnitude is of little practical significance, ensuring that valid comparisons between players can still be made.

Table 6-2 Average disposals by bookmaker rankings

Order	First pair	Second pair	Third pair
Average disposals	23.5±0.5	20.4±0.6	18.8±0.6

Having satisfied the assumptions for normality and independence, the probability that one player will gather more possessions than another can readily be calculated by dividing the difference in predicted disposals by a combined standard deviation and comparing with a standard normal curve.

Assuming that disposals gathered by player A are approximately normally distributed with a mean μ_A and a variance σ^2_A and player B ~ N(μ_B, σ^2_B) then

$(A\text{-}B) \sim N(\mu_A\text{-}\mu_B, \sigma^2_A + \sigma^2_B)$

Thus

$P(A>B) = P((A\text{-}B)>0)$

$= \Phi\,[(\mu_A\text{-}\mu_B)/(\sigma^2_A + \sigma^2_B)^{1/2}]$

An unbiased estimate for $(\mu_A\text{-}\mu_B)$ can readily be given by $Y_A - Y_B$ where Y_A and Y_B are the predicted means derived through the use of linear regression.

An unbiased estimate for the pooled variance $(\sigma^2_A + \sigma^2_B)$ can be created by weighting each player individual standard deviation in accordance with the number of matches played by each player. Thus

$$S^2_{\text{Both}} = \frac{(n-1)S^2_{\text{A}} + (m-1)S^2_{\text{B}}}{(n+m-2)}$$

Where S^2_A is the standard deviation of player A, S^2_B is standard deviation of player B, n is the number of games played by player A and m represents the number of games played by player B.

Given that the number of possessions gathered by each player is well approximated by a Normal distribution it is possible to use each players past data to create an unbiased estimate of the standard deviation. Each player's individual standard deviation can be estimated by

$$S_A^2 = \frac{\sum\left(\mathrm{x_A} - \overline{x_A}\right)^2}{n-1} \quad \text{and} \quad S_B^2 = \frac{\sum\left(\mathrm{x_B} - \overline{x_B}\right)^2}{m-1}$$

where $\mathrm{x_A}$ equals disposals for player A, $\overline{x_A}$ equals the average for player A, n equals the number of games played by player A, $\mathrm{x_B}$ equals the number of disposals by player B, $\overline{x_B}$ equals the average for player B and m equals the number of games played by player B.

In order to ascertain if market inefficiencies exist in HtH betting, it is then necessary to incorporate an appropriate wagering strategy as discussed in section 3.10.

6.9 Market efficiency

Each of the 34 univariate models considered was evaluated for goodness of fit by considering the percentage of winners correctly picked, the AAE in the training dataset, the R-square in the training dataset, and the ROI as defined by total profit divided by total outlay. From Table 6-3 it can be seen that the bottom eight of the 34 models considered were able to produce a positive ROI. Using a Wilcoxon rank sum test, it was possible to ascertain if the amount of profit returned was significantly greater than zero. Three univariate predictors, namely an exponentially smoothed predictor for all performances, an exponentially smoothed predictor for performance at the specific venue and each player’s average for the season, were all able to produce profits that were significantly greater than zero.

Table 6-3 Results for the 34 univariate models, sorted by ROI

Variable	**Number of bets**	**Average bet size**	**Percentage of Winners**	**ROI**	**R^2 training**	**AAE training**
Opposition experience	742	$92±2	46.2%	-19.7%	0.2%	5.59±.02
Opponent	757	$91±2	47.1%	-19.6%	0.2%	5.59±.02
Distance travelled	711	$86±2	33.7%	-19.4%	0.1%	5.59±.02
Team probability	683	$84±2	49.9%	-19.3%	0.2%	5.59±.02
Interstate travel	704	$87±2	33.6%	-18.9%	0.1%	5.59±.02
Days rest	736	$89±2	33.4%	-18.8%	1.4%	5.54±.02
Standard deviation	721	$84±2	51.7%	-18.7%	3.6%	5.44±.02
Difference in experience	717	$84±2	51.5%	-18.3%	0.1%	5.59±.02
Weeks off	684	$98±2	43.7%	-18.1%	1.4%	5.55±.02
Weight	762	$93±2	44.8%	-18.0%	1.6%	5.54±.02
Home ground	734	$85±2	51.5%	-17.9%	0.2%	5.59±.02
Team experience	723	$84±2	51.4%	-17.7%	0.5%	5.58±.02
Body mass index	737	$87±2	50.2%	-17.5%	0.5%	5.59±.02
Height	813	$95±2	45.4%	-14.5%	4.3%	5.46±.02
Age	833	$100±2	49.4%	-14.4%	3.5%	5.53±.02
Games played for season	761	$96±2	34.4%	-13.6%	6.7%	5.39±.02
Experience ever	953	$130±3	51.3%	-8.7%	7.9%	5.37±.02
Games against opposition	972	$127±3	43.5%	-7.8%	4.8%	5.45±.02
Average last 3	858	$116±3	51.2%	-4.7%	31.2%	4.59±.02
Average last 2	955	$131±3	51.7%	-3.3%	29.2%	4.65±.02
Team ranking	785	$101±2	43.5%	-3.2%	28.9%	4.67±.02
Last match	1018	$150±3	48.8%	-3.9%	23.1%	4.86±.02
Games played since 1997	985	$140±3	53.2%	-3.7%	6.9%	5.40±.02
Ave. against opponent	993	$137±3	53.2%	-1.7%	18.4%	5.02±.02
Average last 4	784	$105±3	54.0%	-1.7%	33.2%	4.56±.02
Exp. against opponent	1101	$178±3	54.8%	-0.8%	28.4%	4.69±.02
Position	994	$153±3	38.1%	1.8%	13.2%	5.13±.02
Average last 5	736	$98±2	54.7%	1.8%	33.8%	4.55±.02
Average Ever	962	$130±3	55.4%	3.3%	33.1%	4.55±.02
Average last 6	724	$95±2	54.3%	5.5%	33.1%	4.53±.02
Average for venue	997	$139±3	56.5%	6.7%	24.8%	4.80±.02
Exponential ever	607	$87±3	56.8%	8.6%*	35.6%	4.44±.02
Exponential for venue	994	$153±3	55.5%	8.9%*	31.6%	4.58±.02
Average season	639	$97±2	56.7%	9.9%*	33.1%	4.51±.02

* Average profit per bet significantly greater that zero ($p<.05$)

6.10 Goodness of fit

A clear advantage of considering a range of different prediction models is that it enables a relationship to be established between the quality of predictors in the training data and the goodness of fit of data in the holdout sample. This can be achieved by comparing the AAE in the training data with the number of winners successfully picked.

By assuming that the player with the highest predicted probability would be the winner, it is possible to consider the percentage of outcomes that each model would successfully predict. Using information gathered from the 34 univariate models, it can be seen from Figure 6-31 that an inverse relationship exists with the percentage of winners predicted increasing as the AAE in the training data gets smaller.

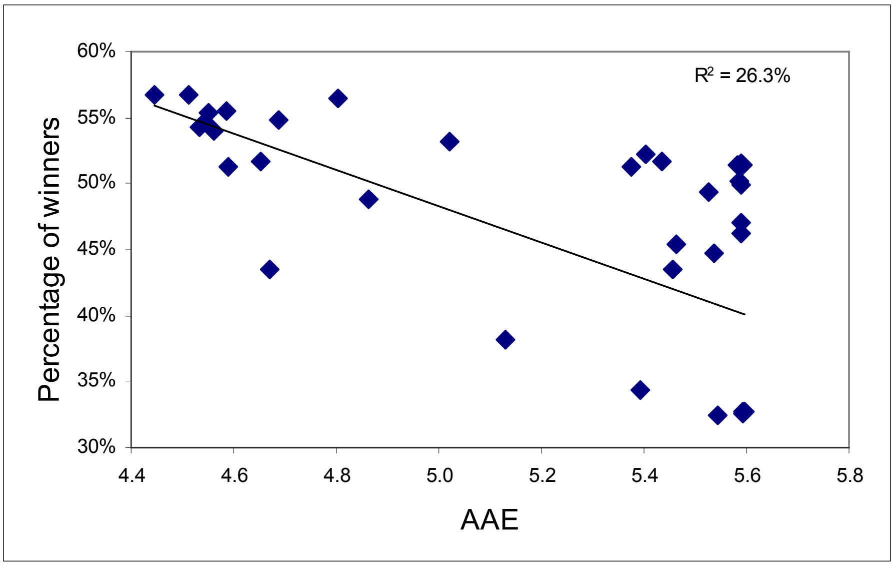

Figure 6-31 Relationship between winning percentage and AAE in training data

The disadvantage to a goodness of fit test based solely on which player had a higher probability of winning is that a substantial amount of information is lost. A model that assigns a winning player a 51% probability of success is considered identical to a model that assigns the same player a 90% probability, whereas in reality, the model that gives the winning player a higher chance of success should be viewed as being a more accurate model. Thus, a more appropriate measure of goodness of fit can be obtained by considering the assigned probability for the winning player as given by the LogProb of the winning player (see section 3.7).

From Figure 6-32, a clear linear relationship can be seen between the AAE in the training data and the LogProb of the winning player. This further enhances the theory that the model producing the lowest AAE in the training data will in fact be the model that produces for the most accurate comparison between players in the holdout data.

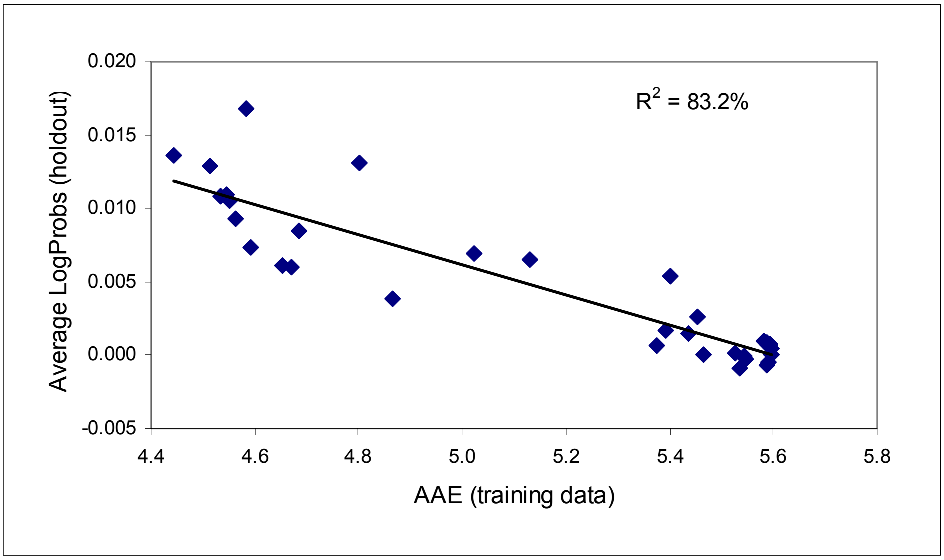

Figure 6-32 Relationship between AAE and goodness of fit (LogProb)

Intuitively, the models that more accurately assign the probability of success will be the models that produce the greatest ROI. This is clearly evident from Figure 6-33 as a strong linear relationship can be seen to exist between the average LogProb of the winning players and the ROI offered.

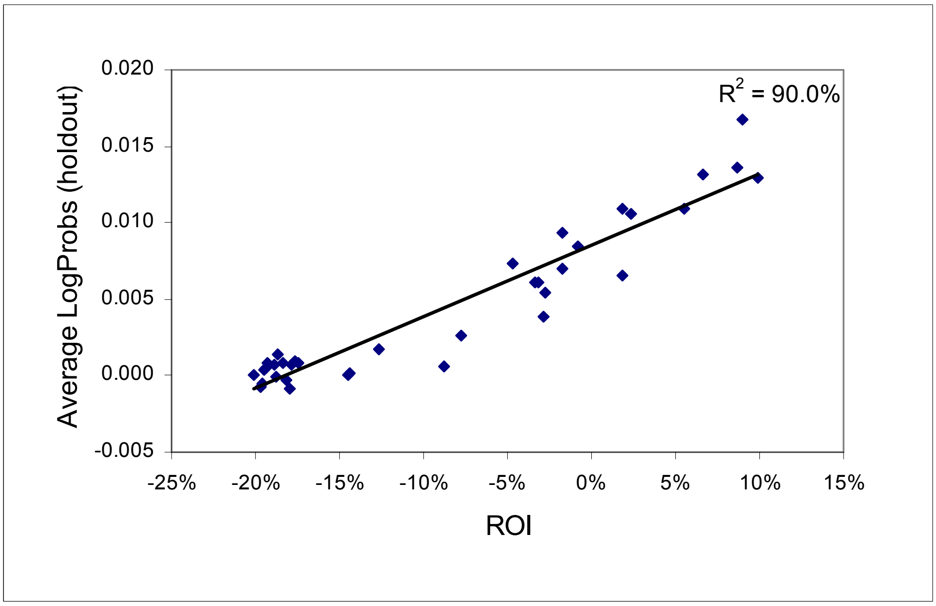

Figure 6-33 Relationship between goodness of fit (LogProb) and ROI

The relationship between goodness of fit in the training data and ROI can also be directly viewed by considering the AAE in relation to ROI. From Figure 6-34, ROI can be seen to increase in accordance with a reduced AAE in the holdout sample, indicating that the models that best fit the training data will produce the highest ROI.

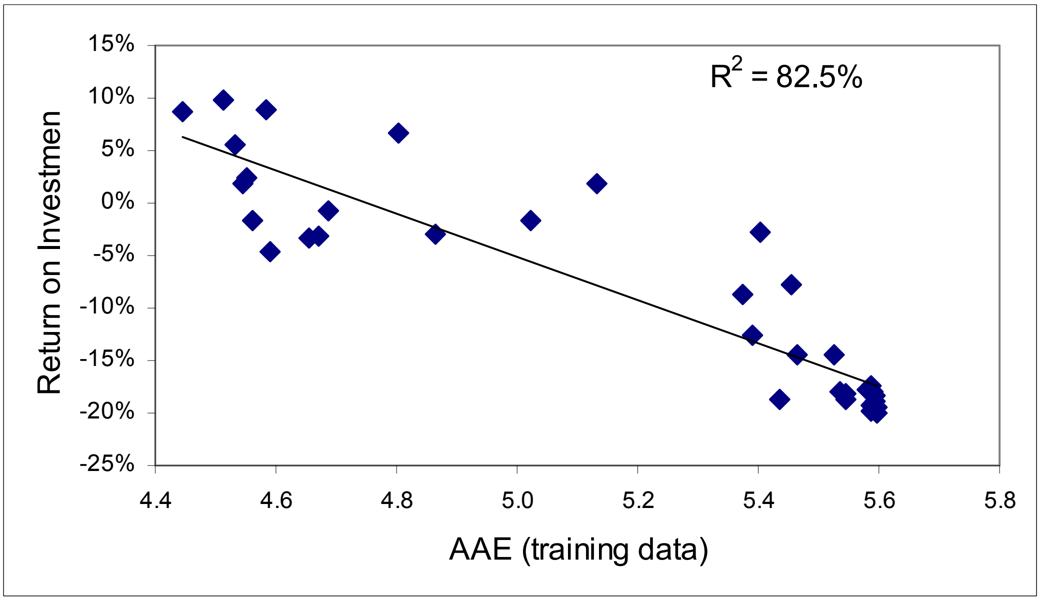

Figure 6-34 Relationship between AAE and ROI

Although is has been possible to identify what single model can best be used predict player performance, more importantly, clear linear relationships have been established between goodness of fit in the holdout sample and ROI. Although intuitive in nature, the confirmation between quality of fit and greater potential for profit opens the door for the use of a multiple linear regression to improve upon goodness of fit.

6.11 Multiple linear regression

Fifteen variables were identified as being independent statistically significant predictors ($p<0.001$) for the number of disposals gathered. Table 6-4 shows the parameter estimates and corresponding partial and total R-square values for the addition of each variable into the multivariate model. All up, the multivariate model could explain 37.7% of the variation in disposals, although the strongest variable, an exponentially smoothed predictor for past performance (x=0.2), could individually explain 36% of the variation in disposals.

Table 6-4 Parameter estimates and model R-square for final 15-variable model

Variable	Step	Parameter Estimate	Partial R^2	Model R^2	P-value
Exponential Ever	1	0.40±0.02	36.02%	36.02%	<0.0001
+ Average Ever	2	0.21±0.02	0.57%	36.58%	<0.0001
+ Position	3	Categorical	0.35%	36.93%	<0.0001
+ Home Ground	4	0.60±0.06	0.18%	37.11%	<0.0001
+Average last 2 games	5	0.05±0.01	0.14%	37.25%	0.0002
+ % played for year	6	0.87±0.16	0.11%	37.36%	<0.0001
+ Height	7	-0.03±0.005	0.06%	37.42%	<0.0001
+ Average Year	8	0.10±0.02	0.06%	37.47%	<0.0001
+ Games Experience	9	0.005±0.001	0.05%	37.52%	<0.0001
+ Team Experience	10	-0.01±0.002	0.05%	37.57%	<0.0001
+ Weeks Off	11	-0.55±0.11	0.04%	37.61%	<0.0001
+ Standard deviation	12	0.04±0.01	0.03%	37.65%	0.0007
+ Key players out	13	0.31±0.07	0.03%	37.68%	<0.0001
+ Last game	14	0.03±0.009	0.03%	37.70%	0.0003
+Diff. in Experience	15	0.008±0.002	0.02%	37.72%	<0.0001

6.12 Results

It was previously observed for Table 6-3 that the first variable in the multivariate model could produce a ROI significantly greater than zero. While the improvement in R-square that is achieved by adding 14 more variables may appear small (1.7%), it can be seen from Figure 6-35 that this enhancement serves to increase the ROI from a figure of 8% up to 20%.

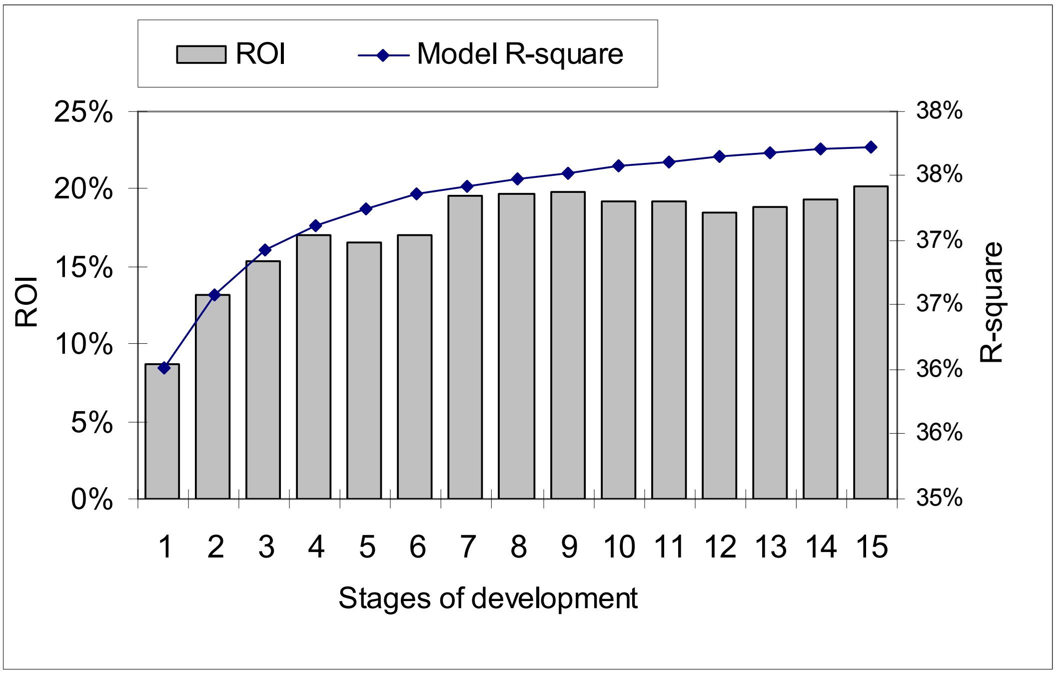

Figure 6-35 Model R-square and ROI for 15 stages of model development

From Figure 6-35, we see that the greatest improvement in ROI was achieved with the addition of the first three variables. Although the final 15 variable model produced the highest ROI, in reality there was very little improvement in ROI obtained after the addition of the first seven variables.

The progressive increase in ROI corresponding to each additional variable being added to the multivariate model can be seen in more detail from Table 6-5. Of the 1597 betting outcomes offered by Centrebet about one third of all HtH match-ups were identified as being inefficient.

Table 6-5 Return on investment for each stage of the multivariate model

Variables	Step	Number of bets (n=1597)	Average bet size	Average profit per bet	ROI
Exponential Ever	1	607	\$87±3	\$7.6±2.7*	8.6%
+ Average Ever	2	591	\$85±2	\$11.0±2.7*	13.1%
+ Position	3	579	\$79±3	\$12.0±2.7*	15.3%
+ Home Ground	4	570	\$78±2	\$13.4±2.7**	17.0%
+Ave. last 2 games	5	574	\$81±2	\$13.0±2.7**	16.5%
+ % played for year	6	569	\$80±2	\$13.7±2.7**	17.1%
+ Height	7	550	\$77±2	\$15.0±2.7**	19.5%
+ Average Year	8	543	\$77±2	\$15.1±2.6**	19.7%
+ Games Experience	9	549	\$77±2	\$15.2±2.6**	19.8%
+ Team Experience	10	555	\$79±2	\$15.1±2.6**	19.3%
+ Weeks Off	11	547	\$77±2	\$15.0±2.6**	19.2%
+Standard deviation	12	563	\$78±2	\$15.0±2.6**	18.5%
+ Key players out	13	544	\$76±2	\$14.8±2.6**	18.9%
+ Last game	14	532	\$76±2	\$14.9±2.6**	19.3%
+Diff. in Experience	15	538	\$75±2	\$15.1±2.6**	20.2%

*probability(profit>0) p-value<0.05, ** probability(profit>0) p value<0.0001

While the profit per bet derived from using only one variable in the prediction model was significantly greater than zero (p<0.05), when at least four variables were included in the multivariate model, the average profit per bet became highly significant (p<0.0001).

6.13 Discussion

Whilst statistical evidence exists of the inefficiency of the player HtH betting markets, the magnitude of this inefficiency is only relative to the quality of the prediction model. The fact that a profit, significantly greater than zero could be identified using only a univariate predictor suggests a high degree of inefficiency in the market. The use of a multiple linear regression to create predictions, further enhances this inefficiency, and confirms the benefits of a statistical approach to predict individual performance in AFL football.

In stage one of this analyses, multivariate models were used to predict the outcome of both AFL football and ODI cricket. While there was clear evidence that a statistical approach could benefit the prediction of match outcomes, the results of this chapter suggest that a multivariate approach is even more beneficial when predicting individual performance. The fact that over 37% of the variation can be explained using past data indicates that individual player performance in AFL football is far more predictable than match outcomes. Because individual player performances in AFL football are essentially independent, the probability that one AFL player will outperform another can be readily made through comparison with a standard normal distribution, giving clear indications that markets for HtH betting in AFL football are inefficient. This beckons the question as to whether player performances in ODI cricket can be evaluated in the same fashion, and if so, are inefficiencies also present in HtH betting for cricket markets.

7 Predicting individual player performance in ODI cricket

7.1 Introduction

Player performances in ODI cricket are quite different from AFL football as are the corresponding betting markets. While the number of possessions gathered in AFL football is used as a guide to performance, in ODI cricket, the differing components of batting and bowling require separate assessment.

Betting on player performance in AFL football has been available since 2001 with Centrebet offering in excess of 500 HtH match-ups per year. Despite more than 120 ODIs being played annually, HtH match-ups for player performance in ODI cricket have only been introduced to the betting public in the last couple of years. As a result betting has generally been restricted to higher profile games, providing insufficient data to fully assess efficiency.

With 50 matches played over a six week period, and player HtH betting available for most matches, the 2003 World Cup of cricket provided an opportunity to explore the efficiency of player HtH betting in ODI cricket. This chapter aims to determine if batting performance could best be assessed through the use of multiple log-linear regression models, and if so, could such a process then be utilised to assess market efficiency.

Using data collected from all ODI matches played prior to the World Cup, mathematical models were constructed to predict the expected scores of individual players. The predicted means were used to make pair-wise comparison between leading players within and between teams. Each player's probability of scoring more runs than his opponent was calculated. These probabilities were then compared with bookmaker prices on offer for the World Cup and market efficiency assessed.

7.2 Background

The World Cup of Cricket is the leading tournament for ODI cricket, with all teams that are sanctioned by the International Cricket Council competing. Held every four years, the 2003 World Cup was held in South Africa and comprises a series of preliminary rounds culminating with finals. Not only does the World Cup provide entertainment to a worldwide television audience, it also creates wide interest for bookmakers and punters alike. In order to evaluate the efficiency of player HtH markets, a multivariate approach is used to predict the number of runs scored.

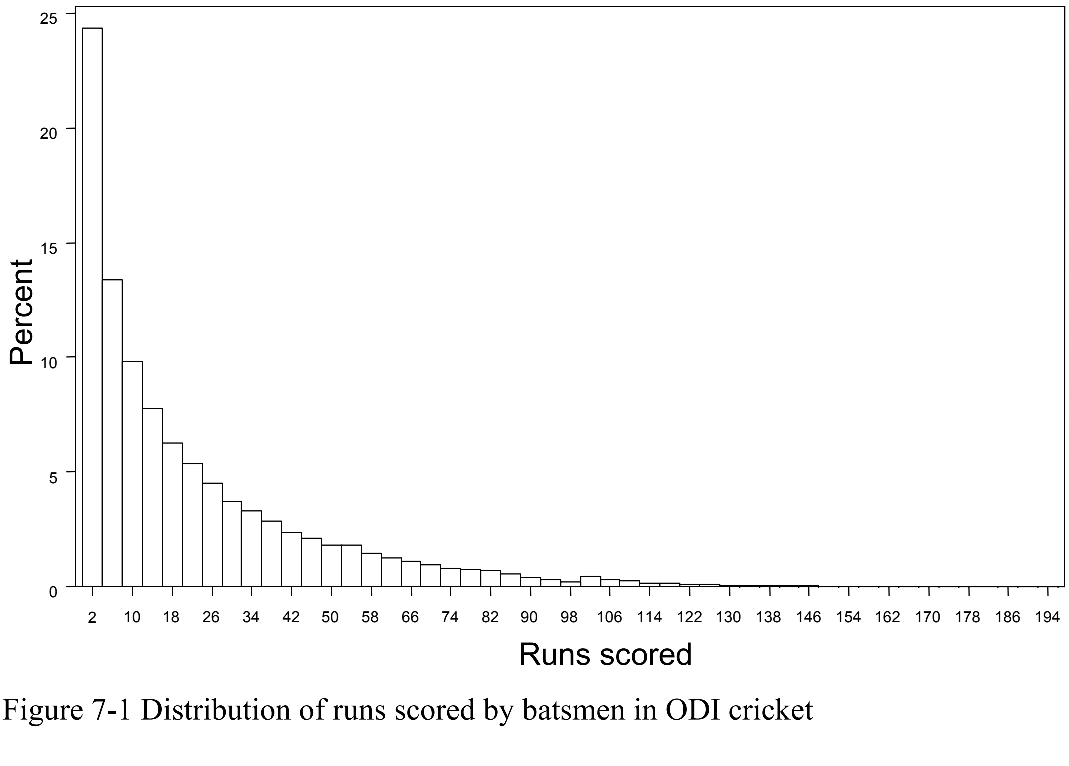

Figure 7-1 Distribution of runs scored by batsmen in ODI cricket

Figure 7-1 clearly shows that the distribution of batsmen scores in ODI cricket does not follow a Normal distribution and confirms previous research that batsmen scores may well follow a geometric distribution. By log-transforming the number of runs scored it is possible to improve the symmetry of the distribution. From Figure 7-2 a training effect for batsmen is apparent with significantly more lower scores present, giving the appearance of a mixture of two distributions. Despite clearly displaying departures from normality, the magnitude of the database and the robustness of a least-squares approach allow for some practical benefit to be gained from the use of a multiple linear regression.

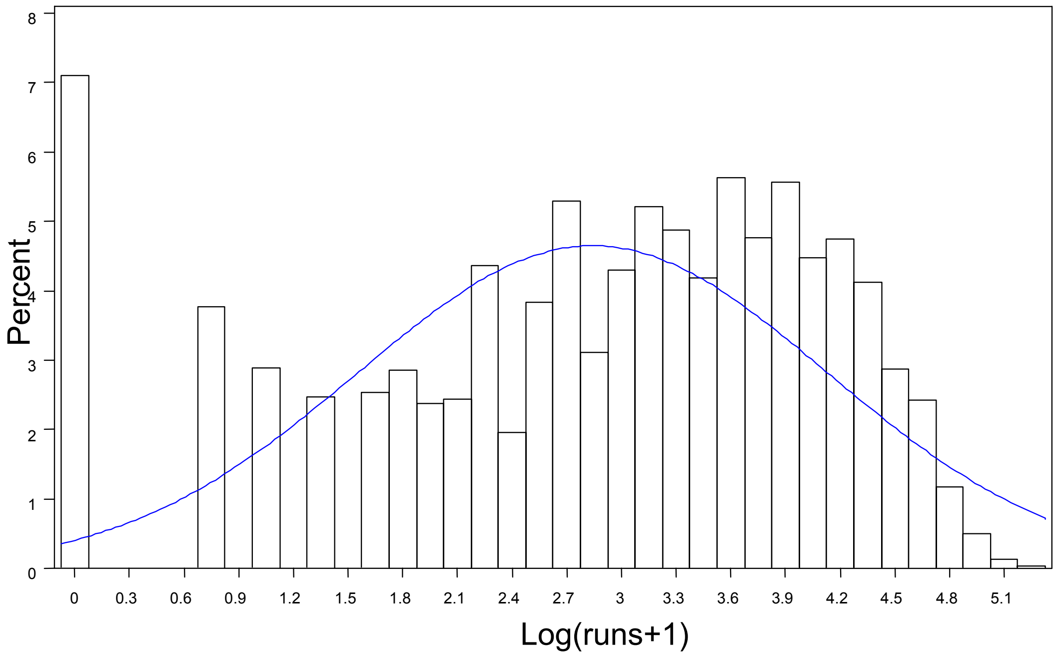

Figure 7-2 Distribution of log(runs+1) for batsmen in ODI cricket

All ODI batting performances prior to the 2003 World Cup were used to identify features thought to affect the number of runs that a batsman will score. Prediction models were constructed to predict scores using a variety of approaches. Rolling averages, exponential smoothing and multiple linear regression models were all used, with goodness of fit determined by comparison of AAE.

Hypothetical match-ups between leading players for past matches were created to explore the best way in which to convert predicted scores into the probability that one player will outscore another. Two alternative methods for converting predicted scores into probabilities were compared by averaging the LogProb of the winning players for each match-up.

Prediction models were applied to 2003 World Cup data and specific bookmaker comparisons calculated. Predicted probabilities were compared with bookmaker prices and potential imbalances determined. Bet sizes were calculated in accordance with the size of the perceived advantage and the probability of winning (see section 3.10). Profitability was determined by calculating the ROI offered by each model. The use of a range of prediction models allowed for exploration into the nature of the relationship between ROI and the quality of the prediction process as measured by AAE.

Because the number of runs scored by batsmen is more closely approximated by a Log-Normal distribution, analysis is conducted using the log(runs+1), with results presented as geometric means with 95% confidence intervals. From the 1936 ODI matches played prior to the 2003 World Cup, there have been 33,606 individual batting performances. These batting performances were used to identify factors having a statistically significant effect on the number of runs scored by each batsman.

7.3 Factors affecting batting performance

7.3.1 Batting position

Given the finite nature of batting time available in ODI cricket, a player's position in the batting line-up is vitally important in determining runs scored. Although batting strategies have evolved over the past 30 years, it is now generally accepted that each team's better players will bat at the top of the order.

When considering the average number of runs scored for each position in the batting order, Figure 7-3 reflects the impact of limited resources.

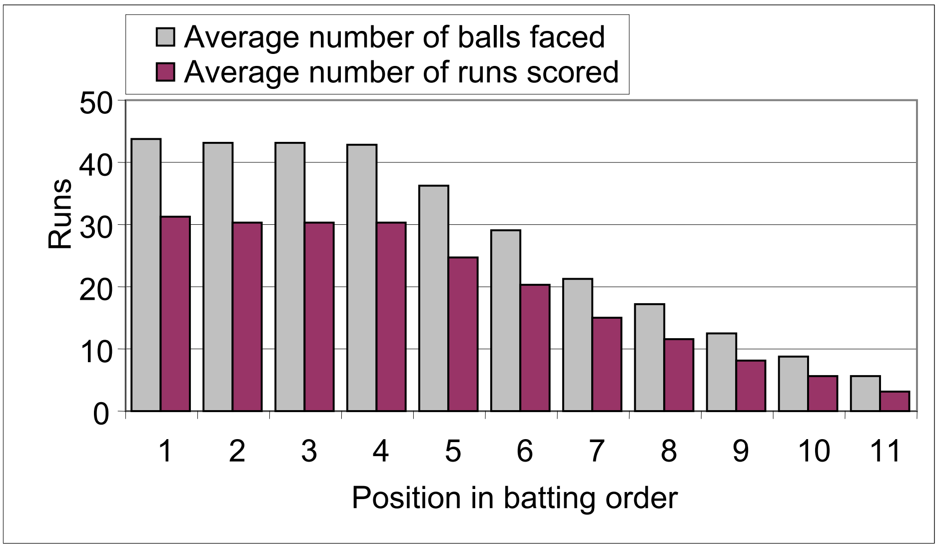

Figure 7-3 Runs scored and balls faced for each position in the batting line up

While a linear like effect is apparent from position five onwards, there was no significant difference for runs scored or balls faced for the first four batsmen from each team. Table 7-1 shows the mean and geometric mean for the number of runs scored per position. Given that resource availability plays such an integral role in determining the number of runs scored and bookmaker HtH prices were for the leading players from each team, further analysis is restricted to only the first four batsmen from each team.

Table 7-1 Average runs for batting position

Order	n	Average runs	Geometric Mean (95%CI)
1	3843	31.4±0.5	16.5 (15.9-17.2)
2	3843	30.3±0.5	15.8 (15.2-16.5)
3	3818	30.3±0.5	16.2 (15.5-16.8)
4	3733	30.2±0.4	17.0 (16.3-17.7)
5	3578	24.6±0.4	14.2 (13.6-14.8)
6	3357	20.4±0.3	12.0 (11.4-12.5)
7	3083	15.0±0.3	8.4 (8.0-8.8)
8	2733	11.6±0.2	6.5 (6.2-6.9)
9	2310	8.1±0.2	4.6 (4.3-4.9)
10	1883	5.7±0.2	3.2 (3.0-3.4)
11	1426	3.0±0.1	1.7 (1.5-1.8)

7.3.2 Experience

Previous chapters suggest experience is a common predictor of performance at both a team and individual level for football and cricket. As can be seen from Figure 7-4, batsmen performance improves at a diminishing rate until a point where the increase in age and subsequent decrease in hand eye coordination outweighs the positive benefits of experience, and performance plateaus out. This result is clearly analogous to performances in AFL football (see Figure 6-3), with the greatest improvement in performance evident at the start of players careers.

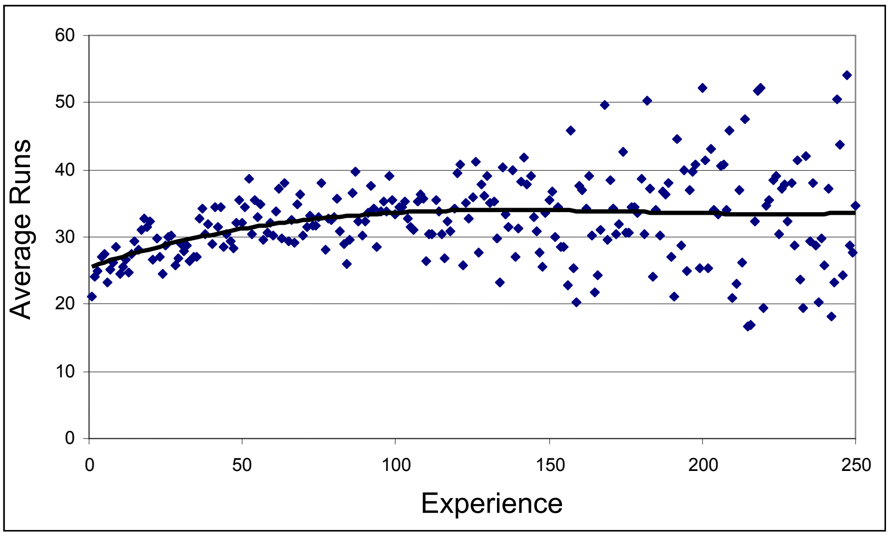

Figure 7-4 Relationship between experience and runs scored

By dividing the number of matches each batsman has played into blocks of 20 and taking the average number of runs scored for each block, it is possible to gain further insight into the effect of experience. From Figure 7-5 we can see that the greatest improvement in performance generally occurs over the first 100 innings, with the average number of runs scored rising from the mid twenties to the low thirties before levelling out.

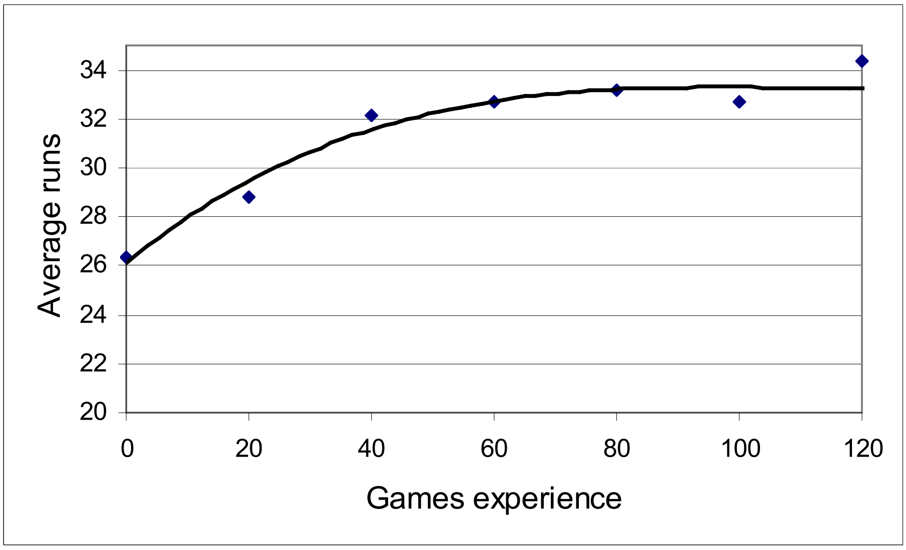

Figure 7-5 Relationship between experience and runs at start of career

7.3.3 Home country advantage

Because ODI cricket is played internationally, any advantage obtained by players playing within their home country, effectively equates to HA. Only 31.7% of all innings played were by batsmen on their home soil. It is not surprising that this figure is quite low, as a large percentage of ODI matches are played in triangular or round robin formats, often held in neutral countries. It can be seen from Table 7-2 that HA for batsmen in ODI matches is highly significant and is equivalent to about three runs per innings when considering the arithmetic mean, but only 1.7 runs when considering a geometric mean.

Table 7-2 Home country advantage for leading batsmen in ODI matches

Batting Location	**N**	**Mean (runs)**	**Geometric mean**	**P-value**
Neutral country	11015	29.1±0.28	15.8 (15.4-16.3)	<.0001
Home country	5124	32.0±0.43	17.5 (16.9-18.2)	

7.3.4 Match time and innings sequence

There are two times at which ODI matches are played. Day games play the first innings prior to lunch and the second innings after lunch, with completion before nightfall. Day/night games commence mid afternoon, break at the end of the first innings for dinner and then resume with the second innings under artificial lighting. Batsmen playing in day matches score slightly more runs than batsmen playing in day/night games, although the difference is not statistically significant (15.9 (15.3-16.4) vs. 16.6 (16.1-17.1) p=0.07).

The role of the team that bats first is to set the highest target possible. In comparison, the team that bats second has a specific target to chase. On average, batsmen make slightly more runs when setting a target in comparison to chasing a target. (16.6 (16.2-17.0) vs. 15.9 (15.2-16.5) p=0.04)

Whilst the time of the match, and the sequence of the innings only appear to have weak effects on the number of runs scored, it can be seen from Table 7-3 that there also appears to be a statistically significant interaction between these two variables, with batsmen chasing runs under artificial lighting averaging about two runs less.

Table 7-3 Interaction between match time and innings sequence

Time	N	Sequence	Mean	Geometric Mean (runs)	P-value
Day	5903	First	30.8±0.4	16.7 (16.1-17.3)	0.007
Day	5677	Second	29.1±0.4	16.5 (15.9-17.1)	
D/N	2304	First	31.9±0.6	16.8 (15.9-17.8)	
D/N	2255	Second	28.1±0.6	14.9 (14.1-15.8)	

7.3.5 Opposition and team effects

Both the quality of the bowlers and the ability of the fielders contribute to how many runs a batsman will score. The quality of the opposition will vary depending upon the composition of the players named in a team. Because the nucleus of a cricket team changes over time, to more accurately determine quality, a comparison of the relative strengths of teams has been restricted to only matches played since the 1999 World Cup. From Figure 7-6 it can be seen that in the preceding four years, Australia has been the hardest team to score runs against.

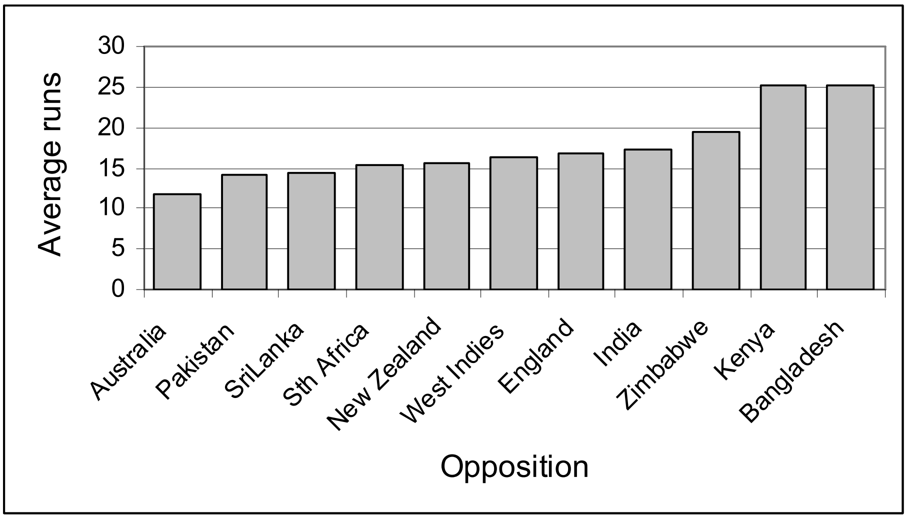

Figure 7-6 Average runs scored per batsmen against each team since 1999

Just as the opposition is a significant predictor of runs scored, the team that each batsman plays for is also a significant predictor of runs. This variable may well reflect different coaching or training techniques used across nations, or may simply be a surrogate marker for quality. Not only have Australian bowlers been the hardest to score runs against, but Australian batsmen have also been the best performed in the four years from 1999 to 2003 (Figure 7-7).

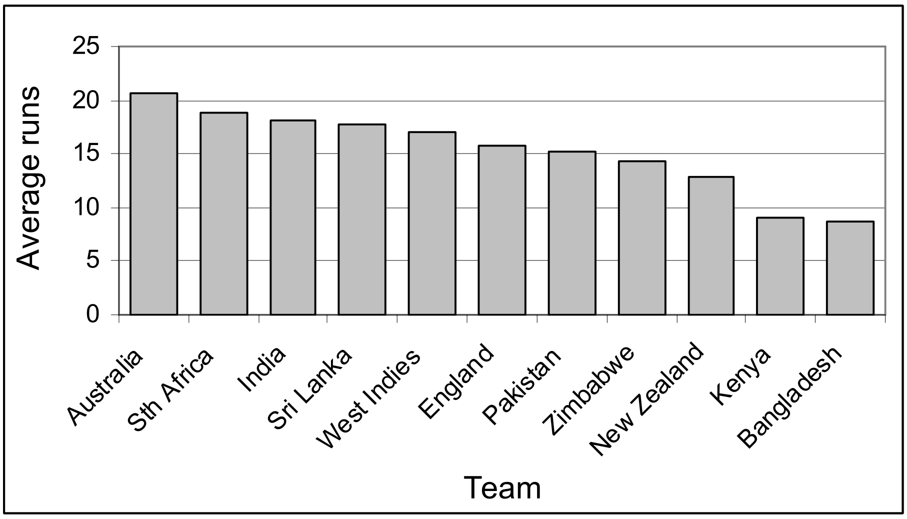

Figure 7-7 Average runs scored per batsmen for each team since 1999

A detailed list of runs scored and runs conceded for each team in the period 1999 to 2003 can be seen in Table 7-4.

Table 7-4 Average runs scored and runs conceded for each team from 1999 - 2003

Team	**Runs scored per batsmen**	**Runs conceded per wicket**
Australia	20.6 (17.7-24.1)	11.8 (10.1-13.8)
Bangladesh	8.7 (6.5-11.6)	25.3 (19.2-33.3)
England	15.8 (13.3-18.9)	16.9 (14.2-20.2)
India	18.0 (15.7-20.7)	17.2 (15.0-19.7)
Kenya	9.1(6.6-12.4)	25.2 (18.6-34.1)
New Zealand	12.8 (11.0-15.0)	15.7 (13.4-18.3)
Pakistan	15.2 (13.2-17.4)	14.3 (12.4-16.4)
Sri Lanka	17.8 (15.5-20.4)	14.5 (12.5-16.7)
South Africa	18.9 (16.4-21.7)	15.3 (13.3-17.6)
West Indies	17.1 (14.5-20.1)	16.3 (13.8-19.2)
Zimbabwe	14.4 (12.4-16.6)	19.5 (16.9-22.5)

Rather than use only fixed constants to indicate opposition and team effects, to further improve the accuracy, an interaction term was explored to allow team effects to vary over time. Although no significant improvement could be found for team ratings, opposition ratings were found to significantly improve the accuracy when allowed to vary linearly with time.

7.3.6 Performance and form

Batsman performance can be measured in various ways. Traditionally, a batsman's career average, defined by runs scored divided by times dismissed, is usually quoted by media outlets as a relative gauge for performance. While Kimber and Hansford (1993) discuss the merits of this, because batting performance in HtH betting is only dependent upon runs scored and not whether a player was dismissed or not, career average is not an ideal predictor of performance. Similarly, because the number of runs scored is better approximated by a log-normal distribution, prediction variables are created to predict the log of runs scored rather than actual runs scored. By taking the exponential of predicted values for the log of runs scored, results can be returned to scale and presented as geometric averages for the prediction of actual number of runs scored.

It is often said that a batsman can be in or out of form - how do we measure this? Does form mean a batsman's last innings, or his last 10 innings? In addition to each batsman's overall geometric average, 10 moving geometric averages (last score, last two scores…last 10 scores) have been used as measures of form. By averaging the absolute difference between the predicted and actual number of runs scored for each individual performance prior to the World Cup, it is possible to compare predictors of form, with the model producing the lowest AAE being the best statistical predictor. From Figure 7-8 we can see that the quality of the predictor is directly related to the numbers of weeks used to create the average, suggesting that any predictor of form is not as efficient as an overall average of performance.

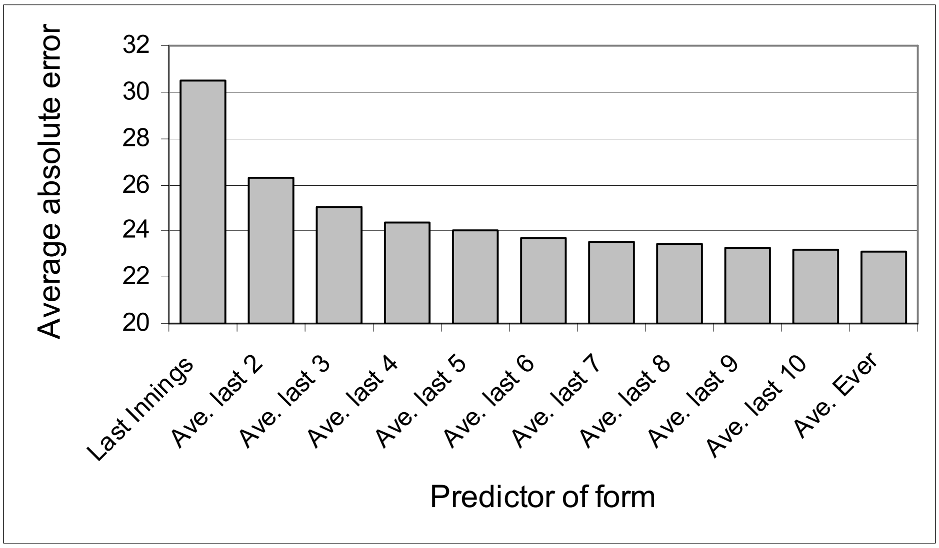

Figure 7-8 Average absolute errors for predictors of form

The process of averaging weights each performance equally, irrespective of when it occurred. An alternative approach is to give more weight to recent performances by exponentially smoothing past scores. It is possible to derive an unbiased prediction for log of runs by using the formula

$$\text{Smoothed score} = (x)\log(\text{actual runs}) + (1-x)\text{Previous smoothed score} \tag{7.1}$$

where x is the smoothing parameter. In the same way that an exponentially smoothed estimate can be obtained by considering all innings played by each player, so too can we derive a smoothed estimate for performance against a specific opposition and performance at a specific venue. Whilst the three smoothing parameters originally chosen ranged from 0.1 to 0.3, it was found that a smoothing parameter of 0.3 was consistently too high to produce accurate results. As a consequence, the three smoothing parameters chosen were (x=0.1, x=0.15 and x=0.2). From Figure 7-9, it can once again be seen that predictors based on all past data produced better results than stratifying the data with regards to opposition or venue. In addition, whilst past averages perform significantly worse than exponential smoothers when stratified by country or venue, there is no significant difference between past averages and exponential smoothing when all past result are used.

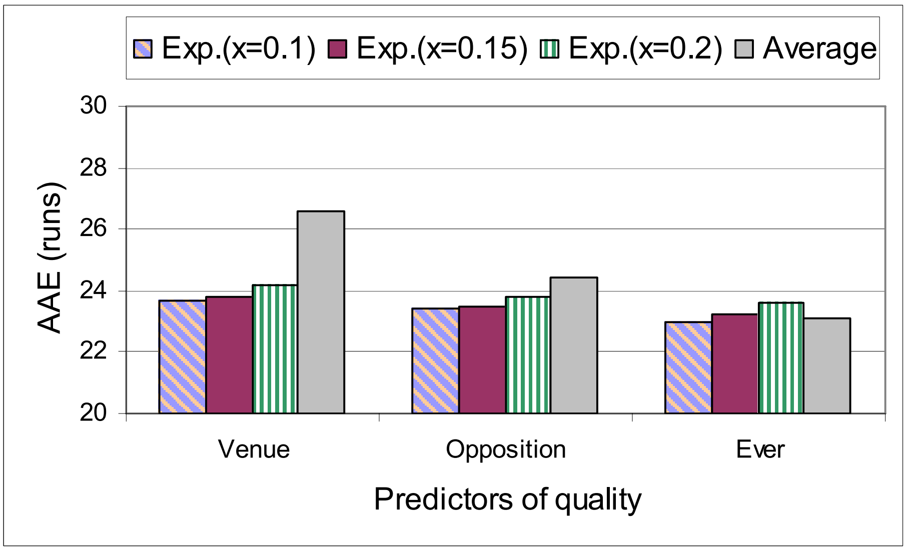

Figure 7-9 Comparison for different predictors of quality

7.4 Multivariate models

Despite batsman scores in cricket not being as predictable as possessions gathered in AFL football, when combined in a multivariate model, nine variables were found to be statistically significant predictors ($p< 0.01$). As seen in Table 7.5, the strongest predictor in the model was an exponentially smoothed predictor for all past performances, followed by a measure of HA. Two other predictors relating to the batsmen were also included, namely the amount of experience gained and each batsmen's past average against the specific opposition. Two match factors were included, namely the batting sequence and whether the match was a day or day/night game; the interaction between batting sequence and time of the match was no longer statistically significant. Categorical parameters relating to the quality of the opposition and the batting team were also highly significant ($p<0.001$).

Table 7-5 Parameter estimates for runs scored by batsmen in ODI matches

Effect	Indicator	Estimate	StdErr	P-value
Intercept		2.55	0.16	<0.0001
Game time	D/N	-0.05	0.02	0.01
Experience		0.0006	0.0002	0.002
Batting on foreign soil	Yes	-0.09	0.02	0.0001
Batting second	Yes	-0.06	0.02	0.005
Exponential ever		0.010	0.00	<0.0001
Team	Australia	0.25	0.05	<0.0001
	Bangladesh	-0.32	0.10	0.0010
	Canada	-0.57	0.45	0.21
	England	0.20	0.06	0.0006
	India	0.17	0.05	0.002
	Kenya	-0.24	0.10	0.02
	Netherlands	-0.40	0.28	0.15
	New Zealand	0.07	0.06	0.19
	Pakistan	0.13	0.05	0.02
	South Africa	0.23	0.06	0.0001
	Scotland	-1.56	0.33	<0.0001
	Sri Lanka	0.12	0.06	0.03
	United Arab Eremites	-0.40	0.27	0.14
	West Indies	0.20	0.05	0.0003
	Zimbabwe	0.0000		
Opposition	Australia	0.03	0.16	0.85
	Bangladesh	0.39	0.32	0.22
	Canada	-0.37	10.00	0.97
	England	-0.09	0.16	0.58
	India	0.10	0.16	0.54
	Kenya	-0.22	0.51	0.66
	Netherlands	0.45	0.92	0.63
	New Zealand	-0.05	0.16	0.75
	Pakistan	-0.02	0.16	0.89
	South Africa	-0.29	0.21	0.17
	Scotland	-10.36	45.61	0.82
	Sri Lanka	0.31	0.17	0.06
	United Arab Eremites	-0.25	3.74	0.95
	West Indies	-0.32	0.16	0.05
	Zimbabwe	0.0000		

Effect	Indicator	Estimate	StdErr	P-value
Opposition* Time	Australia	-0.0002	0.0001	0.001
	Bangladesh	0.0000	0.0002	0.84
	Canada	0.003	0.15	0.99
	England	0.0000	0.0001	0.78
	India	-0.0001	0.0001	0.14
	Kenya	0.0003	0.0003	0.29
	Netherlands	0.0002	0.0007	0.75
	New Zealand	0.0000	0.0001	0.99
	Pakistan	-0.0001	0.0001	0.24
	South Africa	0.0001	0.0001	0.46
	Scotland	0.007	0.03	0.83
	Sri Lanka	-0.0003	0.0001	<0.0001
	United Arab Eremites	0.0007	0.0037	0.84
	West Indies	0.0001	0.0001	0.01
	Zimbabwe	0.0000	0.0001	0.66

To further account for the heterogeneity in runs scored that exists between different teams competing in the World Cup, a second approach was considered. Variables found to be significant in the multivariate model were remodelled at a team level, with different parameter estimates created depending on the relative importance of the past data with respect to each team.

About a half an hour prior to the commencement of each game, captains from opposing teams toss a coin to determine which team will bat first, with the winning captain having the choice. Because it was not always possible or practical to gather this information prior to the placement of bets, a third multivariate model was constructed without the additional benefit of the actual batting sequence.

Although the underlying distribution of runs scored was non-normally distributed, a final linear regression model was created without the benefit of a log-transformation to provide a comparison between transformed and untransformed data.

7.5 Goodness of fit for predicted score

Although past averages and exponential smoothing create unbiased estimates, to accurately measure the goodness of fit for a multivariate model, a hold-out sample separate to that from which the model was constructed, must be used to avoid the bias associated with over fitting.

From the 1936 matches played prior to the World Cup, one quarter of all matches were isolated into a hold-out sample. To avoid heterogeneity due to changes over time, a systematic sampling approach was used with every fourth match allocated into the hold-out sample. AAEs for the four multivariate models and a range of univariate predictors can be seen in Figure 7-10. The original log-linear multivariate model as given in Table 7-5 was the best performing prediction model, with an AAE in the hold-out sample equivalent to 23 runs. The removal of batting time had little impact on the predictive capacity of this model with the 'Early' model also having an AAE of 23 runs. Although the multivariate model constructed at a team level was significantly better performed in the training data set, the AAE in the hold-out sample was about half a run larger reflecting bias due to over fitting. The final multivariate model constructed without the benefit of a log transformation performed better than most univariate models, but was significantly worse than the log-linear approach. At a univariate level, there was little difference between using past average and exponential smoothing to create predictors. The best-performed univariate model was an exponentially smoothed average for performances based on the country in which the match was played. The two worst performing predictors were a batsman's last score, and a batsman's historical average for the venue in which he was playing.

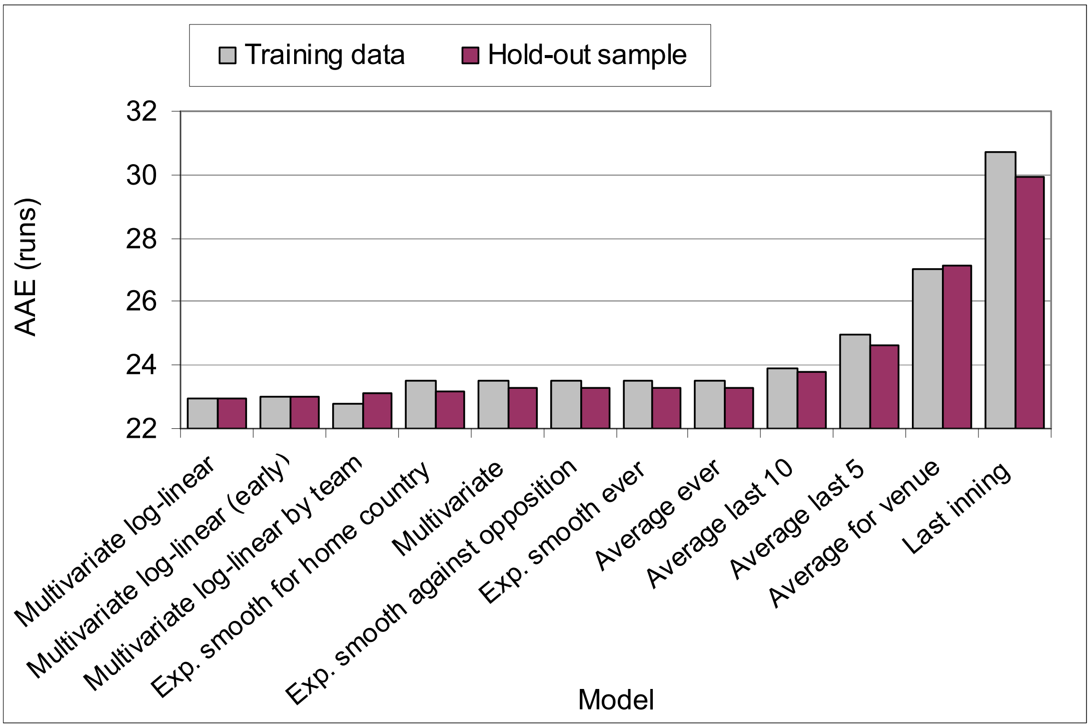

Figure 7-10 AAE for differing prediction models

7.6 Converting predicted scores into probabilities

When investigating the probability of one batsman outscoring another, there may be advantages in assuming that batsmen scores are geometric, with the probability of each score given by

$$\Pr(Score = x) = p\ q^{x}, x = 0,1,2,..........$$

where q = 1-p and p can be thought of as the probability of being dismissed before scoring another run, $\mu = q/p$, $\sigma^2 = q/p^2$, and is fitted using p = 1/ (1+$\bar{x}$). For most top batsman, the mean is at least 30, so q is almost 1. Thus the mean and standard deviation would be approximately the same. It is easily shown the Pr(Score > x) = q^{x+1}, so for two batsman we can calculate the chance one will outscore the other by summing the conditional probabilities the first will outscore the second for each possible score of the first. Thus for two batsman with predicted means and associated parameters p_1 and p_2, we have

$$\Pr(X_2 > X_1) = \sum_{n=0}^{\infty} \mathrm{P}(\mathrm{Score}_1{=}\mathrm{x}).\mathrm{P}(\mathrm{Score}_2{>}\mathrm{x})$$

$$= \sum_{n=0}^{\infty} \mathrm{p}_1\mathrm{q}_1{}^{\mathrm{x}}.\mathrm{q}_2{}^{\mathrm{x}+1}$$

$$= p_1\mathrm{q}_2 \,/\, (1\text{-}q_1q_2)$$

$$= (m) \,/\, (m{+}n{+}1) \qquad (7.2)$$

Similarly

$$\Pr(X_1 > X_2) \;\; = (n) \,/\, (m{+}n{+}1)$$

where n and m represent the predicted means for both batsmen.

This shows that the odds of one batsman outscoring the other are roughly in proportion to their predicted averages. Since departures from the geometric could be expected to affect both batsmen equally, we might expect this relationship to hold approximately.

Alternatively, it can be seen from Figure 7-11 that the difference between scores are approximately normally distributed, so a player's probability of winning can readily be determined by dividing the predicted difference between players by the combined standard deviation for the two players, and comparing with a standard Normal distribution. This approach was previously outlined in section 6.8.

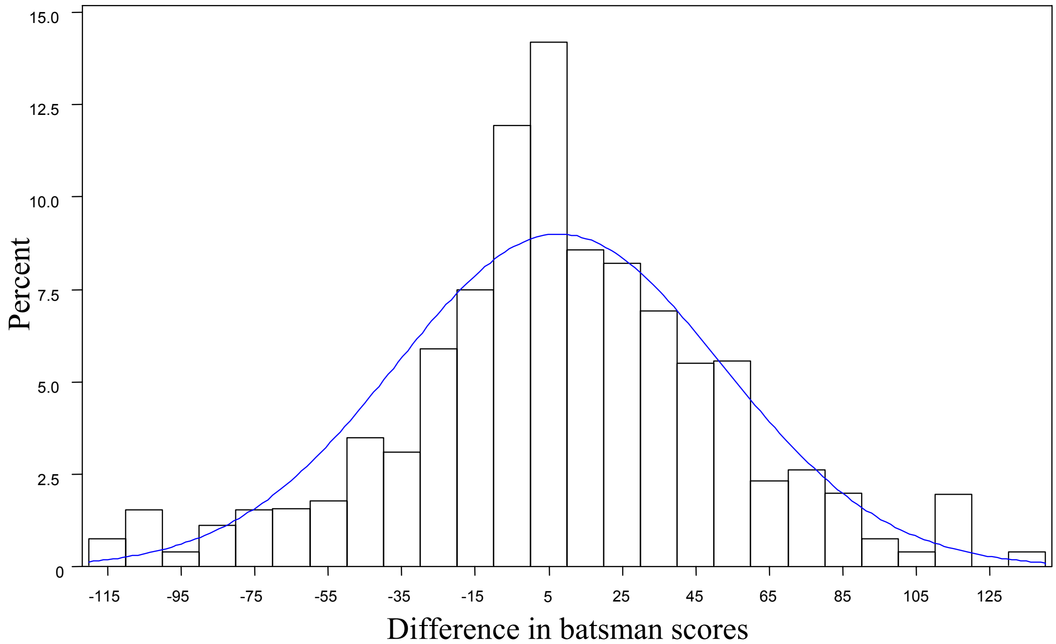

Figure 7-11 Distribution of differences in scores for matched batsmen

7.6.1 Comparison of conversion approaches

Over 50,000 hypothetical match-ups were created by matching the leading four batsmen for both teams for the 1936 matches played prior to the 2003 World cup. By using the 12 prediction approaches shown in Figure 7-10 and summing the log of predicted probabilities for winning players, a measure of goodness of fit could be determined for over 600,000 comparisons. Using a sign rank test to compare between the two methods, a conversion process based solely on the predicted scores of both players as given by equation (7.2) was statistically significantly better than a parametric approach of comparing against a Normal distribution curve. (LogProb(method A) – LogProb(method B) = 0.005, $p<0.0001$). Whilst the large number of comparisons provided statistical evidence to the superiority of one method over another, in reality there was little practical difference between approaches.

7.7 HtH prices for the 2003 World Cup

From the 50 completed matches in the 2003 World Cup, 309 player HtH batting match-ups were collected from three internet sports bookmakers. On 46 occasions either one or both of the competing batsmen failed to reach the crease, leaving a database of 263 HtH match-ups. One hundred and forty seven match-ups (56%) involved players from the same team, as opposed to match-ups from players from competing teams (44%). Prices ranged from $1.50 to $2.40 and were approximately normally distributed with a mean and median of $1.90, and a standard deviation of $0.10. The bookmaker's percentage was approximately seven percent.

7.8 Results

Using a wagering strategy based on the probability of success and the size of the perceived market imbalance as outlined in section 3.10, the number of bets, average bet size and ROI was calculated.

For example, in the World Cup final between India and Australia, the Indian captain Sourav Ganguly ($1.80) was matched against fellow Indian Mohamed Kaif ($2.00) in a Head to Head batting match-up. The multivariate log-linear prediction model predicted Ganguly to make 22 runs whilst Kaif was predicted to make 15 runs. Using equation (7.2),

Pr (Ganguly > Kaif) = (22) / (22+15+1) = 0.58

From (2.1) the perceived advantage $\boldsymbol{A}$ = (0.58*$1.80)-1 = 0.042

From (2.2) the given bet size $\boldsymbol{B}$=0.042/(1.8-1)=0.053

Based on a fixed bank size of $1000 dollars, a bet size of 0.053 *$1000 = $53 was wagered on Ganguly to outscore Kaif. Ganguly actually made 24 runs whilst Kaif made 0, resulting in profit from this wager equal to $53 * (1.8-1) = $42.4.

In addition to the four different multivariate models discussed in section 7.4, a range of univariate prediction models were also considered, allowing for a relationship to be established between goodness of fit and ROI.

Table 7-6 Results from 17 prediction models sorted by return on investment

Model	**n**	**Average bet size ($)**	**Average profit ($)**	**P-value***	**ROI**	**AAE**
Multi Log-linear	108	85±08	24±09	0.001	35%	22.9±0.2
Multi Log-linear Team	137	157±14	37±17	0.01	33%	22.7±0.2
Multi Linear	67	58±07	11±08	0.26	25%	23.5±0.2
Multi Log-linear Early	106	79±07	15±09	0.05	23%	22.9±0.2
Average last 5	223	315±12	5±21	0.69	2%	24.0±0.2
Exponential opposition	199	226±12	-1±17	0.77	-1%	23.5±0.2
Average last 3	239	353±14	-10±23	0.79	-4%	25.1±0.2
Average last 7	219	265±11	-12±18	0.64	-5%	23.6±0.2
Average ever	116	91±08	-4±09	0.88	-7%	23.1±0.2
Exponential ever	204	232±10	-17±17	0.42	-8%	23.3±0.2
Exponential country	212	262±13	-18±18	0.33	-9%	23.4±0.2
Average last 9	217	226±11	-29±16	0.10	-15%	23.4±0.2
Last innings	246	547±15	-74±33	0.01	-20%	30.6±0.2
Average country	198	242±14	-36±18	0.12	-20%	24.5±0.2
Average opposition	191	219±14	-28±18	0.26	-20%	24.4±0.2
Exponential venue	128	225±17	-49±20	0.04	-31%	23.9±0.2
Average venue	139	434±24	-90±36	0.02	-33%	26.6±0.3
*Probability that average profit per bet is significantly different from zero						

From Table 7-6 it can be seen that the average of the last five innings was the sole univariate predictor to produce a positive ROI. In contrast, all four multivariate models produced a ROI in excess of 20%. The most productive of the multivariate models was the log-linear model with a ROI of 35% and an average profit per bet significantly greater than zero (p=0.001). When parameter estimates for the log-linear model were derived at a team level, the resulting ROI was similar (33%) although the number of bets placed and the average bet size were much larger. The loss of information concerning which team would bat first reduced the ROI from 36% down to 25%, although the resulting profit per bet was still significantly greater than zero (p=0.05). Despite violating the underlying assumption of normality, the linear regression model was clearly better performed than any univariate prediction methods. While producing a ROI of 23%, the linear model was not as sensitive in identifying market imbalances, with only 67 betting opportunities identified.

The process of averaging the log of the probabilities for the predicted winner (see section 3.7), provides a useful measure to the quality of the prediction process. Figure 7-12 indicates a strong linear relationship to exist between the AAE in the training data and the goodness of fit for the 263 batsmen match-ups. This finding validates the use of a multivariate modelling approach by establishing a link between past and future data.

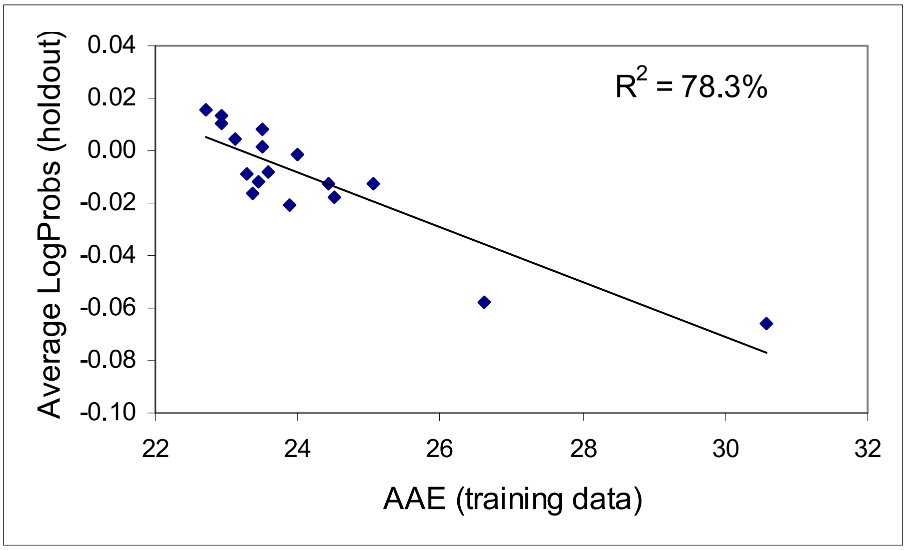

Figure 7-12 Relationship between AAE and goodness of fit (LogProb)

It follows that the more accurate the process of assigning probabilities to batsmen, the greater the ROI. This can be seen from Figure 7-13 in the direct relationship that exists between ROI and the average log of the probability of the winning players.

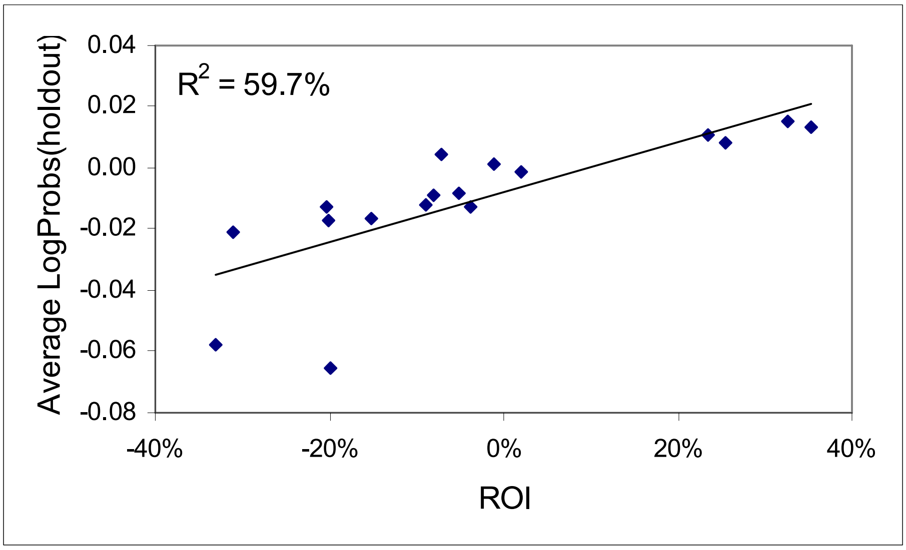

Figure 7-13 Relationship between goodness of fit (LogProbs) and ROI

The relationship between ROI in the training data and goodness of fit in the holdout sample can be viewed directly by examining the relationship between AAE and ROI. Figure 7-14 shows that the models that produce the lowest AAE in the training data consistently produce the highest ROIs. The negative relationship between ROI and AAE suggests that the models that produce the smallest absolute errors for the data prior to World Cup also produce the best ROI for HtH match-ups during the World Cup. The implication of this relationship allows for the construction and improvement of prediction models without the necessity of large amounts of information on bookmaker prices.

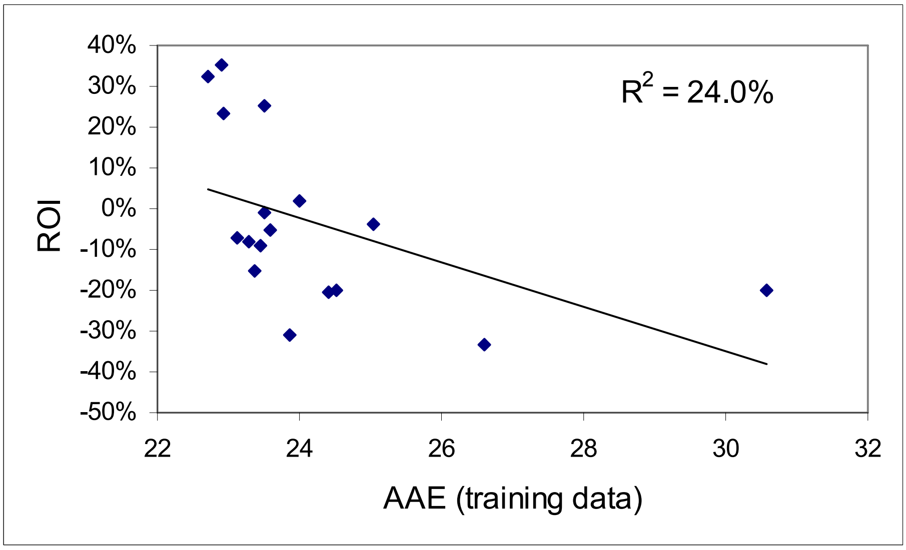

Figure 7-14 Relationship between ROI and AAE

7.9 Discussion

Contrary to popular perception, price setting by bookmakers for sporting outcomes is not an exact science, but rather a combination of experience, knowledge and intuition. With the use of all ODI matches ever played, it is possible to identify features that are known to affect batsman performance. The combination of variables in a multivariate model enables a process of comparison between batsmen that is clearly superior to that offered by bookmakers.

Although ROIs in excess of 20% were achievable in theory, in practice this figure was less. Using a log-linear multinomial approach, 99 bets were actually placed by the author during the World Cup, with a ROI of 16%. This difference between theoretical and practical profit can be primarily attributed to the time delay between the recording of prices and the placing of bets. Bookmaker prices were generally recorded about 24 hours prior to the match, whilst bets were usually placed in the final hours before the start of each game. Because of the dynamic nature of betting markets in which prices change in accordance with supply and demand, the greatest inefficiencies found in the market occur when prices are initially posted. It was often the case that larger inefficiencies identified by the multivariate model were also identified by the general public resulting in a reduction in price and a subsequent reduction in profit.

Obvious differences exist between modelling individual player performance for AFL football and batsman performance in ODI cricket. The number of possessions gathered by AFL footballers is better approximated by a Normal distribution and is more predictable than the number of runs scored by batsmen in ODI matches. This is reflected by the fact that past data can explain almost three times as much variation in AFL performance compared to batsmen in ODI matches. Whilst a multivariate approach is better suited to a more predictable outcome, results from this chapter suggest a practical benefit can be obtained when comparing the performance of two batsmen in ODI cricket.

In combination, these results suggest the use of statistically driven prediction models could readily be applied to predict individual performances for other sporting events. Should sufficient past data exist the relative predictability of a sporting outcome can be determined. A common theme to date has been an underlying assumption of normality, enabling an R-square statistic to provide a measure of predictability.

Determination of the relative predictability of a sport can be seen to have practical benefits for both punters and bookmakers. While a knowledgeable punter might prefer to operate in a more predictable betting market, the bookmaker would prefer to operate in a market where there was less opportunity to exploit inefficiency.

Individual performance in AFL football is clearly the most predictable outcome with over 37% of the variation in possessions explained through the use of past data. Match outcomes of AFL and ODI cricket are the next most predictable with R-square figures of 22% and 20% respectively, whilst player performance in ODI cricket was the least predictable with only 13% of the variation explained through the use of past data. Whilst some sports are more predictable than others, in each case the majority of variation in the outcome is left unexplained. This could be partly attributed to psychological, physical or external factors all of which are difficult to measure. Given the large number of variables considered in this thesis, it is apparent that in football and cricket, a large amount of variation will always remain unexplained. This is what makes sporting events particularly interesting to the viewing public.

Whilst the amount of variation that can be explained by modelling may in general appear to be rather low, these results clearly show that only small improvements in R-square are required to significantly increase ROI. This in turn suggests that ROI is more influenced by the accuracy of the bookmaker than by the predictability of the sporting outcome.

While the use of statistically driven prediction models for team and player performance is of benefit, the application thus far has been restricted by an underlying assumption of normality. In order to establish a wider applicability, alternative outcomes, once again from AFL football and ODI cricket are considered over the next two chapters. Firstly, player and match information will be used to predict the three best players for each AFL match and secondly, previous match information will be used to predict the number of runs scored per over in ODI cricket.

8 Predicting the Brownlow medal winner

8.1 Introduction

The Brownlow medal is the highest individual honour that can be bestowed upon an AFL footballer. In each of the 176 home and away matches for a season, votes are assigned to the three best players (3 – first, 2 – second, 1- third) by the umpires that preside over the game. With the use of an ordinal logistic regression model retrospectively applied to past data, it is possible to identify specific player and match statistics that can aid in the prediction of who will poll votes. Using various measures of goodness of fit, the difference between statistical significance and practical significance is explored. By varying the size of the training data and holdout samples it is possible to determine the optimal size for training data, along with measuring the detrimental effects of overfitting the data. By applying this model to present data it is possible to objectively assign leading players a probability of winning the Brownlow medal. The author has successfully used this approach to identify the leading contenders for each Brownlow medal count since the 2000 season.

8.2 Background

During the 2000 AFL football season, discussion arose as to the best possible way to predict the winner of the Brownlow medal, both before and during the actual count. With a large amount of match performance statistics readily available it was felt that a mathematical modelling process might well assist in the objective assignment of a player's probability of polling votes. Based on data collected from the 1997, 1998, and 1999 seasons, an ordinal logistic regression model was constructed and applied to the 2000 season. Predicted votes for each match were then tallied over the season to provide players predicted totals for the year. During the course of the 2000 Brownlow count, Swinburne Sports Statistics Department provided updated online predictions that combined predicted and actual totals throughout the course of the evening.

Following considerable success and media attention (Anon (2000)), the ordinal regression model was further enhanced for the 2001 season. With the addition of an extra year of data and more statistically significant variables identified, this modelling process was able to clearly identify the three leading candidates for the 2001 Brownlow medal, and was widely publicised prior to the count by Bailey (2001).

Further improvements continued to be added to the modelling process by Bailey and Clarke (2002), and although 25 variables were now identified as being statistically significant predictors, the practical benefit of some variables came into question. Additionally, it was wondered just how much data was optimally required to accurately develop such a model. To answer these questions, a range of practical measures were developed, and with seven complete seasons of data, a comprehensive analysis was conducted exploring the practical benefit of each variable. Training datasets of differing sizes were further incorporated to ascertain the optimal amount of data required to build such models.

8.3 Database

A database was constructed that comprised of data collected from each regular season AFL match played between 1997 and 2003 (1232 games). For each game, an array of individual match statistics is readily available, both in the newspapers and via the internet. An example of what is available can be seen in Table 8-1. Whilst some predictors of votes are created from the past history of the players, most predictors are derived from match statistics.

Table 8-1 Example of individual match statistics available on the Internet

NAME	TK	TH	DI	RE	IN50	MA	HO	CL	TO	FF	FA	TK	G
Akermanis, J	9	10	19	1	0	5	1	2	1	1	0	0	1
Ashcroft, M	10	8	18	1	2	2	0	9	0	1	1	0	0
Bolton, C	7	2	9	3	2	1	0	1	1	0	0	0	0
Bradshaw, D	4	1	5	1	0	0	0	0	1	1	3	0	3
Champion, R	5	2	7	3	0	3	0	0	3	1	2	3	1
Cupido, D	2	2	4	2	0	0	0	0	2	0	2	2	1
Hart, S	6	5	11	3	1	0	0	1	0	0	0	1	0
Key: TK = total kicks, TH = total handballs, DI = disposals, RE = rebounds, IN50 = times inside 50 zone, MA = marks, HO = hit outs, CL = clearances, TO = turnovers, FF = frees for, FA = frees against, TK = tackles, G = goals													

8.4 Predictors for Brownlow voting

8.4.1 Disposals

The number of disposals (kicks + handballs) that a player accumulates during the course of a match is the strongest predictor for polling votes. This is reflected in the fact that the leading possession winner for each match has a 51% chance of polling votes, with the leading possession winner of the winning side having a 64% chance of polling votes.

8.4.2 Won the game / MOV

Umpires have a clear tendency to award votes to the winning side, with 92% of 3-votes being awarded to a player from the winning side. Similarly, 83% of 2-votes and 76% of

1-votes are awarded to players from the winning side. The MOV is also important in determining the voting probabilities.

8.4.3 Hit Outs

Ruckmen have traditionally polled well in the Brownlow medal, as a good ruckman can have a big influence on the match outcome without gathering a lot of possessions. Not surprisingly, the number of hit outs that a ruckman has during a match is a strong predictor for polling votes.

8.4.4 Standout

Based on simple match statistics, there can often be a player that most would agree was the best performed player on the day. Using a second multiple ordinal regression model with match result, disposals, goals and hit outs as predictors, each player was assigned a predicted vote total for the game (3*pr(3votes) + 2*pr(2votes) + pr(1vote)). By ranking players according to their predicted vote total, the difference between consecutive vote totals could be calculated. The larger the gap between each player and the player immediately below him the more likely a player was to poll votes.

8.4.5 Best players

Any prediction model based solely on player statistics would fail to take into account how well a player has performed in comparison to his direct opponent, which would in turn clearly bias against defenders. To compensate against this, the best players (as given by the AFL website) are included in the model, and are a significant predictor for votes, with only 12% of players awarded 3-votes not being named in their team's best six players. Similarly, the order in which players are named is also of importance. This can be seen from Figure 8-1, with the best named player from the winning side receiving an average of 1.5 votes, whereas the best named player from the losing side only receives an average of 0.4 votes.

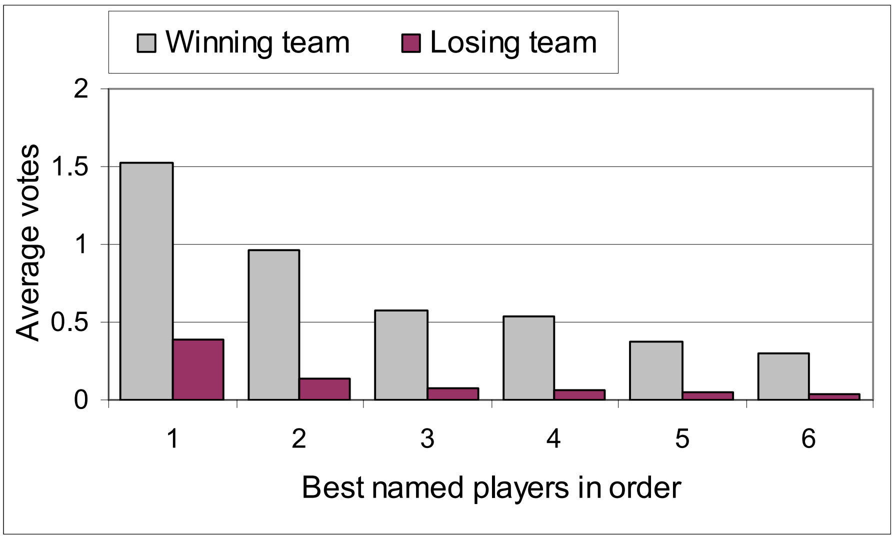

Figure 8-1 Average number of votes polled for the best named players in order

8.4.6 Quality of disposals

Although the number of disposals that a player collects is the strongest predictor for votes, this variable fails to take into account the quality of the disposals. As the quality of a disposal is subjective it will always be difficult to accurately measure. Rather than measure the quality of the disposal, the quality of the player having the disposal has been measured. By constructing a prediction model based solely on the number of disposals, it is possible to identify players that poll significantly more votes in a match than quantity alone would predict. By then measuring the proportion of times that this had previously occurred for each player, it is possible to gain an indirect measurement of the quality of the player in question (Good average). We can also use the same approach to measure the proportion of times each player has not polled votes when disposals alone suggest that he should (Bad average). This helps to accounts for players who gather a lot of possessions but may not dispose of the ball very well.

8.4.7 Goals

Not surprisingly, the number of goals that a player kicks during the course of a match is a significant predictor for votes. What is surprising is that the number of goals does not carry as much weight as some may think, with only 40% of players kicking five goals for the game managing to poll votes! If a player kicks six goals, his chance of polling increase to 62%, and with seven goals, chances of polling increase to 79%. Only one player in the past seven years has kicked eight or more goals and not been awarded at least one vote. The average number of votes polled for corresponding goal totals can be seen from Figure 8-2.

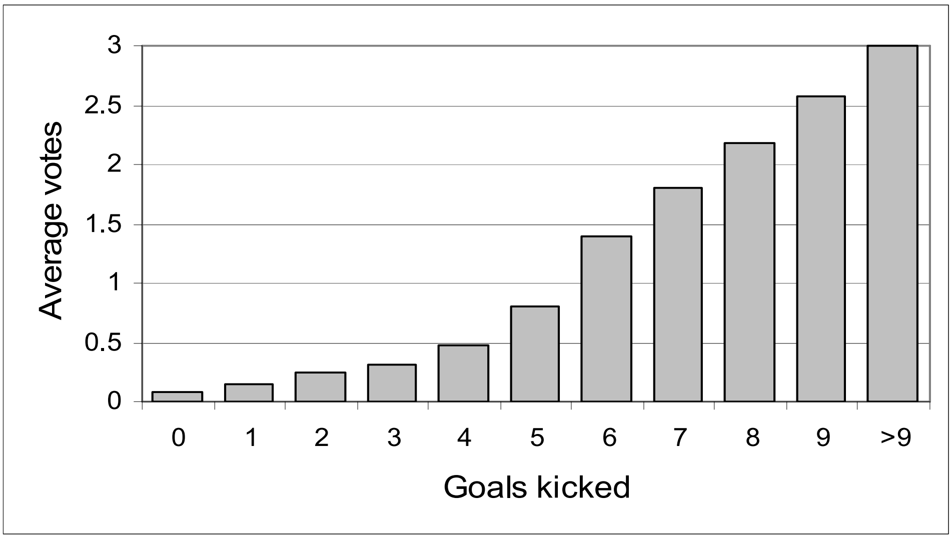

Figure 8-2 Relationship between polling votes & kicking goals

8.4.8 Marks

Whenever a player catches a ball delivered by foot that has travelled more than 15 metres and remained untouched by other players, he is awarded a mark. As can be seen from Figure 8-3, the more often a player marks the ball throughout the course of the game, the more likely he is to poll votes.

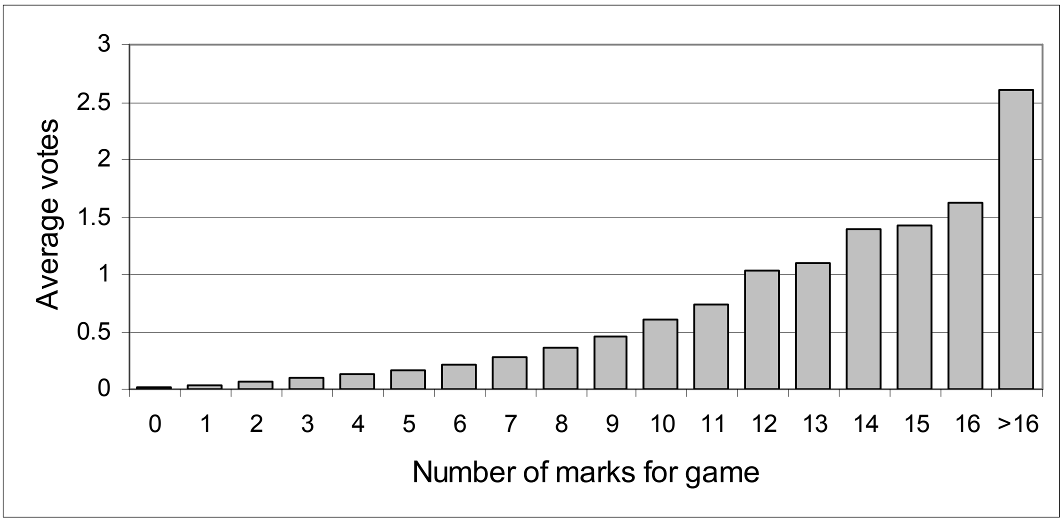

Figure 8-3 Relationship between polling votes and taking marks

8.4.9 Rebounds / Frees for / Tackles / Inside 50

Beside the number of disposals, goals and marks, there are various other player statistics available that can help to quantify player performance.

- Rebounds – the number of times a player repels the ball from defence into attack.
- Frees for – the number of free kicks awarded to a player.
- Tackles – the number of times a player successfully applies a tackle to an opposition player.
- Inside 50 – the number of times that a player propels the ball inside the forward 50 metre arc.

Although only having small effects on a player's probability of polling votes, all of the above four variables are statistically significant predictors.

8.4.10 Position

The position that a player plays in has a significant bearing on votes. Because player positions can change throughout the course of the match, each player's named starting position was used for analysis. From Figure 8-4 it can be seen that on-ball players poll the most votes followed by centre-field players. Players named on the interchange and players named in the backline poll the least number of votes.

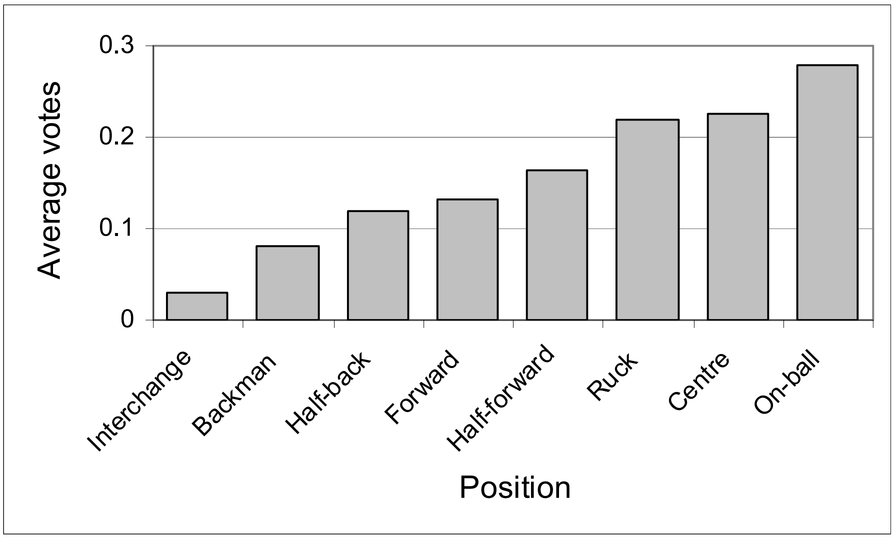

Figure 8-4 Relationship between polling votes and player position

8.4.11 Distinct appearance

Players with distinctive appearance average twice as many votes per game than their non-distinct counterparts (0.24 votes per game vs. 0.12 votes per game $p<0.0001$). Distinctive appearance has been quantified as any player with red or blond hair, or a significantly darker or lighter skin colour.

8.4.12 Team scoring shots

If a team has a lot of scoring shots on goal, this acts as an indirect measure of the number of players in the side that are playing well as both midfielders and forwards are required to be successful for shots on goal to eventuate. Consequently, a greater number of scoring shots reflects more players playing well. With votes spread amongst a greater number of good players, it becomes more difficult for any individual player to poll votes.

8.4.13 Captain

The captain of each club is significantly more likely than all other players to poll votes. (0.45 votes per game vs. 0.12 votes per game $p<0.0001$) At the start of each match, opposing captains meet the umpires and toss a coin to determine choice of direction. Similarly, should players have problems throughout the course of the match; it is the captain's role to liaise with umpires. Even given that captains do have more contact with the umpires than all other players, the primary reason why captains poll more votes is because team captaincy acts as a surrogate marker for player quality with the captain normally one of the best players in the team.

8.4.14 Past votes polled

Another marker of player quality is the average number of Brownlow votes that a player had previously polled prior to the start of the match. The higher the average number of votes polled per game in the past, the more likely a player is to poll in the future.

8.5 Multivariate model

Using the number of votes polled as the outcome, a multiple ordinal logistic regression was constructed. Seven seasons with 176 matches per season and 44 players per match (7744 data points per season) were used to progressively construct a 25-variable multivariate model with all variables in the model statistically significant at a level of $p<0.001$. While backward elimination selection techniques were also used in model construction, to show the relative importance of each variable, results are presented and graphed in stepwise fashion. From Table 8-2 the 25 stages of development can be seen.

In addition to the 19 first order effects that were identified, there were six interactions between variables that were found to be statistically significant ($p<0.001$). Each interaction term was comprised of a first order effects, suggesting that the interaction term was acting as a fine tuning process for relationships that were not perfectly linear.

Goodness of fit can be determined at two different levels relating to the prediction of votes for each game, or the prediction of the votes that each player will poll for the season.

Table 8-2 Goodness of fit for the 25 stages of the multivariate model

Stage	Model	Parameter Estimate*	Chi-square	P-value
1	Disposals	-.08±0.003	549	<0.0001
2	+ Result	-0.02±0.001	235	<0.0001
3	+ Win	-0.35±0.04	68	<0.0001
4	+ (Result*Win)	0.02±0.001	167	<0.0001
5	+ Hit outs	-0.02±0.002	104	<0.0001
6	+ Standout	-0.87±0.08	119	<0.0001
7	+ Best players	-0.20±0.02	132	<0.0001
8	+ Good average	-3.06±0.48	41	<0.0001
9	+ Bad average	0.55±0.17	11	0.001
10	+ (Good*Bad)	5.42±1.51	13	0.0003
11	+ Goals	-0.12±0.02	45	<0.0001
12	+ Marks	-0.02±0.005	11	0.0009
13	+ (Marks*Goals)	-0.03±0.004	36	<0.0001
14	+ (Disposals*Goals)	0.001±0.0001	26	<0.0001
15	+ Position	Categorical	28	0.0004
16	+ (Disposal*Position)	Categorical	16	0.001
17	+ Inside 50	-0.03±0.005	22	<0.0001
18	+ Scoring Shots	0.01±0.003	26	<0.0001
19	+ Distinct	-0.12±0.03	17	<0.0001
20	+ Captain	-0.21±0.04	25	<0.0001
21	+ Average votes	0.64±0.21	11	0.001
22	+ Tackles	-0.02±0.01	13	0.0003
23	+ Rebounds	-0.02±0.01	12	0.001
24	+ Frees for	-0.03±0.01	11	0.001
25	+ (Disposal*Best)	0.004±0.001	22	<0.0001

*Parameter estimates are reported for the final 25 variable model

8.6 Goodness of fit within the game

8.6.1 Log likelihood

When fitting an ordinal logistic regression, a clear guide to the goodness of fit of the model can be gauged from the log likelihood estimate. Because a change in -2log likelihood can be well approximated by a chi-square distribution, the statistical significance of additional variables can readily be determined. By examining Figure 8-5 from left to right, we can see the corresponding reduction in -2log likelihood as each new variable is added to the model, with the greatest improvements occurring over the first seven stages of the model.

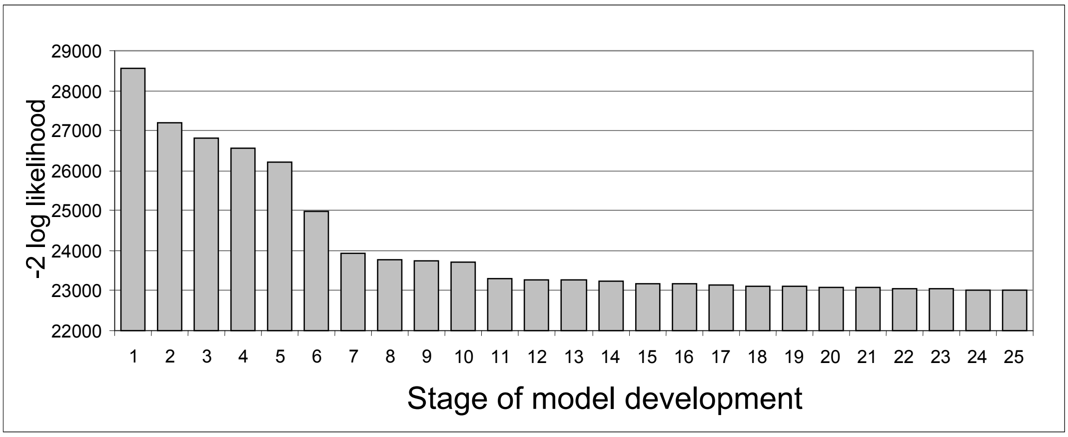

Figure 8-5 -2 log likelihood for each stage of model development

8.6.2 Average Rank

Using the 25-variable multivariate ordinal regression model, each player was assigned a probability of polling three votes, two votes or one vote for each match. A predicted value for the number of votes that each player was expected to get for each game was created by the following formula -

$$Predicted\ Votes = 3 * Pr(3\ votes) + 2 * Pr(2\ votes) + Pr(1\ vote) \quad (8.1)$$

Based on the predicted number of votes, the 44 players for each match could then be ranked in accordance with their predicted vote total. The average rank of players that poll votes provides a practical way to measure the goodness of fit for models, with a lower rank indicating a better fit to the data. For example, from Table 8-2 it can be seen that using a model with only disposals as a predictor, the average rank of all players that polled three votes was about six. With the full 25-variable model the average rank of players that poll three votes is only 3.2.

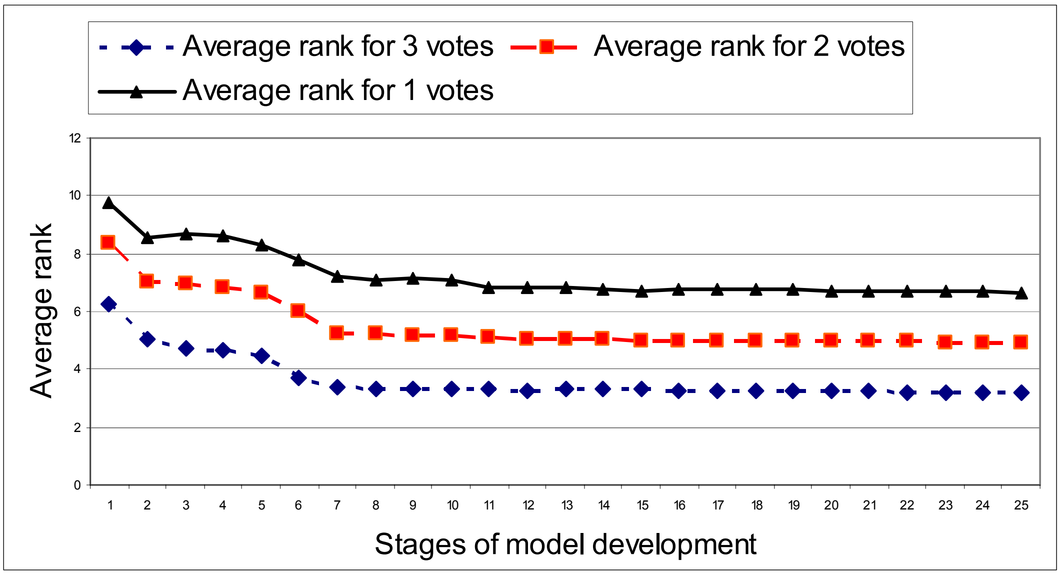

Figure 8-6 Average rank for vote getters as variables are added to the model

8.6.3 Top three for the game

Having ranked players for each match according to their predicted vote total, we can then measure how often the leading ranked player is awarded three votes. From Figure 8-7 we can see that the leading ranked player actually polls three votes, 43% of the time. Similarly, the leading ranked player has a 73% chance of polling any votes (three, two or one). Conversely, the player who polls three votes will be ranked within the top three players in 72% of all games. Once again, there is a clear indication from Figure 8-7 that the majority of improvement occurs over the first 10 stages of model development, with negligible improvement after that.

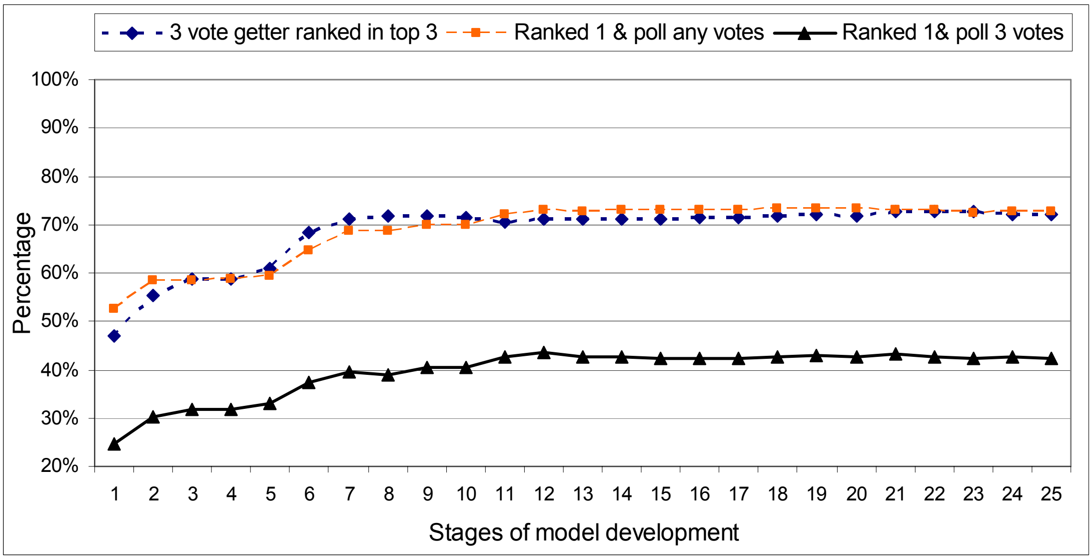

Figure 8-7 Within game goodness of fit measures for stages of model development

In addition to determining goodness of fit at a match level, players predicted vote totals can be aggregated to provide player predictions for the season.

8.7 Goodness of fit for season

8.7.1 Average absolute error

By tallying each player's predicted number of votes for each match as given by equation 8.1, and aggregating for the 22 matches for the season, it was possible to derive a predicted Brownlow total for each player for the year. A simple measure of goodness of fit can then be derived by measuring the AAE between the predicted total and the actual total for each player for the season. When predicting the yearly total for players, of particular interest is the performance of the players most likely to win the Brownlow medal rather than all players, thus the AAE for the leading 20 predicted players is also considered.

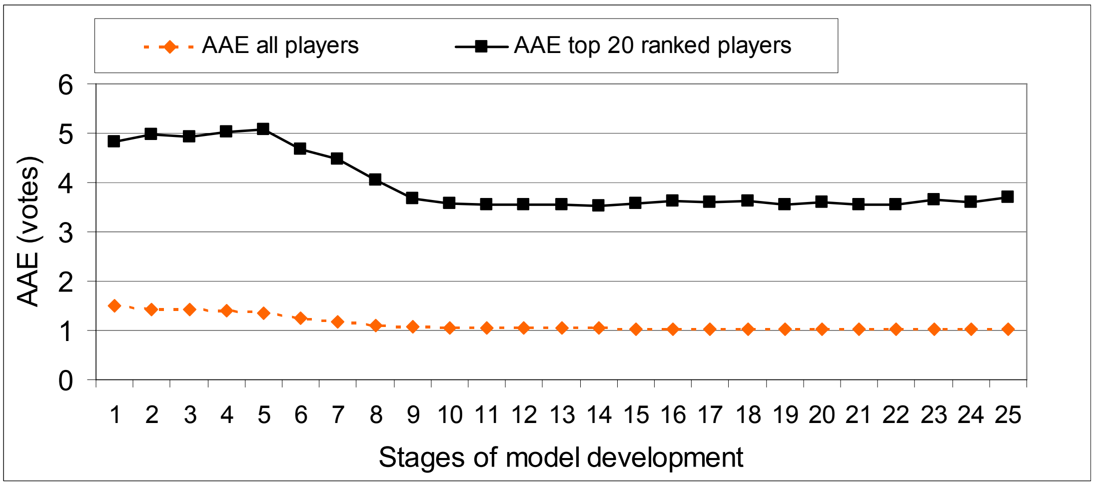

Figure 8-8 AAE for seasonal vote totals for 25 stages of model development

From Figure 8-8 we can see that most additional benefit can be achieved with the first 10 of the 25 statistically significant predictors. The annual AEE for all players can be reduced to approximately one vote, but there is higher variability in the leading players with an AEE of about three and a half votes. Based on the AAE, it is then possible to determine the proportion of players that the modelling approach can accurately predict to within one vote. From Figure 8-9, we can see that the modelling approach tops out after the addition of about 10 variables, successfully predicting 68% of all players to within one vote of the actual total. Similarly, Figure 8-9 also shows that the model can successfully predict 83% of players within two votes, and 90% of players within three votes.

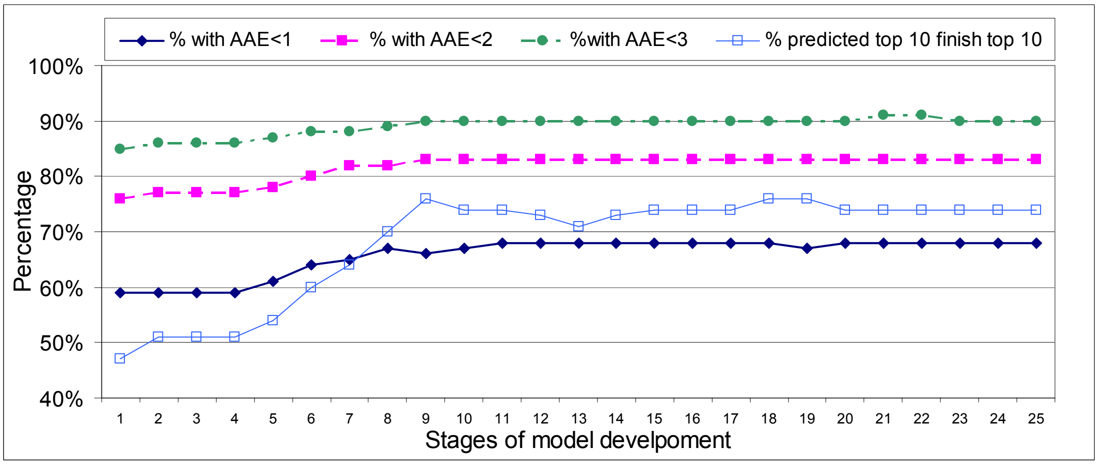

Figure 8-9 Predicted top 10 who finish in top 10 and AAE <1, 2 & 3

8.7.2 Top 10 for season

By measuring the proportion of players that were ranked by the models to finish in the top 10 and actually did finish in the top10, we can further gauge predictive capacity. From Figure 8-9 we can see that after the addition of the first 10 variables, little practical benefit could be achieved with regards to accurately identifying the leading 10 contenders to win the Brownlow medal, with the modelling approach having about 74% accuracy in predicting leading players.

Table 8-3 Seasonal goodness of fit measures for each stage of model development

Stage	AAE	AAE20	Top10	Less1	Less2	Less3
1	1.49	4.83	47%	59%	76%	85%
2	1.42	4.98	51%	59%	77%	86%
3	1.42	4.93	51%	59%	77%	86%
4	1.41	5.02	51%	59%	77%	86%
5	1.35	5.08	54%	61%	78%	87%
6	1.26	4.68	60%	64%	80%	88%
7	1.18	4.47	64%	65%	82%	88%
8	1.10	4.04	70%	67%	82%	89%
9	1.07	3.67	76%	66%	83%	90%
10	1.06	3.58	74%	67%	83%	90%
11	1.04	3.56	74%	68%	83%	90%
12	1.04	3.54	73%	68%	83%	90%
13	1.04	3.55	71%	68%	83%	90%
14	1.04	3.53	73%	68%	83%	90%
15	1.03	3.57	74%	68%	83%	90%
16	1.03	3.62	74%	68%	83%	90%
17	1.03	3.61	74%	68%	83%	90%
18	1.03	3.62	76%	68%	83%	90%
19	1.03	3.56	76%	67%	83%	90%
20	1.02	3.60	74%	68%	83%	90%
21	1.02	3.54	74%	68%	83%	91%
22	1.02	3.55	74%	68%	83%	91%
23	1.02	3.66	74%	68%	83%	90%
24	1.02	3.61	74%	68%	83%	90%
25	1.02	3.69	74%	68%	83%	90%

Key:

AAE = Average absolute error for predicted season total minus actual season total

AAE20 = AAE for the top 20 ranked players for the season

Top 10 = Percentage of predicted top 10 that actually finish in top 10

Less 1 = Percentage of players with an AAE less than 1

Less 2 = Percentage of players with an AAE less than 2

Less 3 = Percentage of players with an AAE less than 3

8.8 Measuring the bias of over-fitting the data

All previously defined criteria for goodness of fit have been determined by using all seven years of data as both a training set to determine parameter estimates and a holdout sample to determine predictability, thus overfitting the data. In order to measure the bias of over-fitting the data, the same 25 parameter model is applied to six additional training datasets, with lengths ranging from one to six years. Goodness of fit has been determined from holdout samples also ranging from one to six years, with the addition of training and holdout samples adding to seven years. Thus, when the training data was 1997, the corresponding prediction model was applied to the remainder of the data (1998-2003). When two years worth of training data was used (1997 & 1998), the holdout sample used was 1999 to 2003.

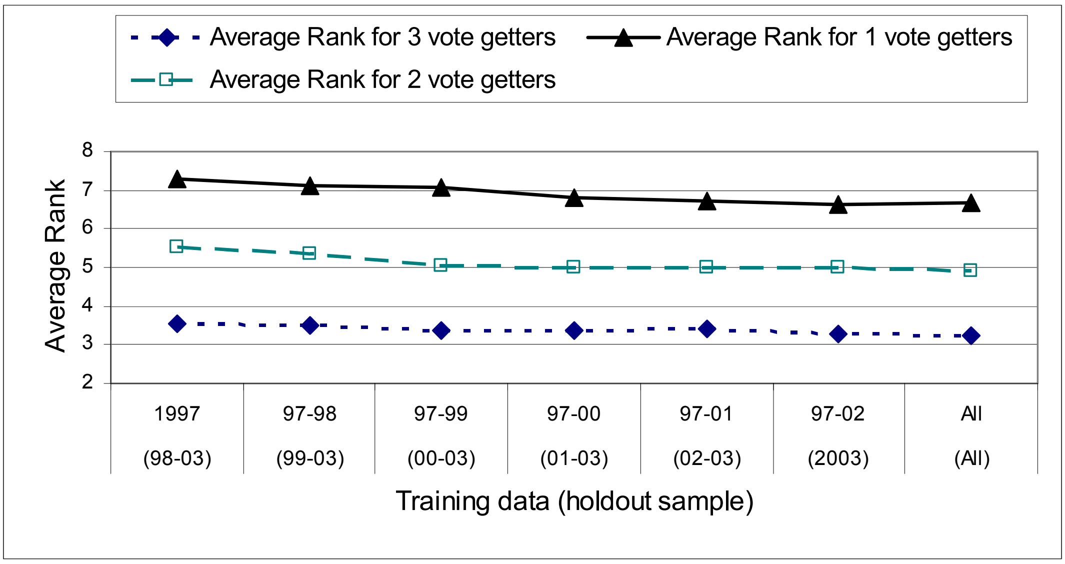

Figure 8-10 Average rank with differing training and holdout sample sizes

By considering the average rank for the players that polled votes, it is possible to gauge the relative accuracy of differing holdout samples. From Figure 8-10 it can be seen that when only the 1997 season was used as a training dataset, the average rank for players polling three votes was 3.5. When three seasons worth of data was used (97-99), the average rank for players polling three votes was reduced to 3.35. When the data was overfitted, by developing on all seven seasons and then reapplying to the same data, the average rank for players polling three votes was reduced to 3.22.

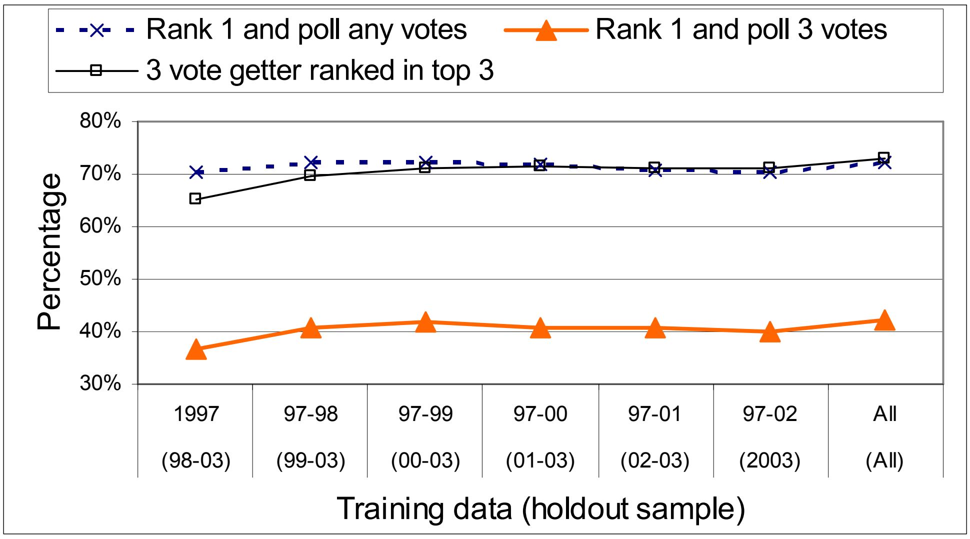

Figure 8-11 Within game measures of goodness of fit

When only the 1997 season was used as a training dataset, the leading ranked player according to the modelling process would poll three votes 36% of the time. By increasing the training data set to two years, the three vote getter could successfully be identified 40% of the time, but it would appear that little benefit could be gained by using more than two years worth of training data, as the percentage of best players successfully identified alters little between two and six years. When the data is overfitted, the bias of over fitting can be approximated at 2% with the three vote getter identified 42% of the time.

When only the 1997 season is used as a holdout sample, the leading ranked player would poll votes 70% of the time. This figure improves slightly to 72% with either two, three of four years worth of holdout data, but drops again slightly when five of six years of training data is used; perhaps reflecting increased variability in the holdout sample. When the data is overfitted, the leading ranked player polls votes 72% of the time.

The final measure considered on Figure 8-11 is how often the player who polls three votes is ranked in the top three positions by the models. When only the 1997 season is used as a training sample, the three vote getter was only ranked in the top three 65% of the time. This figure improves to 70% with two years worth of training data, but shows little improvement after that. When overfitted, the three vote getter is ranked in the top three 72% of the time.

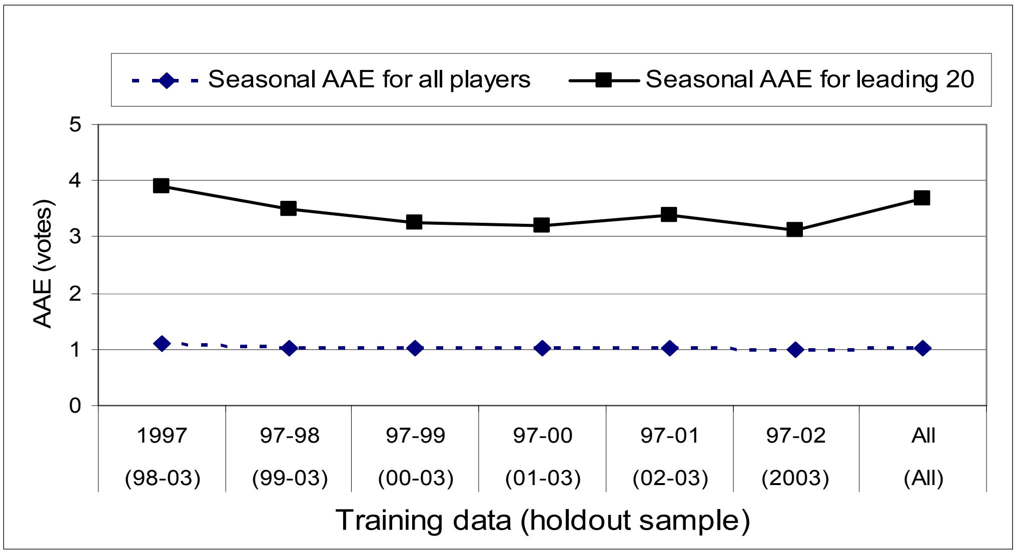

Figure 8-12 Seasonal AAE for all players and leading 20 players

When considering criteria for goodness of fit for the season, the bias associated with overfitting the data is less pronounced. When only the 1997 season is used, the AAE between each player's predicted and actual number of votes polled for the season is 1.1, whilst for the leading 20 ranked players it is 3.9 (see Figure 8-12). When at least two seasons are used for the holdout sample, the AAE for all players is reduced to one, whilst for the leading 20 it is reduced to 3.5. Although there is no difference in AAE for all players when the data is overfitted, the AAE for the leading 20 players is higher when the data is overfitted.

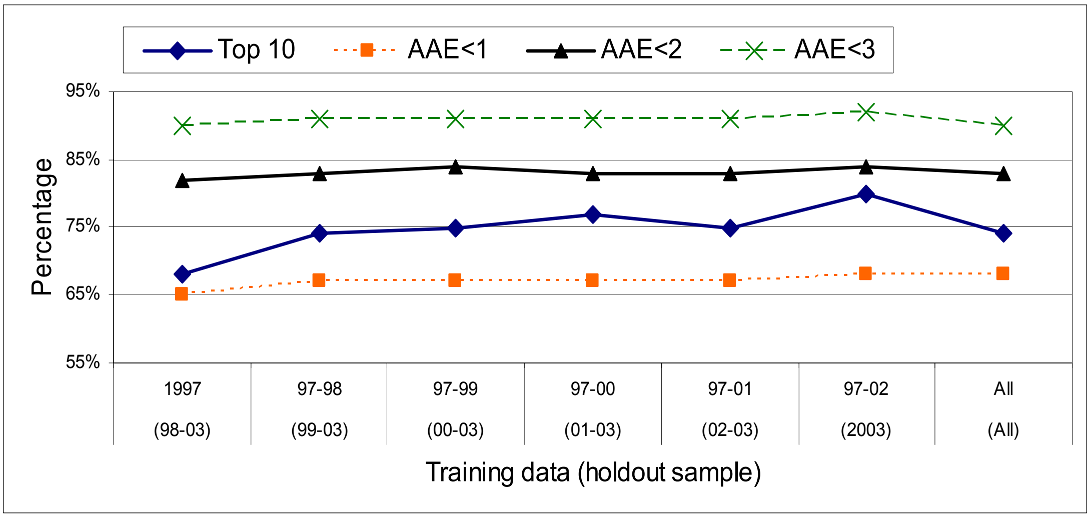

Figure 8-13 Season measures of goodness of fit for differing holdout samples

When only the 1997 season is used as a training dataset, the total votes polled by each player can be identified to within one vote for 65% of all players. It can be seen from Figure 8-13 that when the size of the training dataset is doubled, this figure improves to 67%. When six years of training data is used, this figure improves to 68%, but no additional benefit is apparent by overfitting the data as the overfitted data can still only identify 68% of all players to within one vote of their seasonal tally.

Similarly, the percentage of players that are accurately identified to within two or three votes, appears to improve when the training dataset is increased from one to two years, but little further improvement can then be achieved, and when overfitted, results are slightly worse.

The percentage of players that are accurately identified as finishing the season in the top 10, shows slight improvement as the size of the training data increases, although as only 10 players are considered each year, there is higher variability is this measure of goodness of fit.

8.9 Results

An ordinal logistic regression model was first incorporated to predict Brownlow medal results in the year 2000. The multivariate model has undergone annual improvement from a 15 variable model in the year 2000 to the 25 variable model in 2003. Despite all variables in the model being highly significant ($p<0.001$), the practical benefit of some variables appears negligible.

This research suggests that goodness of fit of prediction models is dependent upon the quality and quantity of data used to construct the model. A minimum of about three seasons (528 games) is required from which to develop models, and although the quality of models does continue to improve with larger holdout samples, the improvement appears minimal. The similarities between the Brownlow medal and a horse race, reaffirms the finding of Benter (1994) in developing multivariate models for horse racing in Hong Kong. Benter states "… the minimum amount of data needed for adequate model development and testing samples is in the range of 500 to 1000 races. More is helpful, but out-of-sample predictive accuracy does not seem to improve with development samples greater than 1000 races".

Using three seasons of holdout data (1997-99) to derive parameter estimates, the 25-parameter ordinal logistic regression model was applied to seasons 2000 to 2003, the results of which can be viewed in Table 8-4.

Table 8-4 Actual and predicted Top 10 Brownlow results 2000-2003

Year	Finishing Position	Actual Vote Total	Team	Player	Predicted Vote Total	Predicted Order
2000	1	24	ME	Woewodin	12.07	12
2000	2	22	WB	West	14.52	6
2000	3	20	AD	McLeod	12.56	10
2000	4	19	CA	Koutoufidis	18.07	2
2000	5	18	CO	Buckley	18.79	1
2000	5	18	CA	Ratten	13.86	7
2000	7	16	BR	Voss	15.12	4
2000	7	16	ES	Hird	14.61	5
2000	7	16	SY	O'Loughlin	9.24	22
2000	10	14	NO	Carey	16.73	3
2001	1	23	BR	Akermains	17.18	2
2001	2	21	AD	McLeod	17.08	3
2001	3	19	BR	Voss	18.15	1
2001	3	19	PA	Francou	12.80	10
2001	5	18	WE	Cousins	15.11	5
2001	6	16	NO	Harvey	14.83	6
2001	7	15	ES	Johnson	14.13	8
2001	7	15	ES	Lloyd	12.18	15
2001	9	14	CO	Buckley	15.89	4
2001	9	14	HA	Crawford	14.15	7
2002	1	25	BR	Black	12.56	10
2002	2	21	PA	Francou	15.15	4
2002	3	17	HA	Crawford	17.90	1
2002	3	17	BR	Voss	16.41	2
2002	3	17	ME	Yze	10.89	14
2002	6	16	WE	Cousins	15.59	3
2002	6	16	AD	McLeod	14.72	5
2002	6	16	WB	Darcy	14.37	6
2002	6	16	SY	Williams	9.17	23
2002	6	16	BR	Headland	8.64	27
2003	1	22	CO	Buckley	21.18	1
2003	1	22	AD	Ricciuto	16.33	5
2003	1	22	SY	Goodes	11.51	16
2003	4	21	HA	Crawford	19.14	3
2003	4	21	PA	Wanganeen	14.70	7
2003	4	21	WE	Cousins	14.30	8
2003	7	19	BR	Voss	20.76	2
2003	7	19	FR	Bell	16.40	4
2003	7	19	ES	Hird	11.22	19
2003	10	18	ST	Harvey	14.83	6

In the year 2000, the Brownlow medal was won by Shane Woewodin. The summation of Woewodin's predicted votes for each game of the 2000 season totalled 12 votes, whilst the number of votes that he actually polled was 24; 12 votes more than expected. In 2001 the actual winner polled six more votes than expected, in 2002 the winner again polled 12 more votes than expected, whilst in 2003, the three equal winners polled one, six and 10 more votes than expected. The fact that the modelling process consistently predicts a lower total than that of the winner is a reflection of two features. Firstly, there is a high degree of variation in voting that a mathematical model cannot explain - at best the modelling process can accurately pick the leading player on the ground only 40% of the time. Secondly, Brownlow voting represents count data and as such could be well characterised by a Poisson distribution. Although the nature of the three, two, one voting pattern may lead to over or under dispersion, the simple principle that variance increases with an increasing mean is apparent. Conversely, it may well be the players with the highest variability have the greatest probability of winning; hence it is of little surprise that the actual winner will have a high AAE.

8.10 Discussion

Although the modelling process can effectively identify the leading players each season, the leading ranked player has only won the Brownlow medal on one occasion. An analogy can be drawn between predicting the winner of the Brownlow and predicting the winner of the Melbourne cup. Although the modelling process can objectively assign a probability of success, the horse with the highest probability does not always win. Thus with respect to the Brownlow, we might deduce that the favourite has only won on one of four occasions.

Whilst indications might suggest that the ordinal regression models may not be ideal for selecting the actual winner, for the vast majority of players, the modelling process is very successful. By accurately predicting 66% of players to within one vote of their actual total, and 90% of players to within three votes of their total, the modelling process provides an objective assignment of probabilities that has many benefits.

As with previously discussed sporting outcomes, bookmaker prices represent a good benchmark in which to compare models. Not only do bookmakers hold bets on the actual winner of the Brownlow, but punters can also bet on a range of betting outcomes such as-

- Top 3
- Top 5
- Individual head to head betting
- Specific groups
- Vote totals
- Leading votes per club
- Team totals

Unfortunately, due to the infrequency in which the Brownlow medal count is held, there is insufficient data to enable valid statistical comparisons between bookmakers and the model. Despite any prolonged evidence of market inefficiency, preliminary indication would suggest that the modelling process is more efficient than bookmakers in assigning probabilities, particularly for individual head to head betting and team totals.

As the voting patterns of leading players tend to follow a Normal distribution, any two given players can be readily matched by comparing the ratio of the difference in votes divided by the standard error, with the standard Normal distribution, creating for each player a probability of outvoting the other (see section 6.8). When considering team totals, the central limit theorem aids in the accuracy of predictions. While relatively large differences can occur between individuals predicted and actual totals, when aggregated over the team, the level of accuracy was found to improve substantially.

While such modelling may have benefit for seasonal totals, clearly the best application of this process is to define probabilities of polling votes for each player in each individual game. Although betting of this nature is currently not available, multivariate modelling paves the way for this to occur in the future.

The Brownlow medal count is held annually on the Monday night prior to the Grand Final. Votes for each game are counted chronologically for the season and presented to a national television audience. As votes were announced for each match, they were entered into the database. Adding each player's predicted votes for future rounds with the actual number of votes that he had already polled, enabled a prediction process that was constantly improving from round to round.

Starting in the year 2000 and continued annually, updated Brownlow predictions were published to the Internet at the end of each round by the Swinburne Sports Statistics department, providing viewers with a more accurate guide to the final vote totals of leading players. These predictions proved so popular, that in 2002 and 2003 the service had to be cut as the university server could no longer handle the volume of traffic coming through the site.

In addition to providing an informative internet service, the statistically derived prediction approach has received much media attention including newspaper, print articles and a national television appearance on "The Footy Show". The success of this approach is further reflected by the ongoing sale of Brownlow prices to an established internet bookmaker, both during and at the completion of each season.

The ability of a mathematical model to provide objective predictions at the push of a button provides clear advantages when time is constrained. Another application of multivariate modelling requiring speed of calculation is the prediction for the number of runs that will be scored in each over during ODI cricket matches.

9 Predicting the number of runs scored per over in ODI cricket matches

9.1 Introduction

ODI cricket is a popular sport worldwide. The advent of the Internet has increased opportunities for punters to wager on differing outcomes associated with each game. Swinburne Sports Statistics department was approached several years ago with the aim to produce a mathematical model that could aid in the prediction of the number of runs scored per over (RPO). It was hoped that such a model could be utilised to offer gamblers the opportunity to bet on the outcome of each individual over. Using information gathered from past ODI matches, specific predictors for RPO were identified and combined into multivariate prediction models. Differing underlying distributions were compared to identify an approach that would provide the best prediction models possible. This process has now been implemented by one of the world's leading bookmakers to provide betting per over during the course of matches.

9.2 Background

Prior to February 2004, 2100 official ODI matches had been played between 20 competing countries. Although match results and player information is available for all matches played, information at an individual over level has only become available in recent years, and has been collected for 627 matches. Although in theory, 100 overs could be bowled in each ODI, due to the nature of the game, this has only occurred about 12% of the time. There are various reasons why a game would not go for the full 100 overs. Rain delays, one or both of the sides being dismissed before using their full resource of 50 overs, the second side completing the required target within the 50 overs, or penalties imposed upon teams for slow play are all reasons why 100 overs would not be bowled. On average, 88 overs were bowled per match, creating a database of 55,000 overs. Table 9-1 shows a truncated version of information that is available at an individual over level from the Internet.[8]

Table 9-1 Example of data collected from the second innings of an ODI match

Over	Score	RO	RR	RR5o	RRreq	Rreq	Brem	Wr
1	1/0	1	1.00		6.20	304	294	10
2	4/0	3	2.00		6.27	301	288	10
3	5/0	1	1.66		6.38	300	282	10
4	11/0	6	2.75		6.39	294	276	10
5	18/0	7	3.60	3.6	6.37	287	270	10
6	31/1	13	5.16	6.0	6.22	274	264	9
7	35/1	4	5.00	6.2	6.27	270	258	9
8	42/1	7	5.25	7.4	6.26	263	252	9
9	46/1	4	5.11	7.0	6.31	259	246	9
10	56/1	10	5.60	7.6	6.22	249	240	9

Key:
RO = Runs off the over
RR = Run-rate for the innings
RR5o = Run-rate from the last 5 overs
RRreq = Run-rate required to win the match
Rreq = Runs required to win
Brem = Balls remaining
Wr = Wickets remaining

Although the number of runs scored per over clearly does not follow a Normal distribution (see Figure 9-1), the enormity of the database allows for some practical benefit to be gained by using a parametric approach to determine prediction variables.

[8] http://aus.cricinfo.com/

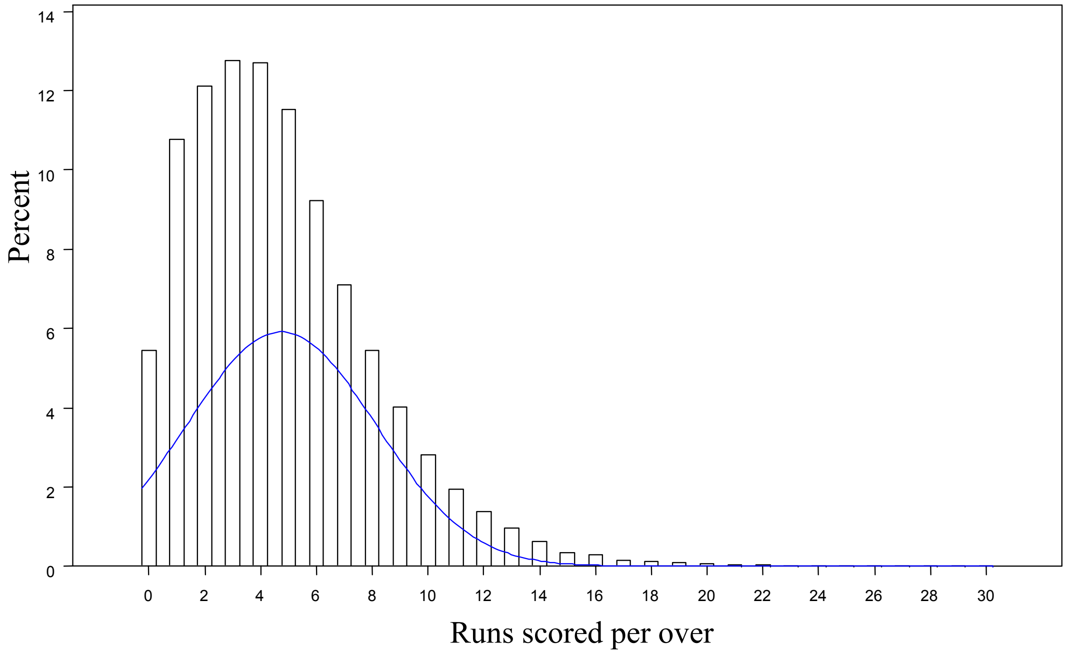

Figure 9-1 Histogram of runs scored per over in ODI cricket

Using information gather from 627 ODI matches (55,000 overs) played between 1998 and 2004, multiple linear regression was used to identify 14 variables that were independently highly significant ($p<0.0001$) predictors of RPO. To determine what underlying distribution might best approximate RPO, five distributions namely Normal, Log-normal, Gamma, Poisson and Negative-binomial were modelled to the data using the 14 previously identified prediction variables. Goodness of fit comparison was made using the AAE between the predicted mean and the actual number of runs scored and the log of the assigned probability for the actual number of runs scored (see section 3.7).

By collapsing all scores greater than nine into the same category, two additional modelling approaches were incorporated using logistic regression. Firstly, an ordinal logistic regression was applied to the 11 possible outcomes (0, 1, 2…9, >9). This model incorporates a different intercept term for each additional run, but has fixed parameter estimates. To further enable variation in the parameter estimates for each run category, a series of 10 binomial models were constructed, each with a different cut-off for success.

The first model predicted whether the number of runs scored would be greater than or equal to one, whilst the second model predicted whether the number of runs would be greater than or equal to two, and so on. By subtracting sequential models the probability for each run category could be determined.

9.3 Predictive factors for runs scored per over

9.3.1 Overview

Prediction variables can be readily divided into 'within match' and 'between match' variables. Not surprisingly, significantly more variation occurs within games than between games. Within game factors that impact on RPO include batting sequence, over number, wickets fallen, duration of current partnership, scoring rate for the previous five overs, wicket fallen in the last over, run rate from the same end and best bowler. Between match variables include the location of the ground and the quality of both the competing teams.

9.3.2 Batting Sequence

More runs are scored per over by the team batting first than by the team batting second (1^{st} inning 4.80±0.02 vs. 2^{nd} inning 4.66±0.02, $p<0.0001$). This may not necessarily mean that runs are easier to score in the first innings. The constraints imposed upon batsmen in the second inning are different to that of the first. In the first inning a batsmen has the aim to score as many runs as possible, whereas in the second innings, batsmen face a specific target to win. There exists a trade off in ODI cricket between risk and reward, whereby to score more runs generally requires the batsmen to increase the risk of losing his wicket. In the second innings, if a target is relatively small, a batsman may opt to score at a slower rate thus reducing risk. The impact of having a target to chase (2^{nd} innings) in comparison to maximising the number of runs scored (1^{st} innings) has a dramatic impact on the predicability of RPO. Over twice as much variation can be explained in the first innings compared to the second, suggesting that external factors increase the variability associated with the second inning. Using linear regression, significant interactions could be found between the batting sequence and several other prediction variables, suggesting the need to model the first and second innings separately.

9.3.3 Overs

A linear predictor applied to RPO would suggest that in the first innings of a game, the expected number of RPO would increase at a rate 0.057±0.001 for each additional over. Interestingly, a linear predictor applied to the second innings

would suggest that the expected number of RPO would increase at a rate of only 0.024±0.001 runs for each additional over. In reality, the effect of the new ball and fielding constraints ensure that a linear relationship between runs and overs is overly simplistic.

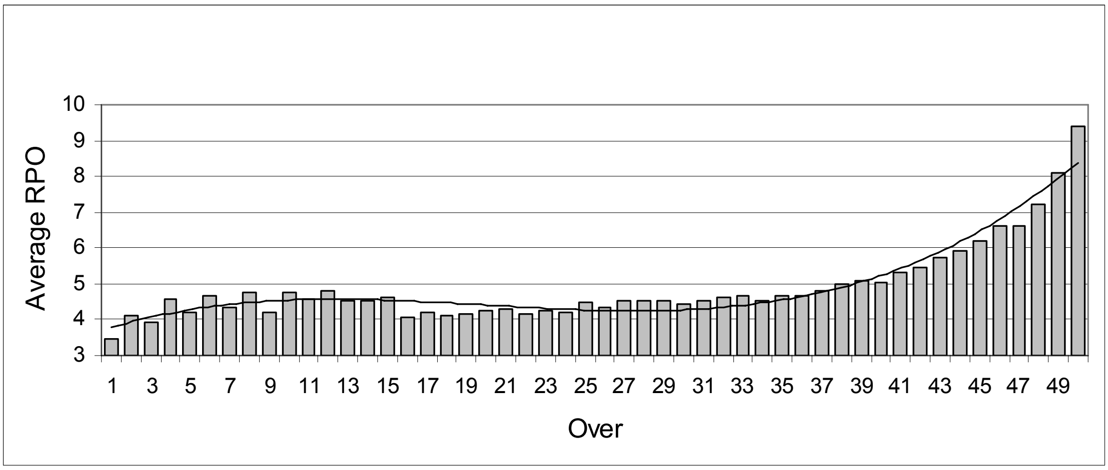

Figure 9-2 Average runs scored per over (both innings combined)

From Figure 9-2 it could be hypothesised that the relationship between overs bowled and runs scored during the course of a match follows a polynomial distribution with three degrees. Although perhaps a little simplistic there is merit in assuming that the course of ODI innings goes through three distinct stages. With fielding restrictions in place for the first 15 overs, a clear distinction can be seen between the fifteenth and sixteenth overs. Further scrutiny reveals that the scoring rate increases more dramatically as the match approaches its conclusion, suggesting a need to identify the turning point in which teams begin to accelerate the scoring rate towards the end of the match. To do this, a series of binomial variables were created, categorising the data as either above or below a given cut-off of overs. By then maximising the likelihood, the cut-off that produces the best fit to the data was identified. By considering a generalised linear model of the following form

$$Runs = A + B(First15) + C(Overs) + D(first15*Overs) + E(Cutoff) + F(Cutoff*Overs) \qquad (9.1)$$

each match could be divided into three sections, with a linear model fitted to each section. A cut-off at the 41 over mark produced the best fit to both the first and second innings. Thus each innings could now be viewed as having three distinct phases, start (overs 1-15), middle (overs 16-41) and end (overs 42-50).

9.3.4 Wickets

Cricket is played between two teams of 11 players – of which two are required to be at the batting crease at any given time. As a general rule, better players bat higher in the batting order so as to optimise the time available from which to score runs, although players batting down the order can often score at a faster rate. Not only is each batting team constrained by the maximum number of overs that they can received, they are also limited by the number of batsmen that they have available. Theoretically, one could expect that as available resources decreased (wickets), so too would the scoring rate.

Using a generalised linear model, with wickets fallen as a continuous variable, the fall of each wicket was found to have significantly more impact towards the end of a game with each additional wicket reducing the average run rate by about a half a run per over. There was no significant difference between the rate of decline between the start and middle stages of the game. This is reflected in Figure 9-3, with the relationship between wickets fallen and RPO differing significantly between the three stages of the first inning. (start:-0.13±0.03 runs per wicket fallen, middle:-0.19±0.03, end:-0.59±0.03).

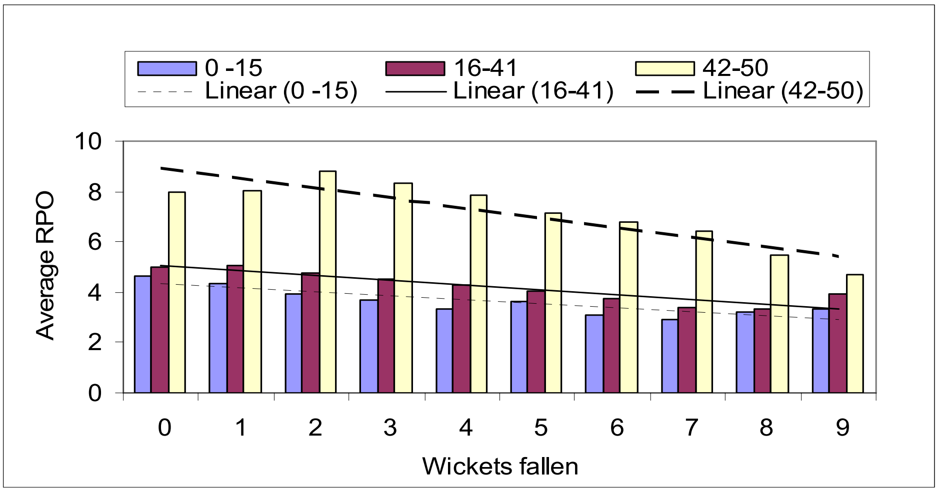

Figure 9-3 RPO for wickets fallen for the three stages of the first inning

From Figure 9-4 it can be seen that in the second innings, the fall of a wicket will reduce the runrate by the greatest amount during the start of the inning (start: -0.18±0.04, middle: -0.06±0.05, end: -0.12±0.04).

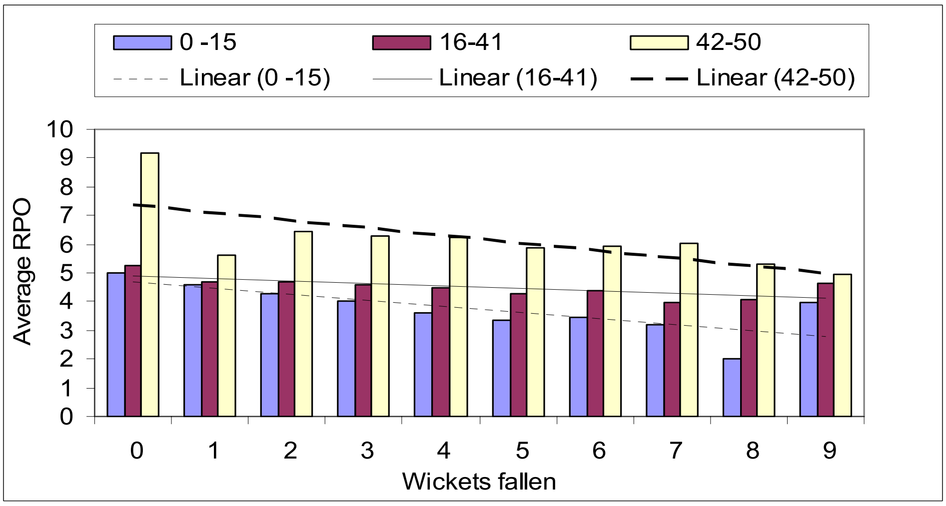

Figure 9-4 RPO for wickets fallen for the three stages of the second inning

9.3.5 Wicket last over

Factors that occur within the game can have enormous effect on RPO. A good example is if a wicket has fallen in the previous over. As seen in Bailey and Clarke (2004), a study of all ODI batting performances revealed that batsmen are most vulnerable when they first come to the crease. This reflects a brief "training" period where batsmen adjust to the conditions and the way the opposition are bowling. If a wicket has fallen in the previous over, a team will score on average one run less in the following over in the first inning and 1.3 runs less in the second inning (1^{st} innings 0.98±0.05 vs. 2^{nd} innings 1.32±0.06, $p<0.0001$).

9.3.6 Partnerships

Another strong predictor for RPO is the magnitude of the partnership for the batsmen at the crease. This equates to the number of runs scored since the fall of the last wicket. It is generally accepted that the longer a batsman spends at the crease, the more comfortable he becomes, thus making it easier to score runs.

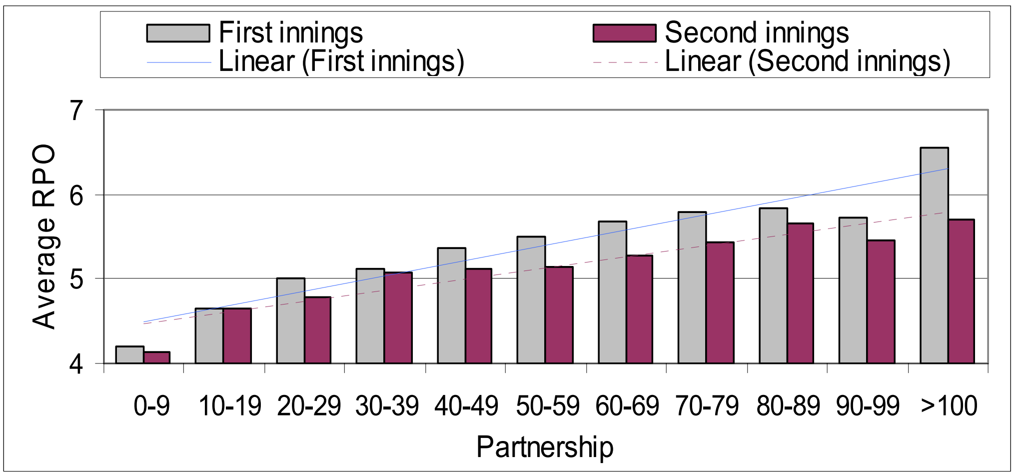

Figure 9-5 Relationship between partnership and runs per over

From Figure 9-5 it is possible to see a clear linear trend between partnership and RPO. Although partnership appears to have a slightly stronger relationship with RPO in the first innings than in the second, this difference was not statistically significant. From Figure 9-5 the greatest difference between innings occurs for partnerships in excess of 100 runs. This is not a surprising result, as in the second innings batsmen are chasing a specific target and are thus not required to maximise the scoring rate to its optimal potential. From the multivariate models, the parameter estimate for partnership in the first innings is equivalent to 0.006±0.001 RPO whilst for the second innings the parameter estimate is equivalent to 0.005±0.001 RPO.

9.3.7 Runs previously scored in the match

Significant auto-correlation exists between consecutive overs in ODI matches. Because bowlers will often bowl consecutive overs from the same end, it is of no surprise that runs scored in the previous over from the same end provide an even better predictor of RPO. By measuring the AAE between actual runs scored and predicted runs scored, it is possible to compare the explanatory capacity of various within game predictors. From Figure 9-6 it can seen that the more information that can be used from within the match, the lower the AAE, thus the better the prediction.

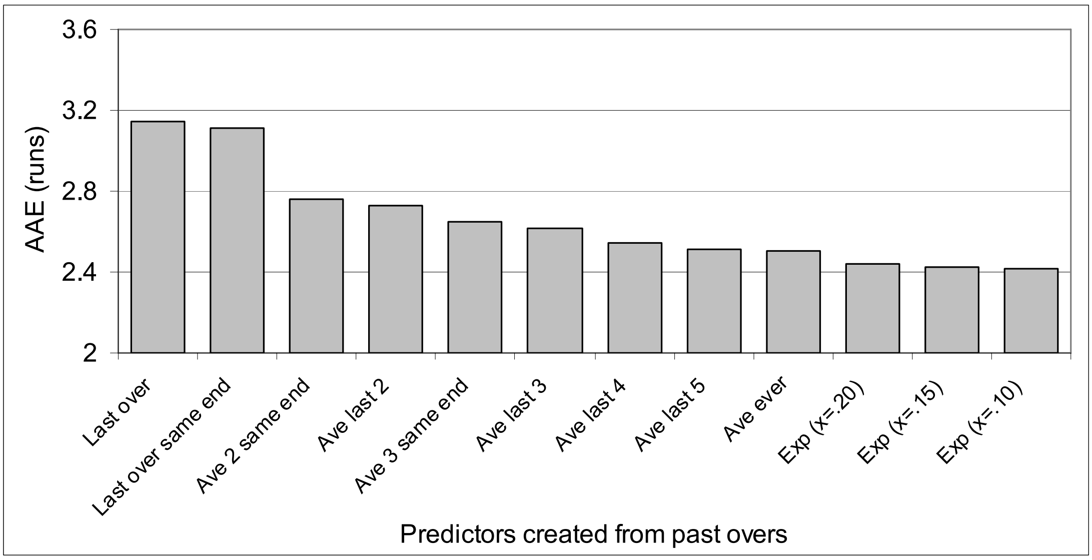

Figure 9-6 AAE for predictors created from previous overs bowled

Arithmetic averages weight each over equally, irrespective of when it occurred. An alternative approach is to give more weight to more recent overs by exponentially smoothing past results. It is possible to derive an unbiased prediction for future data by using the formula

$$\textit{Smoothed score} = (x)\ \textit{actual runs} + (1-x)\ \textit{Previous smoothed score} \qquad (9.2)$$

where x is the smoothing parameter. Both rolling averages and exponentially smoothed predictors were compared. Whilst Figure 9-6 would suggest that exponential smoothing produced the lowest AAE, the run rate for the previous five overs and the average runs scored from the last three overs from the same end were also highly significant predictors in the multivariate model.

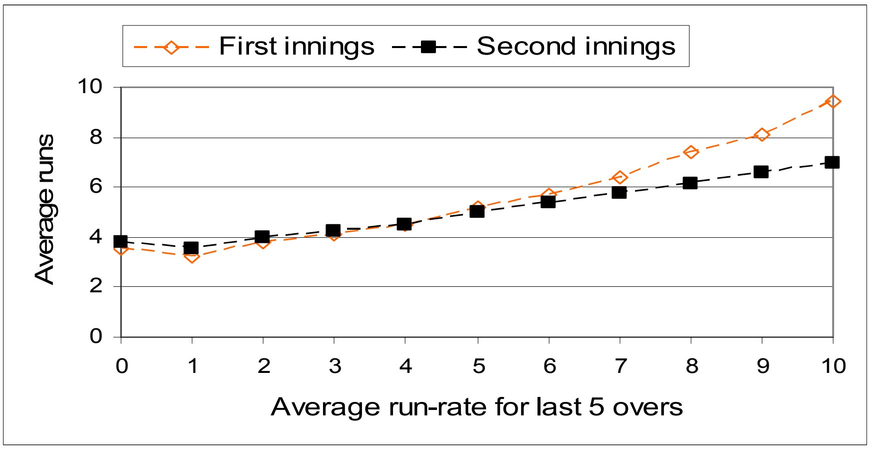

Figure 9-7 RPO and run-rate for the previous five overs

For every one run increase in the run rate for the previous five overs, RPO was found to increase in both the first and second innings at a rate of 0.15±0.03. While Figure 9-7 might suggest the rate to be slightly different between first and second innings, this was not statistically significant. Similarly, for every one run increase in run rate from last three overs from the same end, RPO was found to increase at a rate of 0.11±0.02. There was no significant interaction with innings for either of these two predictors (See Figure 9-8).

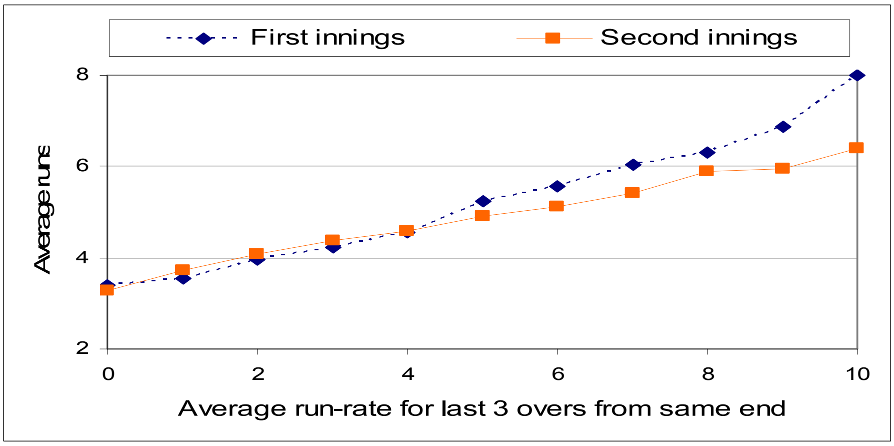

Figure 9-8 RPO and run-rate for the last three overs from the same end

9.3.8 Best Bowler

A closer examination of the first 15 overs reveals some interesting trends. At the commencement of each game, the bowling side is given the use of a new cricket ball. This new ball combined with a joint training effect experienced by both batsmen as they "get their eye in" ensures that significantly less runs are scored on average in the first over than all others (first over 3.44±0.08 vs. all other 4.76±0.01, $p<0.0001$). When considering first inning performance versus second inning performance, the first over effect is greater in the first inning (first over 3.20±0.11 vs. all other 4.83±0.02) than in the second, (first over 3.69±0.11 vs. all others 4.68±0.02), with an interaction term between innings and first over statistically significant ($p<0.0001$).

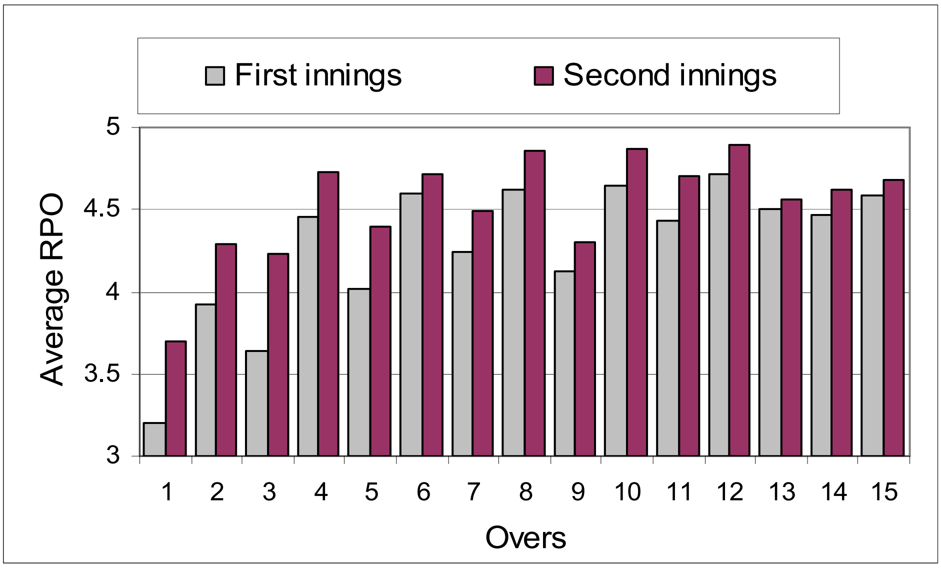

Figure 9-9 Average runs scored in the first fifteen overs

It is normal for the best fast bowler in the side to have first use of the new ball. Each bowler can bowl up to 10 overs per game, but it is unusual for an opening bowler to bowl all 10 overs consecutively. Although dependent upon the performance of the bowler, it would be usual for the best fast bowler in the side to have a spell of bowling that would last about 5-7 overs. This means that it is not unrealistic to expect that the best fast bowler in each side would bowl the first five odd numbered overs (1,3,5,7 & 9) and the second best fast bowler would bowl the first five even numbered overs (2,4,6,8 &10). The fact that bowlers alternate ends can clearly be seen from Figure 9-9 with a significant difference existing between the best and second best fast bowlers from each country. This result can be further confirmed by averaging runs scored for the first five odd numbered overs in comparison to the first five even numbered overs (odd 4.04±0.04 RPO vs. even 4.57±0.04 RPO $p<0.0001$). To account for the best bowler from each country, a binomial variable was created to identify the first five odd numbered overs from each innings. There was no significant interaction between the best bowler and innings.

9.3.9 Host Country

Twenty different host countries are represented in the database, although only 12 host countries have more than 1000 overs of available information. From these 12 countries, runs are primarily scored at the greatest rate in sub-continental countries, where pitches are perceived to be more batsmen friendly. This is reflected in Figure 9-10 with India and Pakistan having the highest average RPO.

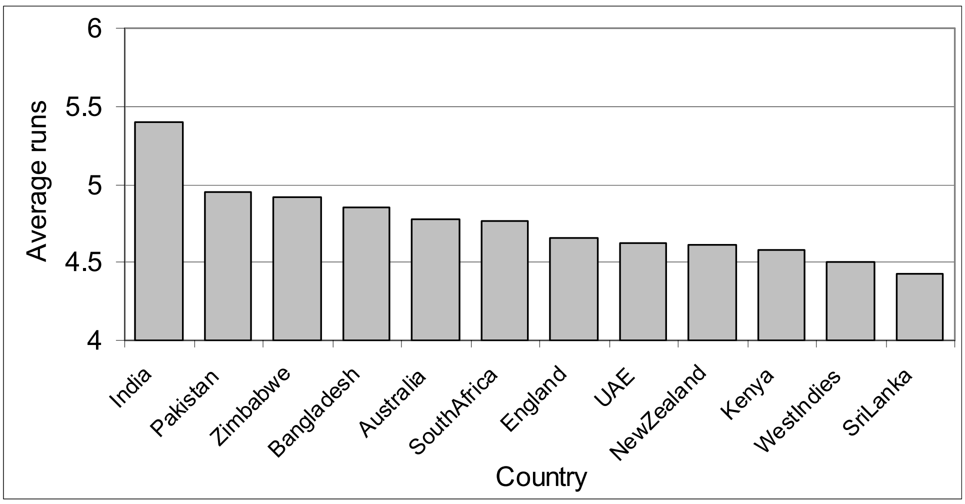

Figure 9-10 Run rate for host country

9.3.10 Team

Significant differences can be seen in the relative strength of the batting teams of competing nations. Interestingly, there is a strong correlation between the average number of runs scored per team and the amount of overs each team has faced over the past six years. From Figure 9-11 a clear distinction can be drawn between the more established cricketing nations and those who play international cricket on a less frequent basis. This may also reflect the fact that the weaker cricketing countries are less likely to bat and bowl for their full 50 overs. Australian batsmen have scored at the fastest rate in the past six years, followed by India and South African batsmen. Second tier teams such Scotland, Canada and the Netherlands, score at the lowest run-rate and have faced the fewest number of overs at an international level in the last six years.

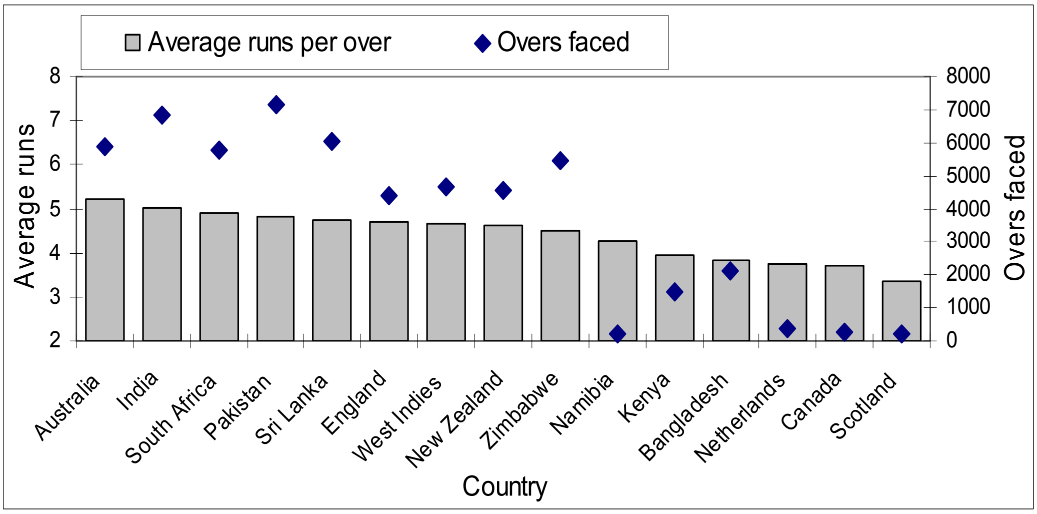

Figure 9-11 RPO and overs faced for the batting team

9.3.11 Team by stage

In addition to differences that exist between scoring rates for different countries, there are also significant differences that exist between countries for each of the three stages in the match (see Figure 9-12). This may well reflect a difference in coaching strategies with some countries such as Australia, opting to follow the advice of Clarke (1988) and score faster earlier in the match, whilst countries such as Pakistan may have a preference to be more conservative earlier in the game, but score faster in the later stages of the match. An alternative explanation could be that Australia has faster scoring batsmen at the start of their batting order, whilst Pakistan has faster scoring batsmen lower down their batting order.

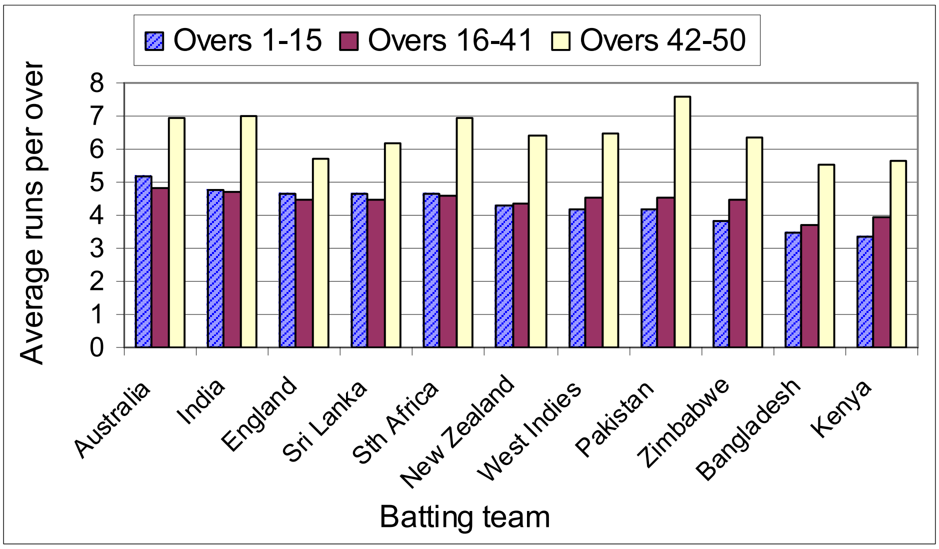

Figure 9-12 Average RPO for each country's batting performances stratified by stage

9.3.12 Opposition by Stage

An important contributing factor to the number of runs scored per over is the quality of the opposition bowlers and fielders. From the 627 matches played, there were 15 countries represented, with 11 of these countries having more than 1000 overs of data available. Figure 9-13 reflects a significant interaction that was found to exist between opposition and stage of the game. Overall, South African and Australian bowler have been the hardest to score runs from, although different countries appear to perform better at differing stages of the match. Once again, this may well reflect the coaching strategies adopted by each country, or may simply be a legacy of team composition.

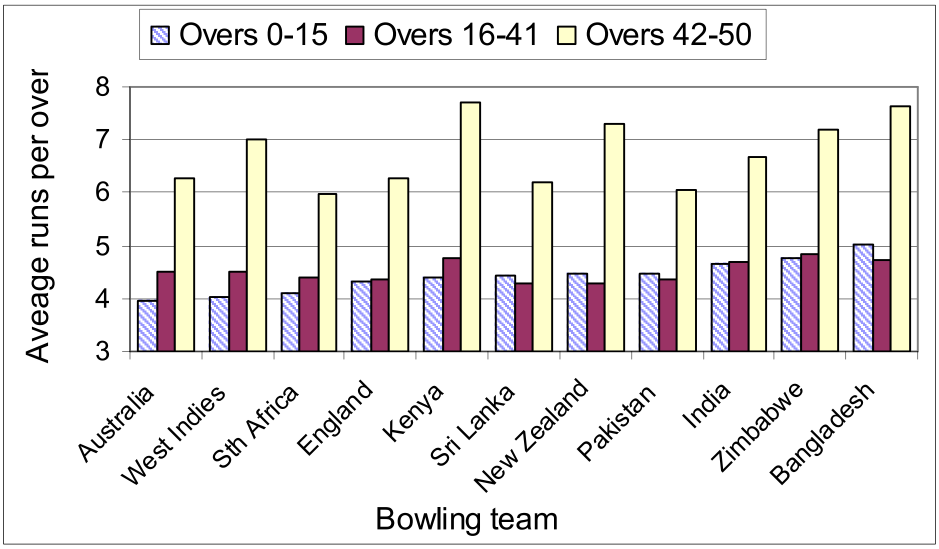

Figure 9-13 RPO for opposition stratified by stage

9.4 Multivariate linear models

A parametric approach to exploring variation associated with RPO enables the use of generalised linear modelling to determine the statistical significance of potential predictive variables. An additional benefit of the linear approach is that it allows practical comparisons to be made through the use of the R-square statistic. 18.6% of the variation in first innings score can be explained by a multivariate model, whereas only 7.8% of the variation in the second innings could be explained. This confirms the need to model innings separately.

Ten first-order variables and four second-order variables were found to be significantly related to RPO in both innings with a p-value less than 0.0001. The corresponding reduction in AAE and -2 log likelihood for each stage of development can be seen in Table 9-2.

Table 9-2 Goodness of fit for stages of model development for both innings

Stage	Variable	First innings			Second innings		
		AAE	-2LL	R^2	AAE	-2LL	R^2
1	Over Number	2.57±.01	157200	5.7%	2.58±.01	132233	0.9%
2	+ Stage	2.53±.01	156187	8.9%	2.57±.01	132012	1.8%
3	+Over X Stage	2.51±.01	155712	10.3%	2.57±.01	131997	1.9%
4	+Wicket	2.46±.01	154204	14.1%	2.54±.01	128976	3.3%
5	+Wicket X Stage	2.44±.01	153911	14.9%	2.54±.01	128950	3.4%
6	+Partnership	2.43±.01	153652	15.7%	2.53±.01	128732	4.3%
7	+Run rate last 5	2.42±.01	153235	16.7%	2.51±.01	128310	5.7%
8	+Wicket last over	2.41±.01	153144	17.0%	2.51±.01	128272	5.9%
9	+Ave. last 3 same end	2.41±.01	152997	17.1%	2.50±.01	128076	6.2%
10	+Best Bowler	2.41±.01	152984	17.2%	2.50±.01	128071	6.2%
11	+Team	2.41±.01	152900	17.5%	2.50±.01	128040	6.4%
12	+Home Country	2.40±.01	152854	17.7%	2.49±.01	127999	6.7%
13	+Opponent X Stage	2.40±.01	152744	18.1%	2.49±.01	127935	7.0%
14	+Team X Stage	2.39±.01	152693	18.4%	2.48±.01	127851	7.5%

Key:
AAE = Average absolute error
-2LL = -2 * log likelihood
R^2 = Percentage of variation explained by model

Both the AAE and the log likelihood have been used to measure goodness of fit. From Figure 9-14 and Figure 9-15, it can be seen that each additional variable brings an improvement to the model. The chronological point in the game has primary importance when predicting RPO, as reflected by the first three parameters, over number, stage and an interaction between the two. Partnerships and wickets and the stage of the game in which they occur, were the next variables of importance. Run rate variables, best bowler, and the recent loss of a wicket complete the list of within match variables that were independently predictive of RPO.

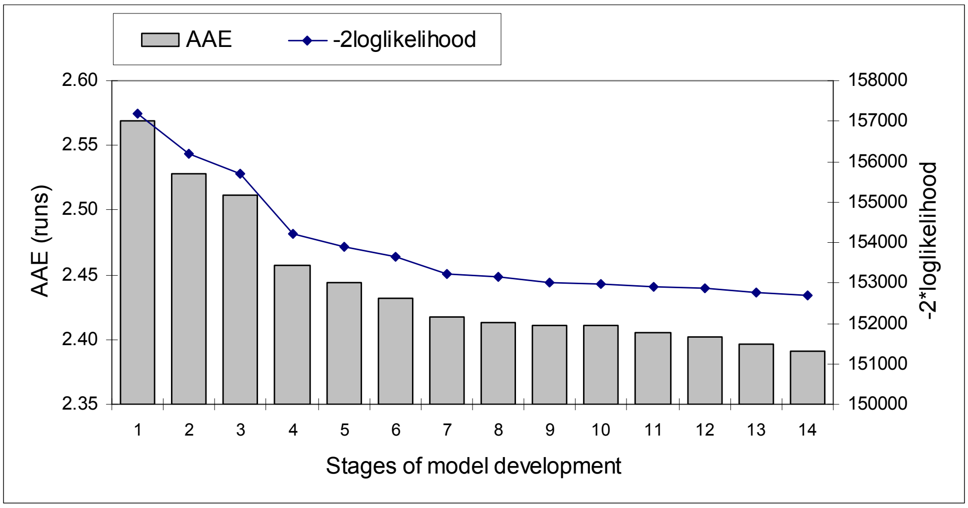

Figure 9-14 AAE and log likelihood for stages of development (1st innings)

Although the host country and the quality of both teams can be shown to be independently significant, from Figure 9-14 and Figure 9-15 it can be seen that their contributions towards improving the model are quite small.

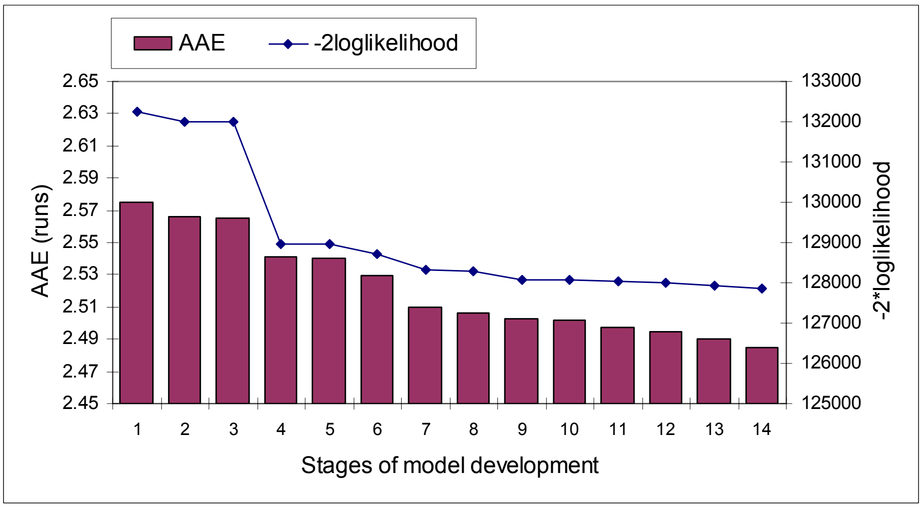

Figure 9-15 AAE and log likelihood for stages of development (2nd innings)

Although parameter estimates differ significantly, the key components for the first innings model, namely, overs, wickets, partnerships and run rates are also the most important predictors for the second inning models. By comparing the R-square statistic for 14 stages of model development between the two innings (Figure 9-16) it can be seen that although the models develop in a similar fashion, far more unexplained variation exists in the second innings.

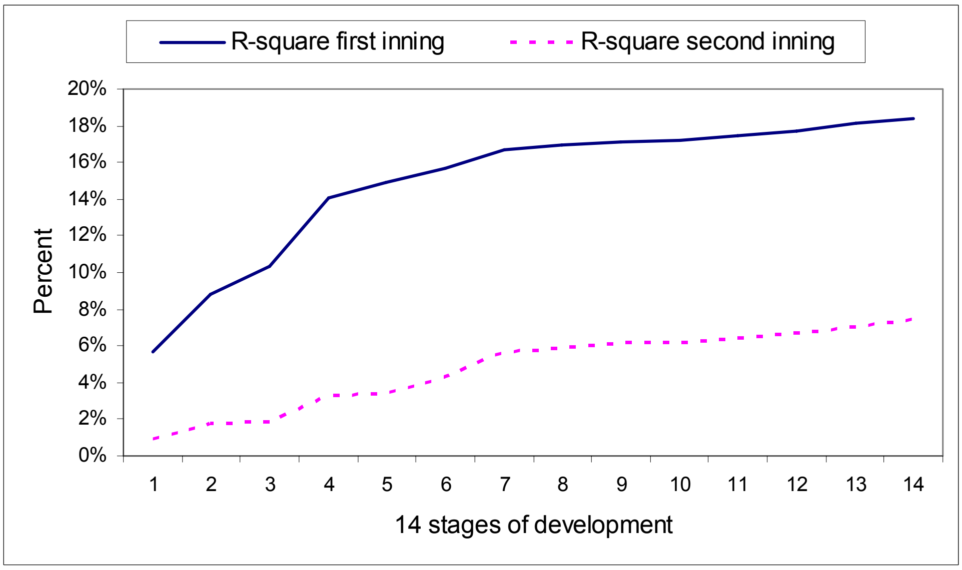

Figure 9-16 R-square statistics showing model development for both innings

9.5 Alternative distributions

Using generalised linear modelling five multivariate models were constructed each assuming a differing underlying distribution, (Normal, Poisson, Gamma, Negative binomial and Log-normal). The 14 variables that were found to be highly significant ($p<0.0001$) using the Normal distribution model were applied to the other four distributions, and found to be equally significant ($p<0.0001$) regardless of what underlying distribution was fitted to the data.

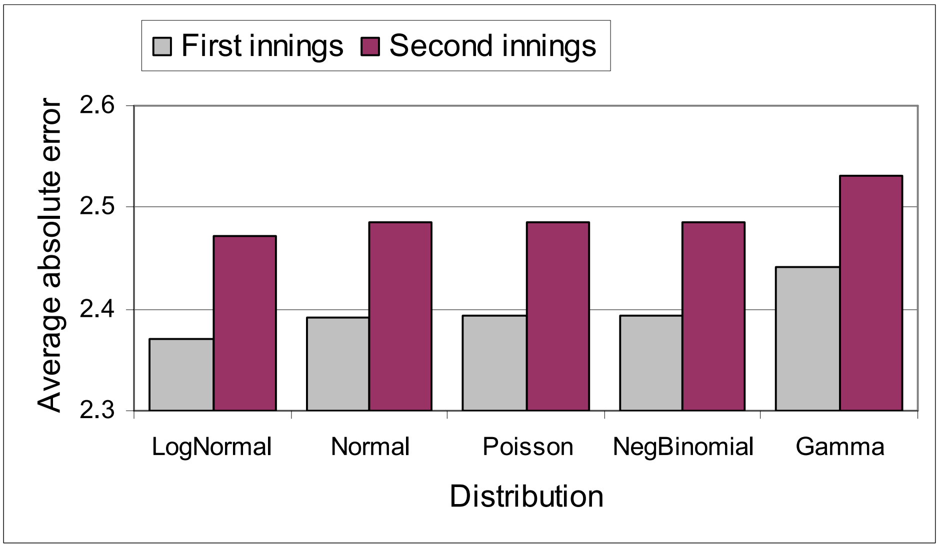

Figure 9-17 Comparison of AAE for the five distributions

The AAE between the predicted and actual scores was compared for the five distributions and can be seen in Figure 9-17. Clear differences exist between the first and second innings with the first innings being more than twice as predictable as the second innings. Overall, the log-normal approach was found to produce the lowest AAE for both the first and second innings.

In additional to modelling five differing distributions, two logistic regression approaches were also used. Firstly, an ordinal logistic regression was applied to the 11 possible outcomes (0, 1, 2…9, >9). This model incorporates a different intercept term for each additional run, but has fixed parameter estimates for each run category. To further enable variation in the parameter estimates for each run category, a series of 10 binomial models were constructed. The first model predicted whether the number of runs scored would be greater than zero, whilst the second model predicted whether the number of runs would be greater than one, and so on. By subtracting sequential models the probability for each run category could be determined.

9.5.1 Hold-out Sample

To avoid the bias associated with over-fitting, predictive capacity must be assessed in a sample of data separate from that which parameter estimates were derived. This was achieved by developing parameter estimates from all data prior to 2003 and applying to matches played in 2003 and 2004 (169 matches). Goodness of fit was determined by averaging the log of the predicted probabilities for the actual runs scored, with the model producing the highest average, indicative of the best fit to the data (see section 3.7). Ordinarily, given the size of the data sets being used (21,000 in training, 7000 in holdout); very little difference should be expected between the holdout and training samples.

9.6 Results

Although the 14-parameter multivariate binomial approach was found to produce the best fit to the training data, from Figure 9-18 and Figure 9-19 the large difference between the training data and the holdout data in both innings indicated over-fitting. To alleviate the bias of over-fitting, the four 'between match' variables (Team, Home country, Team X Stage and Opposition X Stage) were removed, creating a new model assigned the name Reduced Binomial. The similarity between the average log of the probabilities in the training and holdout samples for the Reduced Binomial model, suggest that the source of bias due to over-fitting had been removed. For further comparison, two additional models were incorporated, one that assigned an equal probability of 0.091 to the 11 run categories (Equal), and a second comparison model that assigned the actual probabilities derived from the training data set (Actual). Log transformations of the successful probabilities included a constant so that the Equal model was calibrated to zero.

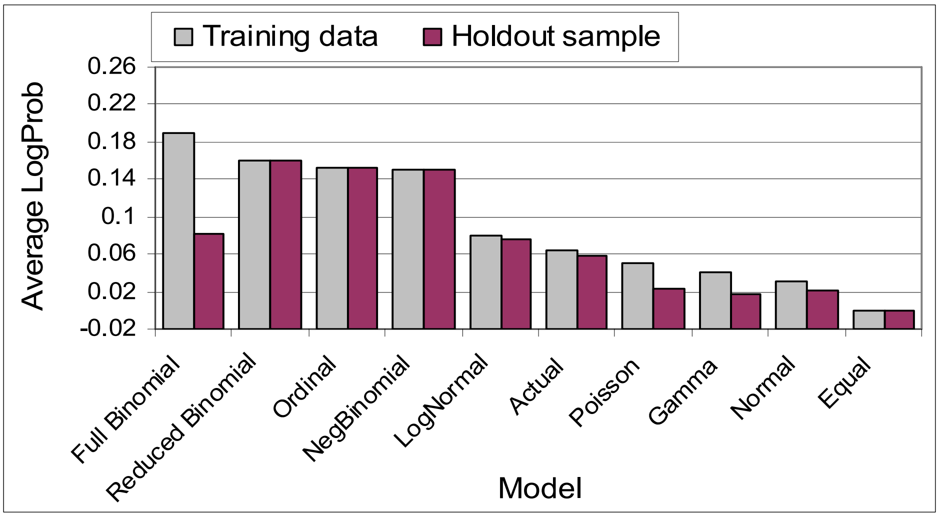

Figure 9-18 Comparison of average LogProb between models for first innings

From Figure 9-18 and Figure 9-19 it is possible to see that the Reduced Binomial model appeared the best performed model in both the first and second innings. Although no statistically significant differences existed between the Reduced Binomial, Ordinal Logistic and Negative Binomial models, all three of these approaches were significantly better than all others ($p<0.0001$) for both first innings and second innings.

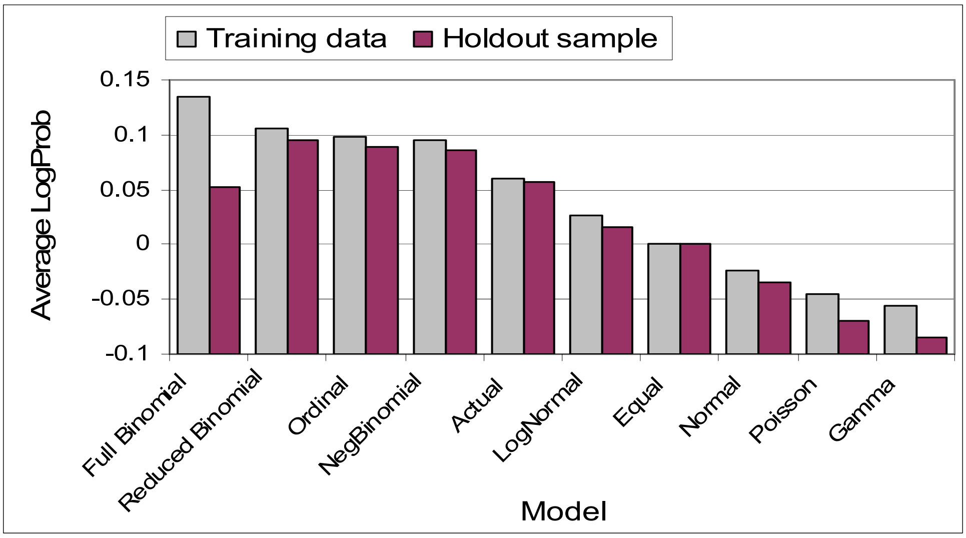

Figure 9-19 Comparison of average LogProb between models for the second innings

9.7 Discussion

Using past data it is possible to identify features of the match that can aid in the prediction of the number of runs scored per over. Of the 10 approaches examined, the slivered binomial approach appeared to produce the best fit to the data, although due to the increased number of models required, considerable care must be taken to ensure that the data is not over-fitted. Over fitting is best avoided by choosing highly significant variables with lower degrees of freedom. Both an ordinal logistic regression and a negative binomial approach produced good fits to the data and were more simplistic in implementation.

The use of mathematical models to predict RPO during the course of the match opens new doors for bookmakers. The speed and objectivity of a mathematical approach supersedes traditional bookmaking methods. A mathematical model, incorporating simple past features of the game enables a dynamic price setting process and creates a much greater scope of betting opportunities for the punter.

Based on a model developed during this thesis, betting on the number of runs to be scored per over was successfully implemented by Ladbrokes of London during the 2003 World Cup of cricket. Probabilities for the first over of each innings were set based on past performance of each team, whilst the probability for all subsequent overs were determined incorporating features of the innings past. At the completion of each over, the number of runs scored and wickets taken were entered and a new set of prices for the next over calculated. Betting was open to punters in the brief period of time between overs, with the short time span to immediately set odds for the next over, upon completion of the last, discounting traditional bookmaking methods. This process has been successful in generating profits close to the calculated bookmaker percentages and has continued to operate successfully in subsequent ODIs.

10 Conclusion

10.1 Summation

Each of the six analyses conducted in this dissertation provide a unique contribution to the literature relating to the prediction of sporting events and the efficiency of the corresponding wagering market, should it be in existence.

Chapter 4 builds on a modelling process developed by Clarke (1993), by providing a prediction for the outcome of AFL football that numerically quantifies two of the three components thought to explain the phenomena of HA. This analysis further incorporates the effect of individual players, to produce predictions that can be shown to be significantly better than the benchmark models created by Clarke. Using a standard wagering strategy, this analysis also shows the AFL betting markets to be statistically inefficient over a seven year period (1997-2003).

Using information gathered from all 2164 ODI matches played prior to July 2004, chapter 5 shows that the MOV developed using a D/L approach displays bias towards the side batting second. This analysis further confirms that the modifications incorporated by de Silva, provide an unbiased prediction of MOV that is well approximated by a Normal distribution. Constructing prediction models at a team and individual level, this analysis further explores HA as it applies to ODI cricket and provides a prediction process that shows inefficiency to exist in the wagering market for the outcome of ODI matches.

Having demonstrated the application of statistically driven models to predict match outcomes, the second stage of the analysis sought to apply a similar process to the prediction of individual player performance.

Using in excess of 50,000 data points, chapter 6 develops prediction models for individual player performance in AFL football. By establishing that the number of possessions gathered by AFL players can be well approximated by a Normal distribution, a range of predictions models are considered, facilitating a linear relationship to be established between goodness of fit and ROI. As a consequence, a 15 variable multiple linear regression is progressively constructed, with each stage of model development indicating a greater degree of inefficiency to exist in the player HtH betting markets for AFL football.

Chapter 7 also develops models for player performance, this time relating to ODI cricket. By using log-linear and linear modelling, univariate and multivariate models were constructed to predict the number of runs to be scored by each batsman in ODI matches. By applying prediction models to bookmaker data collected from the 2003 World cup of cricket, inefficiencies can once again be shown to exist in betting markets for player HtH performances.

Practical applications for statistically driven models can clearly be seen to exist when predicting match outcomes and individual player performances. The third stage of this analysis sought to apply this same prediction approach to a wider range of sporting outcomes. By considering non-normally distributed outcomes and creating real time predictions that can be constantly updated, the consistency and versatility of a statistical driven approach is realised.

Building on work previously developed by the author, chapter 8 uses ordinal logistic regression to identify match and player features that can be linked to the polling of Brownlow votes in AFL football. By applying this prediction model to matches played throughout the course of the AFL season, each individual player was assigned a probability of polling votes. By aggregating individual predictions, each player could then be assigned a probability of winning the Brownlow medal. A series of measures were developed to determine goodness of fit along with the minimal size of holdout samples required to accurately develop such models. While wagering markets are not yet available for Brownlow betting per game, this modelling process establishes a benchmark for future markets.

Chapter 9 uses information from 50,000 overs to develop a prediction model for the number of runs to be scored per over in ODI cricket. A range of different modelling approaches were considered and compared. Using only variables that had previously occurred in the innings, a statistically driven model was subsequently employed by Ladbrokes of London during the 2003 World Cup of cricket to become the first bookmaker to offer punters the opportunity to bet on the number of runs to be scored per over, with betting open until the start of each over.

The common theme amongst all analysis conducted is the wealth of past data used to determine statistical significance. Provided sufficient past data exists and an outcome can be well approximated by a known distribution, a prediction model can be developed. This not only provides a guide to the relative predictability of a given outcome but also creates unbiased predictions for future events. Ultimately, the quality of predictions will be dependent upon the inherit predictability of the outcome and the quality of the data used to develop models.

While the evidence in this thesis would suggests that multivariate modelling is better suited to the prediction of individual performances over team performances, clear benefit can still be obtained when modelling match outcomes. In particular, a greater understanding of HA is now known for both AFL football and ODI cricket. In AFL football, the effects of distance travelled and familiarisation at the venue can clearly be identified as components of HA. In ODI cricket, while HA is a highly significant predictor of outcome, the effects of distance travelled and familiarity cannot be readily identified as components of HA.

Because team sports such as AFL football and cricket, are dependent upon the performance of the individuals within the team, the collection of information at an individual player level provides an increase in the quality of data relating to the team. Despite the increase in time and effort required to collect data at an individual level, this increase in quality does result in more accurate prediction models. While the added benefit of modelling match outcomes at an individual level was small for both AFL football and cricket, it was statistically significant for both. Although it is difficult to measure the impact that specific individuals may have on a particular sport, it could be hypothesised that the added benefit of modelling at an individual player level would have greater impact in sports where teams had less playing members, such as basketball.

10.2 Statistically driven prediction models

The advent of the computer has seen a marked change in many professions. One profession to benefit greatly from the invention of the computer is that of the statistician. The ever increasing speed and memory capacity of the modern computer enables analysis of millions of data points in only a matter of minutes. The advent of the Internet has further increased the availability of the large amounts of data required to successful develop statistically driven prediction models for sporting outcomes.

Drawing on information from the past enables an 'evidence based' approach to the future that is widely applicable to all walks of life. The author Herb Brody was once quoted, "Telling the future by looking at the past assumes that conditions remain constant. This is like driving a car by looking in the rearview mirror." While sporting events are constantly evolving, some features of the event will remain the same. The use of sufficiently large past data enables constants to be identified, so, whilst predicting the future may be difficult, it is indeed possible.

By using multivariate models, a combination of statistically important variables can be combined and weighted to produce a prediction model. In theory, the model that can explain the greatest amount of variation in past results will be the model that produces the best predictions. By reducing the level of statistical significance for inclusion in to the model, the probability of producing a Type I error can be reduced, further increasing the robustness of the prediction approach. Results from this dissertation establish a relationship between the quality of models developed on the training data and the quality of the resulting prediction model. It then follows that the model that best predicts the outcome is also the model that will produce the greatest ROI should the sporting event of interest have a corresponding wagering market. Although only two sports were assessed in this dissertation, the range and diversity of outcomes considered suggest a much wider application to sports in general. While statistical analysis can be conducted on any outcome regardless of the underlying distribution, a statistically driven modelling approach will always be better suited when the chosen outcome can be well approximated by a Normal distribution.

The process of using computers to model sporting outcomes may seem simple in practice, but requires substantial time, effort and diversity of skills. A statistically driven prediction approach can only ever be as good as the data used to derive the results. While automated web collection programs can quickly and easily download web pages from the Internet, extensive programming is inevitably required to ensure that the data is collated into a workable format. Data collection can be divided into the collection of past data used to determine prediction models, and the collection of present data to enable updated predictions. While large amounts of historical data can often be collected and collated in one process, it is the continuing maintenance of data that can be most time consuming. With websites often changing format and structure, a collection processes must be sufficiently versatile enough to ensure that data can be continually retrieved in a consistent format.

Scottish author and poet, Andrew Lang was once heard describing a colleague, “He uses statistics as a drunken man uses lamp-posts - for support rather than for illumination.” The careful application of statistics is essential for obtaining and judging predictions, thus any statistically driven prediction approach requires both a statistical package and the knowledge to use it. While most statistical packages can readily determine statistical significance, few offer the data manipulation capacities of SAS. Of particular benefit is the ability to write and store programs relating to the management and analysis of the data. Over half a million lines of SAS code have been written developing this body of work. Whilst this may reflect the author's programming inefficiency, it also offers the reader an insight into man-hours required to develop a systematic modelling process. This mirrors the beliefs of (Benter 1994) who also noted that several man-years of programming and data analysis would probably be necessary to develop a profitable system for horse racing in Hong Kong.

Whilst establishing a systemic approach to mathematically modelling sporting outcomes may be of great interest to sports fans, it is the determination of market efficiency that remains the most practical application of this body of work.

10.3 Market efficiency

One of the primary purposes for predicting the winner of a sporting event is to obtain financial gain by wagering on a successful outcome. Traditional betting markets have operated for hundreds of years in a sport such as racing, where the sport primarily survives because of the associated betting opportunities. In recent years, betting has expanded into virtually all other sports in which there has been wide public interest prior to and independent of any betting activity. Whilst initial books were set on the winner of matches, the growth of Internet betting and increased competition between bookmakers has resulted in many exotic bets being created. Whether it is possible to successfully profit from exotic betting opportunities is dependent upon the magnitude of the bookmaker percentage and the accuracy with which bookmakers set their prices.

In fixed price bookmaking, the sum of the probabilities that a bookmaker will offer on a given event will exceed one[9]. The amount by which the summed probabilities exceed one is often referred to as the bookmaker margin, and represents the profit the bookmaker stands to win should he balance his books perfectly. The size of the margin is proportional to the number of competitors in the event, with most multinomial outcomes such as horse racing having a margin in excess of 15 %. For competitions in which there

[9] Bookmaker probabilities are given by the reciprocal of the price

are only two outcomes, the margin is generally only about 7-8%. Should competition exist between bookmakers, this margin can be reduced even further.

Market inefficiency represents a difference in the accuracy of prediction between bookmakers and punters, and is thus dependent upon the underlying predictability of the outcome of interest. If a sporting outcome is completely random, such as the tossing of a coin, then it would be impossible for a punter to derive long term profit, as he would not be able to consistently predict the outcome more accurately than a bookmaker. The more predictable an outcome is, the more accurate a bookmaker must be in setting his prices. Results from this dissertation suggest that the more predictable an outcome is, the greater the scope for exploitation of market inefficiency.

Four specific betting markets covering team and player performances for football and cricket have been explored for efficiency. While markets were chosen for the reduced bookmaker percentage present in two outcome events, the primary reason for inefficiencies in these markets is because of the brief period of time in which these fixed betting markets have been available to the public. In statistical terms, a training effect is a known phenomenon affecting all walks of life – the more often you do something, the better at it you become. A good example of this can be observed by the statistical significance of 'experience' when predicting player performance in football and cricket; the greater the experience, the better the performance. With fixed price betting only becoming available in Australia in the last eight years, the relative immaturity of the betting markets coupled with the smaller bookmaker's percentage taken from two outcome events combine to create betting market that have some degree of inefficiency. The limited number of literary contributions relating to the prediction of AFL and ODI cricket may also contribute to this inefficiency. Given the relative efficiency of established betting markets around the world, it is unlikely that the betting markets considered in this doctorate will remain inefficient in the future.

The most established of the four markets considered, is fixed price betting on the outcome of AFL football matches. Whilst results indicate that a statistically significant profit has been available over a seven year period, the magnitude of this profit has decreased annually. This could well be attributed to a greater understanding and acceptance of the benefit of mathematical models in objectively assigning each teams probability of success. The incorporation of such processes in turn leading to an increase in the efficiency of the betting market.

Although evidence of inefficiency is apparent in Australian betting markets, the exact magnitude of this inefficiency is difficult to define. To help facilitate a balanced book, bookmakers will open books a few days prior the start of the sporting event, allowing time to adjust their prices in accordance with supply and demand. It is the author's experience that the most glaring inefficiencies identified by a modelling process are often also identified by astute punters, effectively reducing the price and the magnitude of the inefficiency. For the purpose of this thesis, calculations for market efficiency have been made based on the assumption that the collected price is indeed the price available at the time of bet placement. Despite the speed of a computer based approach, this is clearly not the case, and ROI figures used in this dissertation more accurately reflect the maximal profit that can be made.

To establish consistency, prices used in this thesis were generally collected 24 hours prior to the commencement of the sporting event. To measure market inefficiency more accurately would require prices to be collected at more than one point in time. Ultimately, the best way to determine efficiency would be to collect prices when books first open and then collect prices again, just prior to the event starting. This would enable an accurate reflection of the true efficiency of the wagering market. In addition, the change between opening and closing prices would reflect the demand from the sporting public, which in turn could be used to predict the sporting outcome.

While it is hypothesised that mathematical modelling will replace traditional price setting techniques, this process may occur indirectly. Given the majority of sports betting takes place via the Internet, and complete records are kept for each client, it follows that an astute bookmaker can benefit greatly from the information provided from his client database. Statistical techniques can be utilised to identify punters that can consistently predict outcomes more accurately than the bookmaker. Having identified knowledgeable clients, the bookmaker can then adjust prices accordingly. Undoubtedly, the greatest inefficiencies in betting market will exist when bookmakers originally post their prices. The ensuing time until commencement of the sporting event allow the bookmaker to effectively compensate for price setting mistakes. The establishment of maximum bet sizes further reduce the bookmakers' chance of exploitation by astute punters.

Unless a bookmaker can derive profit from a particular sporting event, he would not operate a book. The existence of long-term market inefficiency suggests that bookmakers can survive with an inefficient price setting process, with the cost of this inefficiency effectively passed from the bookmaker back to the less efficient punters.

The use of statistically derived prediction models can clearly be seen to benefit the sports betting community. While those wishing to wager on sporting outcomes may derive short-term advantage from this approach, in the long-term, the speed and objectivity provided by computer driven models, will plainly benefit prices setting by bookmakers for betting both before and during sporting outcomes.

While the multivariate techniques used in this thesis are well documented in the field of medicine, the application of such techniques to large bodies of sporting data remains limited. As the age of computers continues to evolve, so too will the availability and importance of large data sets. This in turn will bring a greater need for a systematic approach to derive useful information from this data, further enhancing the practical benefits of this body of work. While linear regression techniques may in the future be superceded by data mining approaches such as neural networking, the simplicity of a least squares approach will always appeal to those looking to crawl before they can walk.

11 References

Allsopp, P. and S. R. Clarke (2004). "Rating cricket teams and analysing outcomes in one-day and test cricket." Journal of the Royal Statistical Society (Series A) 167: 657-667.

Anon (2000). Hair's to all those Brownlow hopefuls. The Age. Melbourne: 3 (Sport section).

Bailey, M. (2000). Identifying arbitrage opportunities in AFL betting markets through mathematical modelling. Proceedings of the Fifth Australian conference on Mathematics and Computers in Sport. G. Cohen and T. Langtry. Sydney, University of Technology Sydney: 37-42.

Bailey, M. (2001). Who'll win the Brownlow: Permutations, combinations, probabilities and guesswork. AFL Record. 1st preliminary final: 80-81.

Bailey , M. and S. R. Clarke (2002). Predicting the Brownlow medal winner. Proceedings of the sixth Australian conference on mathematics and computers in sport. Bond University, G. Cohen and T. Langtry: 56-62.

Bailey , M. J. and S. R. Clarke (2004a). Market Inefficiencies in player head to head betting on the 2003 cricket world cup. Economics, Management and Optimization in sport. S. Butenko, J. Gil-Lafuente and P. Pardalos. Heidelberg, Springer-Verlag:185-201.

Bailey, M., and Clarke, S. R. (2004b). Deriving a profit from Australian Rules football: A statistical approach. Proceedings of the Seventh Australian conference on Mathematics and Computers in Sport. Palmerston North, Massey University, R Hugh Morton & S Ganesalingam: 48-56.

Barnett, V. and S. Hilditch (1993). "The effect of an artificial pitch surface on home team performance in football (soccer)." Journal of the Royal Statistical Society (Series A) 156(1): 39-50.

Benter, W. (1994). Computer Based Horse race handicapping and wagering systems: a report. Efficiency of racetrack betting markets. L. Hausch, Ziemba. San Diego, Academic Press Limited:183-198.

Brailsford, T. J., P. K. Gray, et al. (1995). "The Efficiency of Australian Football Betting Markets." Australian Journal of Management 20(2): 167-196.

Brecher, S. L. (1980). Beating the Races with a computer. Long Beach, Software Supply.

Chapman, R. G. (1994). Still searching for positive returns at the track: Empirical results from 2,000 Hong Kong Races. Efficiency of racetrack betting markets. L. Hausch, Ziemba. San Diego, Academic Press Limited: 173-181.

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. (1991). Consistency in sport - with particular reference to cricket. Proceedings 27th NZ OR conference. Wellington, New Zealand Operations Research Society: 30-35.

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

Clarke, S. R. (1998). "Test Statistics." Statistics in Sport. J. Bennett, London, Arnold: 83-101.

Clarke, S. R. (2005). "Home advantage in the AFL." Journal of Sports Sciences in press.

Clarke, S. R. and J. M. Norman (1995). "Home ground advantage of individual clubs in English soccer." The Statistician 44, No 4, p509-521.

Courneya, K. S. and A. V. Carron (1992). "The home advantage in sport competitions: A literature review." Journal of Sport & Exercise Psychology 14: 13-27.

de Silva, B., G. Pond, et al. (2001). "Estimation of the magnitude of victory in one-day cricket." Australian & New Zealand Journal of Statistics 43(3): 259-268.

Davis, C (2000) "Best of the Best." Sydney; ABC Books

Derksen, S. and H. J. Keselman (1992). "Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables." British Journal of Mathematical and Statistical Psychology 45: 265-282.

Dowe, D. L., G. E. Farr, et al. (1996). Information theoretic football tipping. Third conference on: Mathematics & Computers in Sport conference. N. de Mestre. Gold coast, Qld., Bond University: 233-242.

Duckworth, F. and T. Lewis (1998). "A fair method for resetting the target in interrupted one-day cricket matches." Journal of the Operational Research Society 49: 220-227.

Efron, B. and C. Morris (1975). "Data analysis using Stein's estimator and its generalizations." Journal of the American Statistical Association 70(350): 311-318.

Elderton, W. E. (1945). "Cricket scores and some skew correlation distributions." Journal of the Royal Statistical Society (Series A) 108: 1-11.

Elderton, W. P. and E. M. Elderton (1909). Primer of Statistics. London, Black.

Elo, A. E. (1978). The rating of chess players, past and present. London, Batsford.

Fama, E. F. (1970). "Efficient capital markets: a review of theory and empirical work." Journal of Finance 25: 383-475.

Gauss, C. F. (1809). Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium. Perthes et Besser, Hamburg.Z . Perthes et Besser. Hamburg.Z .

Harrell, F. E., K. L. Lee, et al. (1996). "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors." Stat Med. 15(4): 361-87.

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

Hausch, W., V. Lo, et al. (1994). Efficiency of Racetrack Betting Markets, Academic Press, Inc.

Holder, R. L. and A. M. Nevill (1997). "Modelling performance at international tennis and golf tournaments: is there a home advantage?" The Statistician 46(4): 551-559.

Hurvich, C. M. and C. L. Tsai (1990). "The impact of model selection on inference in linear regression." American Statistician 44: 214-217.

James, W. and C. Stein (1981). "Estimation with quadratic loss." Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1: 361-379.

Kelly, J. (1956). "A New Interpretation of Information Rate." Bell System Technical Journal 35: 917-926.

Kimber, A. C. and A. R. Hansford (1993). "A statistical analysis of batting in cricket." Journal of the Royal Statistical Society A 156(3): 443-445.

Kullback, S. and R. A. Leibler (1951). "On Information and Sufficiency 22 pp79-86 1951." Annals of Math. Stats 22: 79-86.

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

Levitt, S. D. (2003). How Do Markets Function? An Empirical Analysis of Gambling on the National Football League. NBER Working Papers9422,. I. National Bureau of Economic Research.

MacLean, L., W. Ziemba, et al. (1992). "Growth Versus Security in Dynamic Investment Analysis." Management Science 38(11): 1562-1585.

Markov, A. A. (1900). Wahrscheinlichkeitsrechnung. Tebrer. Leipzig.

Pollard, R. (1986). "Home advantage in soccer: A retrospective analysis." Journal of Sport Sciences 4: 237-248.

Pollard, R., B. Benjamin, et al. (1977). Sport and the negative binomial distribution. Optimal Strategies in Sports. S. P. Ladany and R. E. Machol. Amsterdam, North Holland: 188-195.

Reep, C., R. Pollard, et al. (1971). "Skill and chance in ball games." Journal of the Royal Statistical Society A 134: 623-629.

Sauer, R. D. (1998). "The economics of wagering markets." Journal of Economic Literature 36: 2021-2060.

Schwartz, B. and S. E. Barsky (1977). "The home advantage." Social Forces 55: 641-661.

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. (1998). A taxonomy and survey of sports rating systems. Statistics In Sport. J. Bennett. New Jersey, Arnold: 253-258.

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

Winkler, R. L. and A. H. Murphy. (1968). "Good probability assessors." J. Applied Meteorology 7: 751–758.

Wood, G. H. (1941). "What do we mean by consistency?" The Cricketer Annual: 22-28.

Wood, G. H. (1945). "Cricket scores and geometrical progression." Journal of the Royal Statistical Society (Series A) 108: 12-22.

Yelas, S. (2003). Analysis of Swinburne computer performance in identiying inefficiencies in the AFL betting market. Mathematical sciences. Melbourne, Swinburne University.

12 Appendices

Appendix 1 Published multivariate analysis conducted by candidate

Abramson, M. J., M. J. Bailey, F. J. Couper, J. S. Driver, O. H. Drummer, A. B. Forbes, J. J. McNeil, E. Haydn Walters, and G. Victorian Asthma Mortality Study, 2001, Are asthma medications and management related to deaths from asthma?[see comment]: American Journal of Respiratory & Critical Care Medicine, v. 163, p. 12-8.

Abramson, M. J., M. J. Bailey, A. B. Forbes, E. H. Walters, and G. Victorian Asthma Mortality Study, 2003, How well do doctors know their patients with severe asthma?: Internal Medicine Journal, v. 33, p. 557-65.

Beach, J. R., J. Raven, C. Ingram, M. Bailey, D. Johns, E. H. Walters, and M. Abramson, 1997, The effects on asthmatics of exposure to a conventional water-based and a volatile organic compound-free paint: European Respiratory Journal, v. 10, p. 563-6.

Brand, C., J. Snaddon, M. Bailey, and F. Cicuttini, 2001, Vitamin E is ineffective for symptomatic relief of knee osteoarthritis: a six month double blind, randomised, placebo controlled study: Annals of the Rheumatic Diseases, v. 60, p. 946-9.

Carter, V. M., J. F. Hoy, M. Bailey, P. G. Colman, I. Nyulasi, and A. M. Mijch, 2001, The prevalence of lipodystrophy in an ambulant HIV-infected population: it all depends on the definition: HIV Medicine, v. 2, p. 174-80.

Chong, A. H., R. D. Sinclair, C. W. Chow, M. Bailey, and E. Fitzpatrick, 2000, Loose anagen hair syndrome shows evidence of an autosomal dominant pattern of inheritance with variable expression and incomplete penetrance: results from a study of 16 families: British Journal of Dermatology, v. 143, p. 127-128.

Cicuttini, F., A. Forbes, K. Morris, S. Darling, M. Bailey, and S. Stuckey, 1999, Gender differences in knee cartilage volume as measured by magnetic resonance imaging: Osteoarthritis & Cartilage, v. 7, p. 265-71.

Cicuttini, F. M., A. Wluka, M. Bailey, R. O'Sullivan, C. Poon, S. Yeung, and P. R. Ebeling, 2003, Factors affecting knee cartilage volume in healthy men: Rheumatology, v. 42, p. 258-62.

Davis, S. R., E. M. Briganti, R. Q. Chen, F. S. Dalais, M. Bailey, and H. G. Burger, 2001, The effects of Chinese medicinal herbs on postmenopausal vasomotor symptoms of Australian women. A randomised controlled trial.[see comment]: Medical Journal of Australia, v. 174, p. 68-71.

de Jong, T. A., S. A. Skinner, C. Malcontenti-Wilson, D. Vogiagis, M. Bailey, I. R. van Driel, and P. E. O'Brien, 2000, Inhibition of rat colon tumors by sulindac and sulindac sulfone is independent of K-ras (codon 12) mutation: American Journal of Physiology - Gastrointestinal & Liver Physiology, v. 278, p. G266-72.

Desmond, C. P., M. Bailey, and S. K. Roberts, 2004a, Ursodeoxycholic acid and liver disease in an adult cystic fibrosis population: Journal of Gastroenterology & Hepatology, v. 19, p. A272.

Desmond, C. P., S. K. Roberts, F. Dudley, J. Mitchell, S. Nguyen, M. Bailey, and S. Pianko, 2004b, Durability of sustained virological response in hepatitis C treated patients: Journal of Gastroenterology & Hepatology, v. 19, p. A251.

Desmond, C. P., S. K. Roberts, S. Nguyen, F. Dudley, J. Mitchell, C. Day, M. Bailey, and S. Pianko, 2004c, Evolution of sustained virologic response rates with interferon based therapies in the clinic setting: Journal of Gastroenterology & Hepatology, v. 19, p. A252.

Dharmage, S., M. Bailey, J. Raven, K. Abeyawickrama, D. Cao, D. Guest, J. Rolland, A. Forbes, F. Thien, M. Abramson, and E. H. Walters, 2002, Mouldy houses influence symptoms of asthma among atopic individuals: Clinical & Experimental Allergy, v. 32, p. 714-20.

Dharmage, S., M. Bailey, J. Raven, A. Cheng, J. Rolland, F. Thien, A. Forbes, M. Abramson, and E. H. Walters, 1999a, Residential characteristics influence Der p 1 levels in homes in Melbourne, Australia.[see comment]: Clinical & Experimental Allergy, v. 29, p. 461-9.

Dharmage, S., M. Bailey, J. Raven, T. Mitakakis, A. Cheng, D. Guest, J. Rolland, A. Forbes, F. Thien, M. Abramson, and E. H. Walters, 2001, Current indoor allergen levels of fungi and cats, but not house dust mites, influence allergy and asthma in adults with high dust mite exposure: American Journal of Respiratory & Critical Care Medicine, v. 164, p. 65-71.

Dharmage, S., M. Bailey, J. Raven, T. Mitakakis, D. Guest, A. Cheng, J. Rolland, F. Thien, M. Abramson, and E. H. Walters, 1999b, A reliable and valid home visit report for studies of asthma in young adults: Indoor Air, v. 9, p. 188-92.

Dharmage, S., M. Bailey, J. Raven, T. Mitakakis, F. Thien, A. Forbes, D. Guest, M. Abramson, and E. H. Walters, 1999c, Prevalence and residential determinants of fungi within homes in Melbourne, Australia.[see comment]: Clinical & Experimental Allergy, v. 29, p. 1481-9.

Egan, M. K., R. G. Langham, M. Richardson, P. Bergin, D. Kaye, M. Bailey, J. Dowling, N. M. Thomson, and A. N. Stein-Oakley, 1999, Quantitative analysis of PDGFA, PDGFB, PDGF receptor beta and TGF-beta 1 mRNA in cardiac transplant biopsies: Transplantation Proceedings, v. 31, p. 131-2.

Finlayson, F. A., D. W. Reid, M. Bailey, V. Jones, M. Lenkic, and J. Wilson, 2002, Severity of disease predicts health care utilisation in adults with cystic fibrosis(CF): Respirology, v. 7, p. A45.

George, J., Y. T. Phun, M. J. Bailey, D. C. Kong, and K. Stewart, 2004, Development and validation of the medication regimen complexity index: Annals of Pharmacotherapy. 38(9):1369-76, 2004 Sep.

Glare, E. M., M. Divjak, M. J. Bailey, and E. H. Walters, 2001, The usefulness of competitive PCR: airway gene expression of IL-5, IL-4, IL-4delta2, IL-2, and IFNgamma in asthma: Thorax, v. 56, p. 541-8.

Glare, E. M., M. Divjak, M. J. Bailey, and E. H. Walters, 2002, beta-Actin and GAPDH housekeeping gene expression in asthmatic airways is variable and not suitable for normalising mRNA levels.[see comment]: Thorax, v. 57, p. 765-70.

Gordon, A., M. J. Bailey, and S. K. Roberts, 2003, The global clinical assessment: an accurate predictor of cirrhosis in chronic hepatitis C: Journal of Gastroenterology & Hepatology, v. 18, p. B15.

Hellard, M. E., M. I. Sinclair, S. C. Dharmage, M. J. Bailey, and C. K. Fairley, 2002, The rate of gastroenteritis in a large city before and after chlorination: International Journal of Environmental Health Research, v. 12, p. 355-60.

Hellard, M. E., M. I. Sinclair, C. K. Fairley, R. M. Andrews, M. Bailey, J. Black, S. C. Dharmage, and M. D. Kirk, 2000, An outbreak of cryptosporidiosis in an urban swimming pool: why are such outbreaks difficult to detect?: Australian & New Zealand Journal of Public Health, v. 24, p. 272-5.

Hurst, W., M. Bailey, and B. Hurst, 2004, Prevalence of external auditory canal exostoses in Australian surfboard riders: Journal of Laryngology & Otology, v. 118, p. 348-51.

Kemp, W. W., S. Pianko, S. Nguyen, M. Bailey, and S. K. Roberts, 2003, Survival in hepatocellular carcinoma (HCC): impact of screening and aetiology of liver disease: Journal of Gastroenterology & Hepatology, v. 18, p. B85.

Keogh, A., S. Fenton, C. Leslie, C. Aboyoun, P. Macdonald, Y. Zhao, M. Bailey, and F. Rosenfeldt, 2003, Randomised Double-Blind, Placebo-Controlled Trial of Coenzyme Q10 Therapy in Class II and III Systolic Heart Failure: Heart Lung Circulation, v. 12, p. 135.

King, S. J., D. J. Topliss, T. Kotsimbos, I. B. Nyulasi, M. Bailey, P. R. Ebeling, and J. W. Wilson, 2005, Reduced bone density in cystic fibrosis: deltaF508 mutation is an independent risk factor: European Respiratory Journal, v. 25, p. 54-61.

Knott, J. C., S. L. Tan, A. C. Street, M. Bailey, and P. Cameron, 2004, Febrile adults presenting to the emergency department: outcomes and markers of serious illness: Emergency Medicine Journal, v. 21, p. 170-174.

Levvey, B., L. Zheng, B. Orsida, L. Law, M. Bailey, T. Kotsimbos, H. Whitford, T. Williams, and G. Snell, 2004, Everolimus favourably alters lung allograft immune responses: results of a prospective randomised trial: Transplantation, v. 78, p. 189.

Li, X., C. Ward, F. Thien, R. Bish, T. Bamford, X. Bao, M. Bailey, J. W. Wilson, and E. Haydn Walters, 1999, An antiinflammatory effect of salmeterol, a long-acting beta(2) agonist, assessed in airway biopsies and bronchoalveolar lavage in asthma.[see comment]: American Journal of Respiratory & Critical Care Medicine, v. 160, p. 1493-9.

Lord, L., S. May, M. Bailey, and L. Smith, 2003, Is one head and neck immobilization system as good as another? One center's experience: Medical Dosimetry, v. 28, p. 39-43.

Mariani, J., R. Ou, M. Bailey, M. Rowland, P. Nagley, F. Rosenfeldt, and S. Pepe, 2000, Tolerance to ischemia and hypoxia is reduced in aged human myocardium: Journal of Thoracic & Cardiovascular Surgery, v. 120, p. 660-7.

Michaelides, A., L. Liolios, E. M. Glare, D. W. Spelman, M. J. Bailey, E. H. Walters, T. J. Williams, G. I. Snell, and T. C. Kotsimbos, 2001, Increased human cytomegalovirus (HCMV) DNA load in peripheral blood leukocytes after lung transplantation correlates with HCMV pneumonitis: Transplantation, v. 72, p. 141-7.

Munro, P. E., M. J. Bailey, J. A. Smith, and G. I. Snell, 2003, Lung volume reduction surgery in Australia and New Zealand. Six years on: registry report: Chest, v. 124, p. 1443-50.

New, P. W., H. B. Rawicki, and M. J. Bailey, 2002, Nontraumatic spinal cord injury: demographic characteristics and complications: Archives of Physical Medicine & Rehabilitation, v. 83, p. 996-1001.
New, P. W., H. B. Rawicki, and M. J. Bailey, 2004, Nontraumatic spinal cord injury rehabilitation: pressure ulcer patterns, prediction, and impact: Archives of Physical Medicine & Rehabilitation, v. 85, p. 87-93.
Opat, A. J., M. M. Cohen, M. J. Bailey, and M. J. Abramson, 2000, A clinical trial of the Buteyko Breathing Technique in asthma as taught by a video: Journal of Asthma, v. 37, p. 557-64.
Padiglione, A. A., J. Willis, M. Bailey, and C. K. Fairley, 1999, Characteristics of patients with community-acquired pneumococcal pneumonia.[see comment]: Medical Journal of Australia, v. 170, p. 165-7.
Palanca, S., D. M. Taylor, M. Bailey, and P. A. Cameron, 2003, Mechanisms of motor vehicle accidents that predict major injury: Emergency Medicine (Fremantle, W A), v. 15, p. 423-8.
Pilcher, D. V., M. J. Bailey, D. F. Treacher, S. Hamid, A. J. Williams, A. C. Davidson, C. P. Desmond, S. K. Roberts, S. Nguyen, F. Dudley, M. Bailey, S. Pianko, J. Mitchell, C. Day, B. Levvey, L. Zheng, B. Orsida, L. Law, T. Kotsimbos, H. Whitford, T. Williams, G. Snell, J. C. Knott, S. L. Tan, A. C. Street, and P. Cameron, 2005, Outcomes, cost and long term survival of patients referred to a regional weaning centre: Thorax, v. 60, p. 187-192.
Rayner, C. R., L. L. Ioannides-Demos, W. J. Spicer, J. D'Agostino, M. Bailey, and H. Krum, 1999, Effect of collection tube on measured gentamicin and vancomycin concentrations: American Journal of Health-System Pharmacy, v. 56, p. 1457-8.
Reid, D. W., N. Misso, R. Bish, M. Bailey, P. Thomson, and E. H. Walters, 2003, Zafirlukast in persistent asthma; evidence for tollerance on treatment and rebound on cessation: Respirology, v. 8, p. A19.
Roebuck, T., P. Solin, D. M. Kaye, P. Bergin, M. Bailey, and M. T. Naughton, 2004, Increased long-term mortality in heart failure due to sleep apnoea is not yet proven.[see comment]: European Respiratory Journal. 23(5):735-40, 2004 May.
Sadafi, H. A., M. J. Bailey, A. W. Wood, and K. Wesnes, 2003, The brain effects of occupational strength 50 hz magnetic field.
Salamonsen, R. F., H.-G. Schneider, M. Bailey, and A. J. Taylor, 2005, Cardiac Troponin I Concentrations, but Not Electrocardiographic Results, Predict an Extended Hospital Stay after Coronary Artery Bypass Graft Surgery: Clin Chem, v. 51, p. 40-46.
Scheinkestel, C. D., F. Adams, L. Mahony, M. Bailey, A. R. Davies, I. Nyulasi, and D. V. Tuxen, 2003a, Impact of increasing parenteral protein loads on amino acid levels and balance in critically ill anuric patients on continuous renal replacement therapy.[see comment]: Nutrition, v. 19, p. 733-40.
Scheinkestel, C. D., M. Bailey, P. S. Myles, K. Jones, D. J. Cooper, I. L. Millar, and D. V. Tuxen, 1999, Hyperbaric or normobaric oxygen for acute carbon monoxide poisoning: a randomised controlled clinical trial.[see comment]: Medical Journal of Australia, v. 170, p. 203-10.
Scheinkestel, C. D., M. Bailey, P. S. Myles, K. Jones, D. J. Cooper, I. L. Millar, and D. V. Tuxen, 2000a, Hyperbaric or normobaric oxygen for acute carbon monoxide

poisoning: a randomized controlled clinical trial.[comment]: Undersea & Hyperbaric Medicine, v. 27, p. 163-4.

Scheinkestel, C. D., L. Kar, K. Marshall, M. Bailey, A. Davies, I. Nyulasi, and D. V. Tuxen, 2003b, Prospective randomized trial to assess caloric and protein needs of critically Ill, anuric, ventilated patients requiring continuous renal replacement therapy.[see comment]: Nutrition, v. 19, p. 909-16.

Scheinkestel, C. D., D. V. Tuxen, M. Bailey, P. S. Myles, K. Jones, D. J. Cooper, and I. L. Millar, 2000b, Hyperbaric oxygen in carbon monoxide poisoning. Authors of study clarify points that they made.[comment]: BMJ, v. 321, p. 109-10; author reply 110-1.

Serpell, J. W., P. W. Carne, and M. Bailey, 2003, Radical lymph node dissection for melanoma: ANZ Journal of Surgery, v. 73, p. 294-9.

Solomon, A., P. U. Cameron, M. Bailey, A. L. Dunne, S. M. Crowe, J. F. Hoy, and S. R. Lewin, 2003, Immunological and virological failure after antiretroviral therapy is associated with enhanced peripheral and thymic pathogenicity: Journal of Infectious Diseases, v. 187, p. 1915-23.

Voigt, T., M. Bailey, and M. Abramson, 1998, Air pollution in the Latrobe Valley and its impact upon respiratory morbidity: Australian & New Zealand Journal of Public Health, v. 22, p. 556-61.

Wilcox, G., P. D. Archer, M. Bailey, L. Dziukas, C. F. Lim, and H. G. Schneider, 2001, Measurement of cardiac troponin I levels in the emergency department: predictive value for cardiac and all-cause mortality.[see comment]: Medical Journal of Australia, v. 174, p. 170-3.

Wluka, A. E., S. R. Davis, M. Bailey, S. L. Stuckey, and F. M. Cicuttini, 2001, Users of oestrogen replacement therapy have more knee cartilage than non-users: Annals of the Rheumatic Diseases, v. 60, p. 332-6.

Wong, L. F., D. M. Taylor, and M. Bailey, 2004, Vagal response varies with Valsalva maneuver technique: a repeated-measures clinical trial in healthy subjects: Annals of Emergency Medicine. 43(4):477-82, 2004 Apr.

Woods, R. K., M. Abramson, M. Bailey, and E. H. Walters, 2001, International prevalences of reported food allergies and intolerances. Comparisons arising from the European Community Respiratory Health Survey (ECRHS) 1991-1994: European Journal of Clinical Nutrition, v. 55, p. 298-304.

Woods, R. K., M. Abramson, J. M. Raven, M. Bailey, J. M. Weiner, and E. H. Walters, 1998, Reported food intolerance and respiratory symptoms in young adults.[see comment]: European Respiratory Journal, v. 11, p. 151-5.

Woods, R. K., R. M. Stoney, P. D. Ireland, M. J. Bailey, J. M. Raven, F. C. Thien, E. H. Walters, and M. J. Abramson, 2002, A valid food frequency questionnaire for measuring dietary fish intake: Asia Pacific Journal of Clinical Nutrition, v. 11, p. 56-61.

Xiang, S., R. Denver, M. Bailey, and H. Krum, 2003, Physiologic determinants of endothelin concentrations in human saliva: Clinical Chemistry, v. 49, p. 2012-9.

Zheng L, W. E., Wang N, Whitford H, Orsida B, Levvey B, Bailey M, Williams TJ, Snell GI., 2004, Effect of inhaled fluticasone propionate on BAL TGF-beta(1) and bFGF concentrations in clinically stable lung transplant recipients.: Heart Lung Transplant, v. 23, p. 446-455.